Human Reproduction Open, pp. 1–14, 2020
doi:10.1093/hropen/hoaa058

human reproduction open

REVIEW

Concordance between systematic reviews of randomized controlled trials in assisted reproduction: an overview

Mariano Mascarenhas[1], Theodoros Kalampokas[2], Sesh Kamal Sunkara[3], and Mohan S. Kamath [4,*]

[1]Glasgow Centre for Reproductive Medicine, Glasgow, UK [2]2nd Department of Obstetrics and Gynecology-"Rea Maternity Hospital", University of Athens, Athens, Greece [3]Division of Women's Health, Faculty of Life Sciences and Medicine, King's College London, London, UK [4]Department of Reproductive Medicine, Christian Medical College, Vellore, India

*Correspondence address. Department of Reproductive Medicine, Christian Medical College, Vellore, India. E-mail:dockamz@gmail.com
https://orcid.org/0000-0002-0316-636X

Submitted on August 6, 2020; resubmitted on October 14, 2020; editorial decision on October 26, 2020

STUDY QUESTION: Are systematic reviews published within a 3-year period on interventions in ART concordant in their conclusions?

SUMMARY ANSWER: The majority of the systematic reviews published within a 3-year period in the field of assisted reproduction on the same topic had discordant conclusions.

WHAT IS KNOWN ALREADY: Systematic reviews and meta-analyses have now replaced individual randomized controlled trials (RCTs) at the top of the evidence pyramid. There has been a proliferation of systematic reviews and meta-analyses, many of which suffer from methodological issues and provide varying conclusions.

STUDY DESIGN, SIZE, DURATION: We assessed nine interventions in women undergoing ART with at least three systematic reviews each, published from January 2015 to December 2017.

PARTICIPANTS/MATERIALS, SETTING, METHODS: The systematic reviews which included RCTs were considered eligible for inclusion. The primary outcome was extent of concordance between systematic reviews on the same topic. Secondary outcomes included assessment of quality of systematic reviews, differences in included studies in meta-analyses covering the same search period, selective reporting and reporting the quality of evidence.

MAIN RESULTS AND THE ROLE OF CHANCE: Concordant results and conclusions were found in only one topic, with reviews in the remaining eight topics displaying partial discordance. The AMSTAR grading for the majority of the non-Cochrane reviews was critically low whilst it was categorized as high for all of the Cochrane reviews. For three of the nine topics, none of the included systematic reviews assessed the quality of evidence. We were unable to assess selective reporting as most of the reviews did not have a pre-specified published protocol.

LIMITATIONS, REASONS FOR CAUTION: We were limited by the high proportion of reviews lacking a pre-specified protocol, which made it impossible to assess for selective reporting. Furthermore, many reviews did not specify primary and secondary outcomes which made it difficult to assess reporting bias. All the authors of this review were Cochrane review authors which may introduce some assessment bias. The categorization of the review's conclusions as beneficial, harmful or neutral was subjective, depending on the tone and wording of the conclusion section of the review.

WIDER IMPLICATIONS OF THE FINDINGS: The majority of the systematic reviews published within a 3-year period on the same topic in the field of assisted reproduction revealed discordant conclusions and suffered from serious methodological issues, hindering the process of informed healthcare decision-making.

STUDY FUNDING/COMPETING INTEREST(S): All the authors are Cochrane authors. M.S.K. is an editorial board member of Cochrane Gynaecology and Fertility group. No grant from funding agencies in the public, commercial or not-for-profit sectors was obtained.

Key words: assisted reproduction / meta-analysis / infertility / overview / systematic review / concordance

WHAT DOES THIS MEAN FOR PATIENTS?

The number of studies being published on any intervention in fertility treatment has increased exponentially. Systematic reviews compile similar studies to provide a concise answer as to whether an intervention improves the success rates with fertility treatment or not. Recently, it has been noted that systematic reviews on the same topic published around the same time have different conclusions. We planned to assess the quality and the degree of agreement between systematic reviews on the same topic published around the same time. We systematically searched for all systematic reviews focusing on interventions involving IVF treatment published between January 2015 and December 2017 and found that nine interventions had at least three reviews published within this period. We found that most reviews on interventions in IVF treatment were of suboptimal quality as they were not methodologically sound and often gave contradictory answers to the same clinical question. This might lead to more confusion than clarity in determining whether an intervention is actually useful or not during IVF.

Introduction

One of the central underpinnings of 'evidence-based medicine' is the performance of a thorough literature review of the available scientific publications to answer a clinical question (Guyatt *et al.*, 1992). This has become progressively difficult due to the huge increase in research output resulting in an exponential increase in the number of scientific papers being published. There is hence a need to systematically search and appraise the studies and present the output on a topic in a manner that could be used as a 'ready reckoner' for clinicians, patients and healthcare providers.

Systematic reviews involve performing a systematic unbiased search of the available scientific literature to find studies that answer a clinical question and then dissect, analyse and compare these studies. Additionally, the results of these similar studies can be condensed through meta-analysis to provide a more precise estimate. The major advantage of systematic reviews is that the 'hard work' of the literature review is performed by the review authors and a condensed version of the research is available along with an unbiased critique of the available studies, which is essential for clinicians to make an informed choice.

Systematic reviews and meta-analyses have now replaced individual randomized controlled trials (RCTs) at the top of the evidence pyramid (Phillips *et al.*, 2009). Meta-analyses are also increasingly used by committees generating guidance and funding bodies to make decisions on healthcare policy decisions. However, increasing numbers of systematic reviews are now being published, frequently for the same clinical question and sometimes within a few months of each other. There has been a concern regarding the 'proliferation' of meta-analysis, which quite frequently vary in their included studies and sometimes dramatically vary in their conclusions (Greco *et al.*, 2015). Concerns have been raised that the proliferation of suboptimally conducted meta-analyses could compromise the value of this valuable tool (Humaidan and Polyzos, 2012).

Meta-analyses may suffer from discrepancies in the included studies and some exhibit bias in selective reporting and analyses which may mislead the reader (Page *et al.*, 2014). It is important to assess the concordance between meta-analyses especially in fast-evolving fields such as ART. As there has yet to be an overview assessing systematic review published in the field of assisted reproduction, this research project was planned to examine systematic reviews covering interventions in the context of ART. We aimed to systematically examine the quality of systematic reviews and the extent of agreement of conclusions between the meta-analyses published on the same topic in the similar time period.

Materials and methods

A comprehensive search of the literature was performed with a predefined protocol and search strategy to identify systematic reviews of interventions in ART over a 3-year period (January 2015 to December 2017). To allow meaningful comparisons between meta-analyses on the same topic, we only included topics with three or more published systematic reviews with similar eligibility criterias (assessed using the population, intervention, comparison and outcomes -PICO), each of which had to have included RCTs. If there were clinical relevant differences in more than one domain of the PICO framework (for e.g. population and intervention), we considered omitting those reviews as this would have led to differences in study selection leading to discordant results and conclusions. We excluded those reviews which included only non-randomized or cohort studies. The protocol for the literature search was pre-specified and registration of the protocol was initiated with PROSPERO on 2 March 2019. However, the PROSPERO administrators replied on 19 March 2019 that the protocol could not be registered due to lack of at least one outcome of direct patient or clinical relevance. Since it is an overview of published systematic reviews, ethics approval was not required.

Search strategy

The search terms (assisted reproductive techn*) OR IVF OR ICSI AND (meta-analysis OR systematic review) along with explosion of the MeSH terms was used to systematically search PUBMED/MEDLINE, EMBASE/SCOPUS, Cochrane Central Register of Controlled Trials (CENTRAL), The Cochrane Library (www.cochranelibrary.com), Database of Abstracts of Reviews of Effects (DARE), Citation indexes (http://scientific.thomson.com/products/sci/) and Conference abstracts in the Web of Science (http://wokinfo.com/).

We restricted the included studies to those published in the English language. We also searched using Google for grey literature, hand searched the reference lists of the included studies, and contacted experts in the field to identify any additional relevant reviews.

Two authors (M.M. and M.S.K.) independently scrutinized the titles and abstracts of the studies retrieved, and identified potential eligible meta-analyses for inclusion in this review. Disagreements were resolved by consensus or by discussing with the third author (S.K.S.). The meta-analyses were then grouped by topic or intervention. Those topics or interventions which had at least three meta-analyses were selected and full text of the articles were obtained. A data extraction form (detailed in Supplementary Table SI) was designed by consensus between the authors prior to performing the literature search.

Primary outcome

The primary outcome was the extent of concordance or discordance between systematic reviews on the same topic in their results (assessing the effect size and direction of effect) and conclusion.

Secondary outcomes

Secondary outcomes were: (i) discrepancies in the included randomized trials in reviews (trials missed in reviews covering the same time period); (ii) discrepancies between systematic review registry entries or protocols with the published systematic reviews in the PICO, study design change or new subgroup introduction/change (*post hoc* changes); (iii) the association between statistical significance and selective reporting such as the addition of a newer outcome or upgrading/downgrading; (iv) pooling of studies and outcome analysis performed as per intention to treat (ITT) or as per protocol analysis; (v) reporting the quality of evidence for outcomes using GRADE; and (vi) the overall quality of the systematic review as assessed by AMSTAR2.

Differences between pre-specified protocol and final review

We had initially planned to include topics with at least two reviews, but since the authors felt that two reviews may not be sufficient to allow for meaningful comparison, a decision was made to only include topics with at least three reviews.

Results

The literature search yielded 1748 records of which 148 were systematic reviews. These 148 systematic reviews were grouped by topics, with a total of 74 topics (Fig. 1). After initial screening, 79 records with fewer than two reviews on a topic were excluded. After further scrutiny, we excluded another 38 records for reasons as mentioned in Supplementary Table SII. Finally, we had nine topics with 31 reviews which were included in the current overview (Tables I and II).

Acupuncture in IVF

Three systematic reviews, all of which included only RCTs, examined the impact of acupuncture on IVF. Two of these reviews (Qian and Meng, 2016; Jo and Lee, 2017) had a similar search period and analysed similar populations with concordant conclusions that acupuncture is useful as an adjunct to IVF, but one of these reviews (Qian and Meng, 2016) included unselected IVF population whilst the other (Jo and Lee, 2017) included only women with polycystic ovary syndrome undergoing IVF. The authors of one review (Qian and Meng, 2016) cited their analyses of clinical pregnancy rates as evidence for the benefit of acupuncture. However, their analyses of the other primary outcome (i.e. live birth rate) failed to show a significant difference and this was not highlighted in the conclusion, raising the possibility of reporting bias. AMSTAR grading of all three reviews was critically low, with none of them reporting a pre-defined protocol and only one review (Jo and Lee, 2017) reporting the quality of evidence (Table I and Supplementary Table SIII).

Time lapse imaging

Four systematic reviews examined this intervention and all of them included only RCTs. There were partially discordant conclusions as three reviews were neutral (Armstrong *et al.*, 2015; Racowsky *et al.*, 2015; Chen *et al.*, 2017) and one was in favour of the intervention (Pribenszky *et al.*, 2017). Two of the systematic reviews with similar search periods (Armstrong *et al.*, 2015; Racowsky *et al.*, 2015) had partially discordant results but concordant conclusions. The other two systematic reviews (Chen *et al.*, 2017; Pribenszky *et al.*, 2017) had a similar search period, but the trials included were different, which led to a difference in the pooled results and partially discordant conclusions. AMSTAR grading was critically low for three systematic reviews (Racowsky *et al.*, 2015; Chen *et al.*, 2017; Pribenszky *et al.*, 2017) and high for the Cochrane review (Armstrong *et al.*, 2015). All four reviews reported the quality of evidence, but only the Cochrane review had a pre-specified protocol in place (Table I and Supplementary Table SIV).

Cleavage stage embryo transfer versus blastocyst transfer

Three reviews examined this intervention and conclusions were partially discordant as one review was neutral (Martins *et al.*, 2017) whilst the other two reviews were in favour of blastocyst stage over cleavage stage embryo transfers (Glujovsky *et al.*, 2016; Zhang *et al.*, 2017). One of the reviews included RCTs and non-randomized studies while the other two included only RCTs. Two systematic reviews (Glujovsky *et al.*, 2016; Martins *et al.*, 2017) with similar search periods had a difference in the number of included studies leading to difference in pooled results and partially discordant conclusions. AMSTAR grading was high for two systematic reviews, both of which had also reported the quality of evidence and had a pre-specified protocol (Glujovsky *et al.*, 2016; Martins *et al.*, 2017) and critically low for one systematic review, which did not report the quality of evidence or a pre-specified protocol (Zhang *et al.*, 2017) (Table I and Supplementary Table SV).

Preimplantation genetic screening

Four reviews examined preimplantation genetic screening as an intervention. The conclusions were partially discordant as one review was neutral (Lee *et al.*, 2015), and three reviews concluded that the intervention was beneficial (Chen *et al.*, 2015; Dahdouh *et al.*, 2015a,b). Three of the reviews included RCTs and cohort studies. Two reviews (Chen *et al.*, 2015; Dahdouh *et al.*, 2015a) had the same search period, but with different number of studies, and only one of them

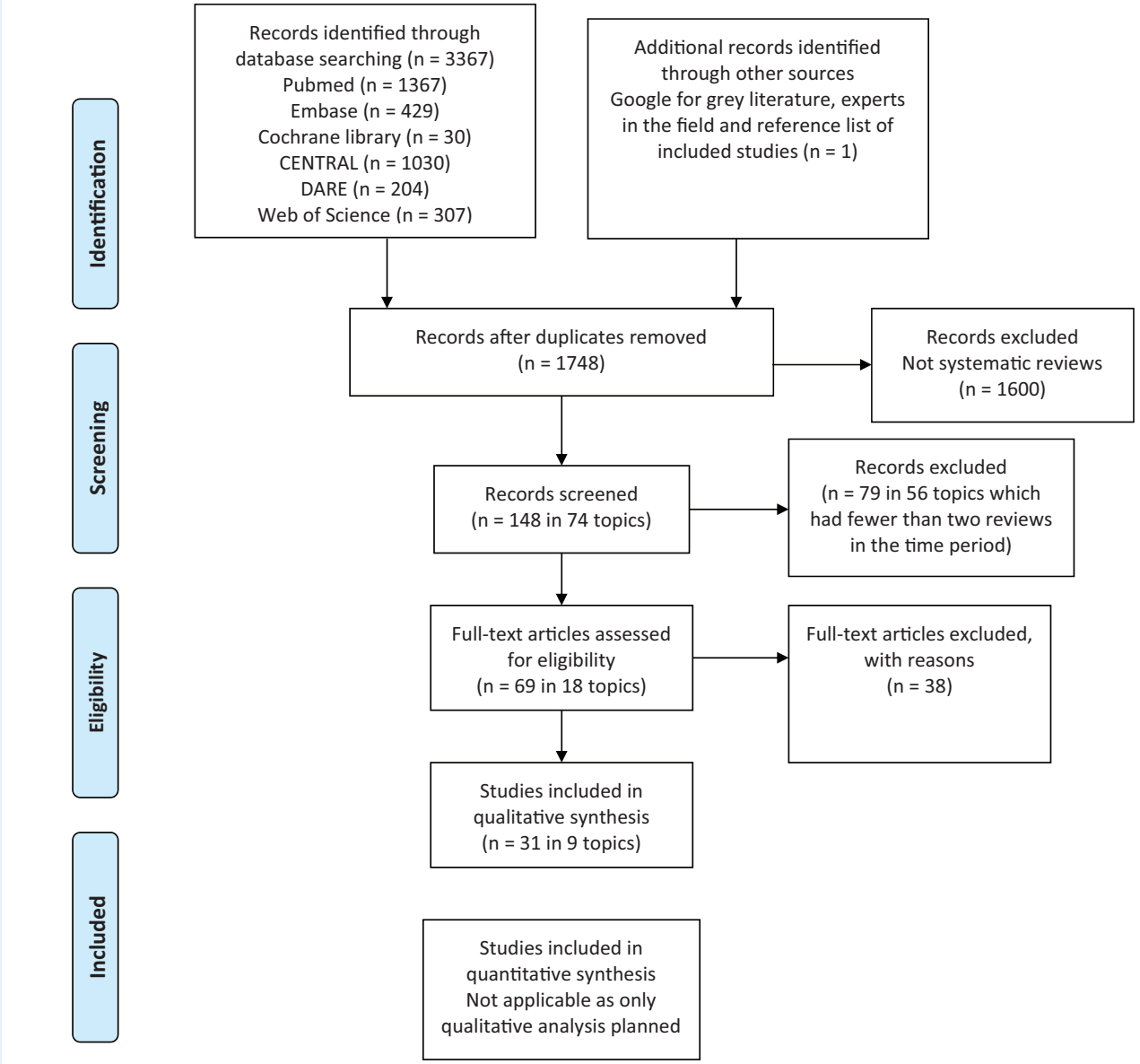

Figure 1. PRISMA flow diagram.

(Chen *et al.*, 2015) provided pooled results for clinical pregnancy and live birth. None of the systematic reviews reported quality of evidence or a pre-specified protocol and all had a critically low AMSTAR grading (Table I and Supplementary Table SVI).

Oestradiol supplementation in luteal phase

Partially discordant conclusions were noted from the three reviews on this topic with two being neutral (Huang *et al.*, 2015; Pinheiro *et al.*, 2017) and one in favour of the intervention (Zhang *et al.*, 2015). All the three reviews included only RCTs. Two systematic reviews had similar search periods, but had different numbers of included studies with one of these reviews (Huang *et al.*, 2015) providing separate results according to route of drug delivery, whilst the other review (Zhang *et al.*, 2015) pooled results across studies with different routes of drug delivery. AMSTAR score was critically low for all three reviews, with none reporting the quality of evidence or a pre-specified protocol (Table I and Supplementary Table SVII).

Growth hormone

Three reviews examining growth hormone had partially discordant conclusions with one being neutral (Yu *et al.*, 2015) whilst the other two were in favour of the intervention (Hart *et al.*, 2017; Li *et al.*, 2017b). One of the reviews included both RCTs and non-randomized studies (Yu *et al.*, 2015). Two systematic reviews were published in the same year (neither specified the search period), but had differences in the included studies and the method of data synthesis (per

Table I Summarizing final assessment of the systematic reviews published in field of assisted reproduction between January 2015 and December 2017.

Intervention	Number of SRs	Concordant or discordant or partial discordant (results)*	Reason	Concordant or discordant (conclusion)#	Reason (if discordant)	Remarks
Acupuncture	3	Partially discordant	Two SRs (Shen *et al.*, 2015; Qian and Meng, 2016) performed subgroup analyses based on timing of intervention but the study periods were different, leading to a differences in the included studies. The third SR (Jo and Lee, 2017) included only women with PCOS leading to differences with the other two reviews and did not perform subgroup analyses for the timing of intervention.	Partially discordant. Neutral (Shen *et al.*, 2015), in favour of intervention (Qian and Meng, 2016; Jo and Lee, 2017)	One SR (Shen *et al.*, 2015) concluded that the impact of the intervention depended on the timing, suggesting lack of evidence of benefit if performed during ET, despite their analyses in actuality showing a benefit if performed during ET. Other two SRs (Qian and Meng, 2016; Jo and Lee, 2017) highlighted overall benefit of the intervention in the population.	One SR reported quality of evidence (Jo and Lee, 2017). AMSTAR grading was critically low for all three SRs. Two SRs (Shen *et al.*, 2015; Qian and Meng, 2016) which performed subgroup analyses according to timing of intervention focused on different timings during an IVF cycle and had different search periods, while the third SR (Jo and Lee, 2017) did not assess timing of intervention which made comparison of the SRs difficult.
Time lapse imaging	4	Partially discordant	Two SRs (Armstrong *et al.*, 2015; Racowsky *et al.*, 2015) with similar search period had differences in included studies and outcomes reported, leading to differences in final results. Other two SRs (Chen *et al.*, 2017; Pribenszky *et al.*, 2017) with similar search period also had differences in included trials for main outcome (LBR). Different outcomes reported by two SRs (clinical pregnancy vs ongoing pregnancy rate) leading to different results.	Partially discordant. Neutral (Armstrong *et al.*, 2015; Racowsky *et al.*, 2015; Chen *et al.*, 2017), in favour of intervention (Pribenszky *et al.*, 2017)	Two SRs (Armstrong *et al.*, 2015; Racowsky *et al.*, 2015) with similar search period had concordant conclusions even though the results were not concordant. The other two SRs (Chen *et al.*, 2017; Pribenszky *et al.*, 2017) had different conclusions due to differences in the results.	All four SRs reported quality of evidence. AMSTAR grading was critically low for three SRs (Racowsky *et al.*, 2015; Chen *et al.*, 2017; Pribenszky *et al.*, 2017) and high for one (Armstrong *et al.*, 2015).
Cleavage stage embryo transfer vs blastocyst transfer	3	Partially discordant	Two SRs (Glujovsky *et al.*, 2016; Martins *et al.*, 2017) had discordant results due to differences in number of included studies despite search periods being identical. Differences in denominators while pooling results since one of the SR (Glujovsky *et al.*, 2016) did ITT analysis, while other (Martins *et al.*, 2017) performed per protocol analysis. Third SR (Zhang *et al.*, 2017) did not report main outcomes of interest and reported only one outcome.	Partially discordant. Neutral (Martins *et al.*, 2017), in favour of intervention (Glujovsky *et al.*, 2016; Zhang *et al.*, 2017)	One SR (Martins *et al.*, 2017) concluded no benefit of intervention, while the other SR (Glujovsky *et al.*, 2016) suggested improvement in LBR but highlighted uncertainty due to no improvement in cumulative LBR. Third SR (Zhang *et al.*, 2017) suggested intervention (blastocyst transfer) is beneficial due to lower ectopic pregnancies.	Two of the SRs (Glujovsky *et al.*, 2016; Martins *et al.*, 2017) reported level of evidence. AMSTAR grading was high for two SRs (Glujovsky *et al.*, 2016; Martins *et al.*, 2017) and critically low for one SR (Zhang *et al.*, 2017). One SR (Zhang *et al.*, 2017) had a very narrow focus, and reported only one outcome (ectopic pregnancy), while other clinically relevant outcomes were available such as LBR/CPR.

(continued)

Table I Continued

Intervention	Number of SRs	Concordant or discordant or partial discordant (results)*	Reason	Concordant or discordant (conclusion)#	Reason (if discordant)	Remarks
Pre-implantation genetic screening	4	Unable to assess (two SRs did not perform the meta-analysis, and other two (Dahdouh *et al.*, 2015b; Lee *et al.*, 2015) did not perform meta-analysis for the same outcomes)	Two of SRs (Lee *et al.*, 2015; Dahdouh *et al.*, 2015b) did not perform meta-analysis even though more than two RCTs were available. One SR (Dahdouh *et al.*, 2015a) performed meta-analysis for less common outcome (clinical IR and sustained IR). Only fourth SR (Chen *et al.*, 2015) performed meta-analysis for outcomes of interest.	Partial discordant. Neutral (Lee *et al.*, 2015), in favour of intervention (Chen *et al.*, 2015; Dahdouh *et al.*, 2015a,b)	One SR (Lee *et al.*, 2015) was inconclusive about the beneficial role of PGS, while other three SRs (Chen *et al.*, 2015; Dahdouh *et al.*, 2015a,b) suggested beneficial role of the intervention with one of SR (Dahdouh *et al.*, 2015b) drawing this conclusion based on only qualitative assessment.	None of SRs assessed the quality of evidence. All four SRs were critically low when AMSTAR grading was done. Two of SRs did not perform quantitative synthesis even though >2 RCTs were available.
Oestradiol supplementation	3	Partially discordant	One SR (Huang *et al.*, 2015) reported pooled results from different subgroups while other SR (Zhang *et al.*, 2015) reported combined results irrespective of route of drug delivery. Differences in number of included studies noted despite similar search period. Third SR did not pool the results (Pinheiro *et al.*, 2017).	Partially discordant. Neutral (Huang *et al.*, 2015; Pinheiro *et al.*, 2017), in favour of intervention (Zhang *et al.*, 2015)	One SR(Huang *et al.*, 2015) concluded that there was no benefit with the intervention, while the other SR (Zhang *et al.*, 2015) concluded that there was a benefit. Third SR (Pinheiro *et al.*, 2017) had a balanced conclusion highlighting lack of benefit in terms of improvement in clinical pregnancy.	None of the SRs described quality of evidence. AMSTAR score was critically low for all three SRs (Huang *et al.*, 2015; Zhang *et al.*, 2015; Pinheiro *et al.*, 2017). One SR (Pinheiro *et al.*, 2017) did not conduct a meta-analysis despite four study results being available for pooling.
Growth hormone	3	Partially discordant	One SR (Li *et al.*, 2017b) reported per protocol analysis, while another SR reported ITT (Hart *et al.*, 2017). Both the SRs were published in the same year but there were differences in the included studies. Third SR (Yu *et al.*, 2015) did not report primary outcome of LBR and reported CPR as risk difference.	Partially discordant. Neutral (Yu *et al.*, 2015), in favour of intervention (Hart *et al.*, 2017; Li *et al.*, 2017b)	One SR (Hart *et al.*, 2017) concluded that LBR did not improve but there was improvement in CPR and presented a balanced conclusion. Other SR (Li *et al.*, 2017b) concluded significant improvement in LBR and CPR. Third SR (Yu *et al.*, 2015) focused primarily embryological outcomes and found no significant differences in CPR.	None of the SRs described quality of evidence. AMSTAR grading was critically low for all three SRs.
Inositol	3	Partially discordant	One SR (Zheng *et al.*, 2017) included RCTs and cohort study, while other SRs (Mendoza *et al.*, 2017; Showell *et al.*, 2017) restricted to only RCTs. Two SRs (Mendoza *et al.*, 2017; Zheng *et al.*, 2017) had similar search period, but had differences in the number of included studies and differences in denominator despite similar (per protocol) analysis being done by both SRs. Third SR (Showell *et al.*, 2017) presented ITT analysis and differences was noted in number of studies included compared to other SRs.	Partially discordant. Neutral (Mendoza *et al.*, 2017; Showell *et al.*, 2017), in favour of intervention (Zheng *et al.*, 2017)	One of SRs (Zheng *et al.*, 2017) suggested improvement in clinical pregnancy, while other SR (Mendoza *et al.*, 2017) suggested no benefit. The third SR (Showell *et al.*, 2017) did not specifically focus on Inositol hence not considered under conclusion.	One SR (Showell *et al.*, 2017) described the grade of evidence. AMSTAR grading was high for one SR (Showell *et al.*, 2017), while critically low for other two SRs (Mendoza *et al.*, 2017; Zheng *et al.*, 2017).

(continued)

Table I Continued

Intervention	Number of SRs	Concordant or discordant or partial discordant (results)*	Reason	Concordant or discordant (conclusion)#	Reason (if discordant)	Remarks
Granulocyte-colony stimulating factor	4	Concordant	Three SRs (Zhao *et al.*, 2016; Xie *et al.*, 2017; Li *et al.*, 2017a) included RCTs and non-RCTs and did not pool results from RCTs separately leading to large differences in final pooled results. Furthermore, two SRs (Zhao *et al.*, 2016; Li *et al.*, 2017a) pooled studies with heterogeneous population (RIF, thin endometrium, unselected population), while one SR (Xie *et al.*, 2017) included studies with only thin endometrium. Fourth SR (Kamath *et al.*, 2017) presented pooled results separately for RIF and thin endometrium. Two SRs (Zhao *et al.*, 2016; Kamath *et al.*, 2017), which reported pooled results for RIF had some differences in included studies and differences in numbers from the same study.	Concordant. All four SRs in favour of intervention (Zhao *et al.*, 2016; Kamath *et al.*, 2017; Xie *et al.*, 2017; Li *et al.*, 2017a)	While there was differences in inclusion criteria (difference designs, subpopulations), there was concordance between the reviews of a possible benefit with G-CSF.	Only one SR (Kamath *et al.*, 2017) described the grade of evidence. AMSTAR grading for one of the SR (Kamath *et al.*, 2017) was categorized as low and critically low for the three remaining SRs (Zhao *et al.*, 2016; Xie *et al.*, 2017; Li *et al.*, 2017a).
Dehydroepiandrosterone	4	Partially discordant	Three of the SRs (Nagels *et al.*, 2015; Zhang *et al.*, 2016; Qin *et al.*, 2017) had similar search periods but showed differences in the included studies. Two SRs (Nagels *et al.*, 2015; Qin *et al.*, 2017) which had presented pooled data from RCTs separately had different results due to significant differences in number of included RCTs despite the search periods being identical. However, the results from pooled results of all included studies (RCT and non RCTs) from the same SR (Qin *et al.*, 2017) were identical to other three SRs (Li *et al.*, 2015; Nagels *et al.*, 2015; Zhang *et al.*, 2016).	Concordant. All four SRs in favour of intervention (Li *et al.*, 2015; Nagels *et al.*, 2015; Zhang *et al.*, 2016; Qin *et al.*, 2017)	Heterogeneity in study designs between SRs was noted with one SR (Nagels *et al.*, 2015) including only RCTs on both poor and normal responders. The other three SRs (Li *et al.*, 2015; Zhang *et al.*, 2016; Qin *et al.*, 2017) included non-RCTs and RCTs but only studies examining poor responders/diminished ovarian reserve. Despite such heterogeneity in included studies, all four SRs had similar conclusions indicating benefit of the intervention.	Only one SR (Nagels *et al.*, 2015) described the GRADE of evidence. AMSTAR grading was high for one of the SRs (Nagels *et al.*, 2015), while it was critically low for other three SRs (Li *et al.*, 2015; Zhang *et al.*, 2016; Qin *et al.*, 2017).

CPR, clinical pregnancy rate; ET, embryo transfer; G-CSF, granulocyte colony-stimulating factor; IR, implantation rate; ITT, intention to treat; LBR, live birth rate; PCOS, polycystic ovary syndrome; PGS, preimplantation genetic screening; RCT, randomized controlled trial; RIF, recurrent implantation failure; SR, systematic review.

*Results were considered to be concordant when the direction of effect on pooling the studies were similar between systematic reviews (all the studies showing significant benefit or harm), discordant when the direction of effect was opposite to one another (such as one study showing a significant benefit from the intervention and another showing a significant evidence of harm) and partially discordant when the direction of effect was neutral (no apparent benefit or harm due to the intervention) in some studies and in one direction in others(significant benefit or harm).

#Conclusions were considered concordant when the reviews have similar conclusions on the usefulness of the intervention—in favour, neutral or against an intervention. Conclusions were considered discordant when they were in opposition (one review in favour and the other against an intervention), and conclusions were considered partially discordant when some reviews were in favour or against an interventions and others were neutral.

Table II Summarizing characteristics of the systematic reviews published in field of assisted reproduction between January 2015 and December 2017.

Intervention	Systematic review	Number of studies included	Method of data synthesis	Pooled results for live birth rate	Conclusion	AMSTAR grading
Acupuncture	Shen *et al.* (2015)	14 RCTs	ITT	Not assessed	No significant difference if acupuncture performed only at embryo transfer. Improves clinical pregnancy rate if performed at other times.	Critically low
	Jo and Lee (2017)	4 RCTs	Not specified but appears to be ITT	RR 1.61 (95% CI 0.73 to 3.58); 1 RCT; n = 66	Overall, acupuncture improves clinical pregnancy rate, ongoing pregnancy rates and reduces risk of OHSS.	Critically low
	Qian and Meng (2016)	30 RCTs	Per protocol	OR 0.83 (95% CI 0.68 to 1.02); 5 RCTs; n = 1747	Overall, acupuncture improves clinical pregnancy rate during IVF.	Critically low
Time lapse imaging	Racowsky *et al.* (2015)	4 RCTs	Appears per protocol (not clearly specified)	LBR not assessed	The authors did not support the use of routine TLI due to low-quality evidence of benefit.	Critically low
	Armstrong *et al.* (2015)	3 RCTs	Per couple—ITT	OR 1.1, (95% CI 0.45 to 2.73); 1 RCT; n = 76	Authors reported insufficient evidence for use of TLI.	High
	Chen *et al.* (2017)	10 RCTs	Per woman (per oocyte was also separately reported in the full study)	RR 1.23, (95% CI 1.06 to 1.44); 1 RCT; n = 842	Authors reported insufficient evidence for use of TLI.	Critically low
	Pribenszky *et al.* (2017)	5 RCTs	ITT	OR: 1.66, (95% CI 1.13 to 2.45); 3 RCTs; n = 481	Authors reported significant improvement in clinical outcomes.	Critically low
Cleavage stage embryo transfer vs. blastocyst transfer	Martins *et al.* (2017)	12 RCTs	Not specified; does not appear to be ITT	RR 1.11 (95% CI 0.92 to 1.35); 10 RCTs; n = 1940	No superiority of blastocyst stage over cleavage stage transfer.	High
	Glujovsky *et al.* (2016)	27 RCTs	ITT	OR 1.48 (95% CI 1.2 to 1.82); 13 RCTs; n = 1630	Low-quality evidence for live birth and moderate-quality evidence for clinical pregnancy that fresh blastocyst stage transfer is associated with higher rates than cleavage stage transfer.	High
	Zhang *et al.* (2017)	23 studies (7 RCTs)	ITT not done; main outcome (ectopic pregnancy) reported per pregnancy	Not assessed	Blastocyst stage reduced ectopic pregnancy compared to cleavage stage.	Critically low
Pre-implantation genetic screening	Dahdouh *et al.* (2015b)	3 RCTs	Meta-analysis not performed	Not assessed	PGS improves implantation and ongoing pregnancy rate.	Critically low
	Lee *et al.* (2015)	3 RCTs and 16 observational studies	Meta-analysis not performed	Not assessed	PGS increases implantation rates but insufficient clinical data available for evaluation.	Critically low
	Dahdouh *et al.* (2015a)	3 RCTs (8 observational studies additionally analysed)	Implantation rates reported as gestational sacs per embryo transferred	Not assessed	PGS improves clinical and sustained implantation rate.	Critically low
	Chen *et al.* (2015)	4 RCTs (7 cohort studies additionally analysed)	ITT	RR 1.26 (95% CI 1.05 to 1.50); 1 RCT; n = 155	PGS improves implantation rate.	Critically low
Oestradiol supplementation	Huang *et al.* (2015)	15 RCTs	ITT	Not assessed	No improvement in IVF outcomes.	Critically low
	Zhang *et al.* (2015)	11 RCTs	ITT	Not assessed	Significantly better clinical pregnancy rates following E2.	Critically low
	Pinheiro *et al.* (2017)	4 RCTs	Meta-analysis not performed	Not assessed	Higher implantation rate but no increased clinical pregnancy rate.	Critically low

(continued)

Table II Continued

Intervention	Systematic review	Number of studies included	Method of data synthesis	Pooled results for live birth rate	Conclusion	AMSTAR grading
Growth hormone	Hart *et al.* (2017)	12 RCTs	Not specified (appears to be ITT)	OR 1.54 (95% CI 0.86 to 2.74); 8 RCTs; n = 540	No difference in LBR, better CPR; increase in oocyte yield.	Critically low
	Li *et al.* (2017b)	11 RCTs	Not specified (appears to be per protocol analysis)	RR 1.65 (95% CI 1.23 to 2.22); 10 RCTs; n = 579	Higher LBR and CPR following GH.	Critically low
	Yu *et al.* (2015)	11 studies (6 RCTs)	Not specified (unclear whether ITT was done)	Risk difference 0.051 (95% CI −0.033 to 0.134); 5 studies; n = 422	No significant improvement in CPR; improvement in oocyte/embryo yield.	Critically low
Inositol	Zheng *et al.* (2017)	7 studies (6 RCTs)	Per protocol	Not assessed	Myoinositol increased clinical pregnancy rate in IVF and may improve embryo quality.	Critically low
	Mendoza *et al.* (2017)	8 RCTs	Per protocol	Not assessed	Myoinositol does not improve oocyte or embryo quality or clinical pregnancy rate for women with PCOS undergoing ICSI.	Critically low
	Showell *et al.* (2017)	4 RCTs	ITT	Not assessed	Conclusion was based on combined impact of all different type of antioxidants on IVF treatment outcome.	High
Granulocyte-colony stimulating factor	Kamath *et al.* (2017)	4 studies (including 2 RCTs)	ITT	Not assessed	Low-quality evidence for improvement in clinical pregnancy rate in RIF.	Low
	Zhao *et al.* (2016)	6 studies (including 2 RCTs for in recurrent implantation failure (results presented separately) and 1 RCT for unselected population)	Per protocol	Not assessed	Subcutaneous GCSF(RIF), but not intrauterine improves clinical pregnancy rate.	Critically low
	Xie *et al.* (2017)	11 studies (including 1 RCT)	Per protocol	Not assessed	G-CSF improves endometrial thickness, clinical pregnancy rate and implantation rate in thin endometrium.	Critically low
	Li *et al.* (2017a)	6 studies (including 3 RCTs)	Unclear whether ITT was done	Not assessed	G-CSF improves clinical pregnancy rate in IVF especially in women with thin endometrium and RIF.	Critically low
Dehydroepiandrosterone	Li *et al.* (2015)	8 studies (including 2 RCTs)	Unclear if ITT was done	Not assessed	DHEA increases clinical pregnancy rate in poor responders and has a positive impact on outcomes.	Critically low
	Nagels *et al.* (2015)	12 RCTs	ITT	composite LBR/OPR, OR 1.81, (95% CI 1.25 to 2.62); 8 RCTs; n = 878 participants	Moderate-quality evidence for improvement in live birth rates with DHEA in women identified as poor responders (however, the pooled results included few trials with normal responders as well).	High
	Qin *et al.* (2017)	9 studies (including 4 RCTs)	Per protocol analysis	Not assessed	DHEA might improve outcomes for poor responders having IVF/ICSI.	Critically low
	Zhang *et al.* (2016)	21 studies (including 8 RCTs)	ITT	RR 1.87, (95% CI 1.22 to 2.88); 6 studies; n = 528	DHEA associated with increased CPR/LBR and seems to improve ovarian reserve and IVF/ICSI outcomes.	Critically low

CPR, clinical pregnancy rate; DHEA, dehydroepiandrosterone; E2, oestradiol; G-CSF, granulocyte colony-stimulating factor; ITT, intention to treat; LBR, live birth rate; OHSS, ovarian hyperstimulation syndrome; OPR, ongoing pregnancy rate; OR, odds ratio; PCOS, polycystic ovary syndrome; PGS, preimplantation genetic screening; RCT, randomized controlled trial; RR, relative risk; TLI, time lapse imaging.

protocol versus ITT). AMSTAR grading was critically low for all three systematic reviews, with none reporting the quality of evidence or a pre-specified protocol (Table I and Supplementary Table SVIII).

Inositol

The conclusions were partially discordant between the three reviews examining this intervention, with two systematic reviews being neutral (Mendoza et al., 2017; Showell et al., 2017) and one being in favour of inositol (Zheng et al., 2017). One of the reviews included RCTs and non-randomized studies (Zheng et al., 2017). One of these three reviews was a Cochrane review which had a high AMSTAR grading (Showell et al., 2017) whilst the other two reviews had a critically low AMSTAR grading (Mendoza et al., 2017; Zheng et al., 2017). Only the Cochrane review (Showell et al., 2017) amongst the three reviews reported a pre-specified protocol and the quality of evidence (Table I and Supplementary Table SIX).

Granulocyte-colony stimulating factor

All four reviews had concordant conclusions in favour of the intervention (Zhao et al., 2016; Kamath et al., 2017; Xie et al., 2017; Li et al., 2017a). All four reviews included RCTs and non-randomized studies. There were differences in populations analysed (recurrent implantation failure, thin endometrium), route of administration and designs of studies (RCTs and non-RCTs) between the reviews. One of the reviews (Kamath et al., 2017) had a low AMSTAR grading and reported a pre-specified protocol and provided quality of evidence, whilst the other three reviews (Zhao et al., 2016; Xie et al., 2017; Li et al., 2017a) had critically low AMSTAR grading and none provided a pre-specified protocol nor the quality of evidence (Table I and Supplementary Table SX).

Dehydroepiandrosterone

All four systematic reviews on this topic were in favour of the intervention (Li et al., 2015; Nagels et al., 2015; Zhang et al., 2016; Qin et al., 2017) with concordant conclusions but partially discordant results. Three of the four reviews included RCTs and non-randomized studies. Three of the systematic reviews (Nagels et al., 2015; Zhang et al., 2016; Qin et al., 2017) had similar search periods but showed discrepancies in the included studies. One of the four reviews was a Cochrane review (Nagels et al., 2015) which reported the quality of evidence, had a pre-specified protocol and had a high AMSTAR grading. The other three systematic reviews (Li et al., 2015; Zhang et al., 2016; Qin et al., 2017) had a critically low AMSTAR gradings and none provided a pre-specified protocol nor the quality of evidence (Table I and Supplementary Table SXI).

Discussion

This overview critically appraised between three to four systematic reviews covering nine topics on ART and found concordant results and conclusions for only one topic with reviews on the remaining eight topics displaying partial discordance. The AMSTAR gradings for the majority of the non-Cochrane reviews were critically low whilst categorized as high for all the Cochrane reviews. For three of the nine topics, none of the included systematic reviews assessed the quality of evidence. The primary reasons for discordant results and conclusions amongst reviews with the same search period appeared to be differences in included trial numbers, discrepancies in data extraction, different approaches for data synthesis and absence of quality assessment tools leading to heterogeneity in final conclusions. Only two non-Cochrane reviews had a pre-specified protocol.

In the field of assisted reproduction, an opinion paper highlighted the issues surrounding quality of meta-analysis by citing an example of three systematic reviews published on the same topic, 'endometrial scratch', within a year i.e., 2012, with each review using different eligibility and study designs to answer the same clinical question (Simon and Bellver, 2014). There are few overviews which have compared inconsistencies in findings of systematic reviews published on identical topics with similar search periods in other specialities but none in the field of assisted reproduction (Linde and Willich, 2003; Shrier et al., 2008; Bolland and Grey, 2014). A study examining whether systematic reviews on the same clinical question can produce different answers included 17 topics in complimentary medicine which were addressed by two to five reviews published within 4-year time period (Linde and Willich, 2003). The most common discrepancy was differences in number of included studies due to differences in the inclusion criteria. The quality assessment tool for primary studies varied between the reviews, but there was broad agreement about the overall quality. The effect sizes after pooling of study results varied mainly due to differences in number of included studies. The authors concluded that there are large scale discrepancies in the conclusions of systematic reviews examining the same clinical question. They attributed the discordance to decisions taken during planning, performance and interpretation. These conclusions are broadly in agreement with the current study findings of widespread discordance in results and conclusions between the systematic reviews. We found discordance even when reviews with similar search period were compared, which ideally would have resulted in lower discordance in included trial numbers.

There may be discordance in conclusion due to differences in interpretation of identical data from pooled results by different reviewers. This difference could be due to varying level of expertise, value systems and personal preferences (Shrier et al., 2008). A study in 2008 investigated the level of discrepancies in interpreting pooled data which were identical (Shrier et al., 2008). Pooled data from 23 RCTs (meta-analysis from 3, and subsequently, 5, 10, 20 and 23 RCTs performed in a stepwise manner) on a clinical question was presented to eight experienced reviewers. The reviewers were asked to respond to questions which were linked to the effectiveness of the intervention and whether they would recommend the same in clinical practice. The authors found significant disagreements in the responses from the reviewers for the initial pooled results (10 RCTs) which had similar effect sizes (whether a fixed or random model was used) and minimal statistical heterogeneity. The disagreement among the reviewers increased when pooled results from 20 RCTs with divergent effect sizes (due to use of different models) and higher statistical heterogeneity were presented. The authors highlighted the subjective nature of interpreting identical data even when experienced reviewers were involved. Furthermore, there was a lack of consensus even when GRADE was used. Another overview studied seven overlapping meta-analyses from high-quality journals published on the same topic using AMSTAR tool to assess the quality of reviews (Bolland and Grey, 2014). Despite

majority of the reviews being categorized as high quality after applying the AMSTAR, the authors found discrepancies in study inclusion, data extraction and analytical methods. The relative risks estimated for the same trial varied between the meta-analysis due to use of different approaches, i.e. per protocol analysis versus ITT which is the recommended approach. There was a trend for reviews which adhered to recommended approaches and fulfilled the majority of the AMSTAR criteria to report more conservative conclusions. The current study found that for similar pooled results, the conclusions in the systematic reviews were more guarded when the quality of evidence was described using GRADE compared with those reviews which did not assess the quality of evidence.

Each of the steps involved in planning, executing and interpreting data during a systematic review is a potential source for discordance (Jadad *et al.*, 1997). For the same topic, quite often, there are differences in eligibility criteria with some reviews having a narrow focus as compared with others which have broader eligibility. This is clearly illustrated in numerous systematic reviews published in the last 3 years on 'endometrial scratch for IVF' with populations varying from unselected groups to those women with one or more than two IVF failures (Vitagliano *et al.*, 2018, 2019; Gui *et al.*, 2019; Sar-Shalom Nahshon *et al.*, 2019; van Hoogenhuijze *et al.*, 2019). Due to the publishing pressures in the academic world, researchers might be tempted to make subtle changes in eligibility criteria to bring in the novelty factor leading to varying 'themes' of reviews on the same topic being published (Rawat and Meena, 2014). These minor changes may not be readily apparent to the casual reader leading to a lack of clarity in understanding the impact of interventions. This can be clearly avoided by embarking on comprehensive systematic review capturing various subpopulations of interest. Replication of scientific work is an important part of research methodology. Since systematic reviews often form the basis for health policies and clinical decision-making, there is on-going effort to build consensus on when to replicate reviews and reduce research wastage (Tugwell *et al.*, 2020). Recently, a consensus checklist has been published as a guidance tool to decide whether a systematic review needs replication (Tugwell *et al.*, 2020). The checklist incorporates factors such as priority for replication, addressing uncertainties arising from previous review, potentially sizeable impact on a large population and optimization of resources. Another source of variation in the included trial numbers between the reviews is due to differences in the number of databases searched, language restrictions and inclusion of data from ongoing trials. Data extraction is an important source of discrepancy. This could be due to lack of clarity in the data presented in the primary study, efforts made to contact authors for clarification and use different approaches for data synthesis such as 'per protocol analysis' or 'as treated'. This calls for uniform measures such as meticulous and transparent data extraction which can be duplicated, actively seeking clarification from authors whenever required and adherence to recommended approach, i.e. ITT analysis for data synthesis. For peer reviewers evaluating systematic reviews with large number of trials, it may not be possible to cross check the validity of extracting data and, very often, discrepancies in the data extraction are unravelled by authors working on another systematic review or updating the previous review. A contributory factor to differences in conclusion might stem from an incorrect interpretation of statistical results by the reviewers. In particular, it is possible for an imprecise, non-significant result of a meta-analyses secondary to insufficient sample size to be interpreted as 'no effect' or 'no difference' whilst the correct interpretation would be that there is insufficient data to know whether there is an impact or not.

In the current overview, we found the majority of the reviews did not assess the quality of evidence. The lack of assessment of quality increases heterogeneity in the interpretation of results and magnifies the discordances in the conclusions. The systematic reviews should incorporate risk of bias and quality assessment which can help in bringing uniformity in the way results are summarized. There have been efforts to standardize the way results are described depending on the effect size and quality of evidence (Ryan *et al.*, 2016). These measures will help reduce the discordance within different reviews with similar results.

The current overview is the first to assess and compare systematic reviews in the field of ART, where there is a predilection for introducing newer unproven interventions in routine clinical practice without rigorous vetting (Wilkinson *et al.*, 2019). AMSTAR was used for objective grading of the reviews. Importantly, we captured each review's key conclusion as to whether the intervention discussed was beneficial, harmful or neutral (unclear impact), as that would be the key message that clinicians would be seeking from a systematic review. We were limited by the high proportion of reviews lacking a pre-specified protocol, which made it impossible to assess for selective reporting. Furthermore, many reviews did not specify primary and secondary outcomes which made it difficult to assess reporting bias. All the authors of this review were Cochrane review authors which may introduce some assessment bias. However, AMSTAR grading was used as an objective method for assessing the review quality. The selection of reviews was based on broad similarities in eligibility criteria based on the PICO framework. However, it is possible that the systematic reviews with some heterogeneity due to subtle differences in eligibility criterias may have been included for the comparison and the discordant results obtained could have been due to those minor variations in the selection criteria. This 'selection bias' could be considered a possible limitation of the current study. The categorization of the review's conclusions as beneficial, harmful or neutral was subjective depending on the tone and wording of the conclusion section of the review. Despite the search being comprehensive, it is possible that the reviews might have been missed as we had restricted to reviews published in English language. Restricting the analyses to topics with at least three systematic reviews published within a defined time-frame might have skewed the framework of this project towards topics which are emerging or contentious rather than relatively established techniques. The acceptability of the findings of a review might be influenced by pre-held notions, leading to either confirmation bias (when the findings are in concordance with their notion) or cognitive dissonance (when the findings are discordant to their pre-held notion). This might lead to initiation of a new review process to either 'disprove' or 'prove' findings from an earlier review.

Finally, we would like to emphasize that the current study findings do not undermine the importance of appropriately conducted systematic reviews as a research tool. The 'weaknesses' of systematic reviews that we have pointed out are primarily due to lack of strict adherence to proper research methodology when conducting systematic reviews. We suggest that only comprehensive systematic reviews with prospectively registered protocols and good methodological rigour should be considered for publication.

Conclusion

The majority of the systematic reviews published within a 3-year time period in the field of assisted reproduction on the same topic revealed discordant conclusions and suffered from serious methodological issues, hindering the process of informed health care decision-making. Differences in search results (despite similar search periods) and a lack of methodological rigour in quality assessment of the included studies and data synthesis, appear to lead on to differences in the interpretation of the results, which in turn leads to discordant conclusions amongst reviews on the same topic. The journal editors and reviewers need to be more stringent during the peer review process and allow only methodologically sound, comprehensive systematic reviews for publication. Clinicians need to be cautious in applying the results of a systematic review of 'questionable provenance' in their clinical practice.

Supplementary data

Supplementary data are available at *Human Reproduction* online.

Acknowledgements

The authors thanks Richard Kirubakaran, from Prof. B.V. Moses Centre for Evidence Informed Health Care & Health Policy, Christian Medical College, Vellore, for initial inputs.

Authors' roles

S.K.S. and M.S.K. conceived the idea. M.S.K. and M.M. performed the systematic search and M.M., T.K. and M.S.K. performed the data extraction and analyses. M.M., M.S.K., T.K. and S.K.S. contributed to the writing of the article and gave critical inputs. All authors approved the final version of the manuscript.

Funding

No grant from funding agencies in the public, commercial or not-for-profit sectors was obtained.

Conflict of interest

All the authors are Cochrane authors. M.S.K. is an editorial board member of the Cochrane Gynaecology and Fertility group.

References

Armstrong S, Arroll N, Cree LM, Jordan V, Farquhar C. Time-lapse systems for embryo incubation and assessment in assisted reproduction. *Cochrane Database Syst Rev* 2015;**2**:CD011320.

Bolland MJ, Grey A. A case study of discordant overlapping meta-analyses: vitamin D supplements and fracture. *PLoS One* 2014;**9**: e115934.

Chen M, Wei S, Hu J, Quan S. Can comprehensive chromosome screening technology improve IVF/ICSI outcomes? A meta-analysis. *PLoS One* 2015;**10**:e0140779.

Chen M, Wei S, Hu J, Yuan J, Liu F. Does time-lapse imaging have favorable results for embryo incubation and selection compared with conventional methods in clinical *in vitro* fertilization? A meta-analysis and systematic review of randomized controlled trials. *PLoS One* 2017;**12**:e0178720.

Dahdouh EM, Balayla J, Garcia-Velasco JA. Comprehensive chromosome screening improves embryo selection: a meta-analysis. *Fertil Steril* 2015a;**104**:1503–1512.

Dahdouh EM, Balayla J, García-Velasco JA. Impact of blastocyst biopsy and comprehensive chromosome screening technology on preimplantation genetic screening: a systematic review of randomized controlled trials. *Reprod Biomed Online* 2015b;**30**:281–289.

Glujovsky D, Farquhar C, Quinteiro Retamar AM, Alvarez Sedo CR, Blake D. Cleavage stage versus blastocyst stage embryo transfer in assisted reproductive technology. *Cochrane Database Syst Rev* 2016;**6**:CD002118.

Greco E, Minasi MG, Fiorentino F. Healthy babies after intrauterine transfer of mosaic aneuploid blastocysts. *N Engl J Med* 2015;**373**: 2089–2090.

Gui J, Xu W, Yang J, Feng L, Jia J. Impact of local endometrial injury on *in vitro* fertilization/intracytoplasmic sperm injection outcomes: a systematic review and meta-analysis. *J Obstet Gynaecol Res* 2019; **45**:57–68.

Guyatt G, Cairns J, Churchill D, Cook D, Haynes B, Hirsh J, Irvine J, Levine M, Levine M, Nishikawa J *et al.* Evidence-based medicine: a new approach to teaching the practice of medicine. *JAMA* 1992; **268**:2420–2425.

Hart RJ, Rombauts L, Norman RJ. Growth hormone in IVF cycles: any hope? *Curr Opin Obstet Gynecol* 2017;**29**:119–125.

Huang N, Situ B, Chen X, Liu J, Yan P, Kang X, Kong S, Huang M. Meta-analysis of estradiol for luteal phase support in *in vitro* fertilization/intracytoplasmic sperm injection. *Fertil Steril* 2015;**103**: 367–373.e5.

Humaidan P, Polyzos NP. (Meta)analyze this: systematic reviews might lose credibility. *Nat Med* 2012;**18**:1321–1321.

Jadad AR, Cook DJ, Browman GP. A guide to interpreting discordant systematic reviews. *CMAJ* 1997;**156**:1411–1416.

Jo J, Lee Y. Effectiveness of acupuncture in women with polycystic ovarian syndrome undergoing *in vitro* fertilisation or intracytoplasmic sperm injection: a systematic review and meta-analysis. *Acupunct Med* 2017;**35**:162–170.

Kamath MS, Chittawar PB, Kirubakaran R, Mascarenhas M. Use of granulocyte-colony stimulating factor in assisted reproductive technology: a systematic review and meta-analysis. *Eur J Obstet Gynecol Reprod Biol* 2017;**214**:16–24.

Lee E, Illingworth P, Wilton L, Chambers G. The clinical effectiveness of preimplantation genetic diagnosis for aneuploidy in all 24 chromosomes (PGD-A): systematic review. *Hum Reprod* 2015;**30**: 473–483.

Li J, Mo S, Chen Y. The effect of G-CSF on infertile women undergoing IVF treatment: a meta-analysis. *Syst Biol Reprod Med* 2017a;**63**: 239–247.

Li J, Yuan H, Chen Y, Wu H, Wu H, Li L. A meta-analysis of dehydroepiandrosterone supplementation among women with diminished ovarian reserve undergoing *in vitro* fertilization or intracytoplasmic sperm injection. *Int J Gynaecol Obstet* 2015;**131**: 240–245.

Li X-L, Wang L, Lv F, Huang X-M, Wang L-P, Pan Y, Zhang X-M. The influence of different growth hormone addition protocols to poor ovarian responders on clinical outcomes in controlled ovary stimulation cycles: a systematic review and meta-analysis. *Medicine (Baltimore)* 2017b;**96**:e6443.

Linde K, Willich SN. How objective are systematic reviews? Differences between reviews on complementary medicine. *J R Soc Med* 2003;**96**:17–22.

Martins WP, Nastri CO, Rienzi L, van der Poel SZ, Gracia C, Racowsky C. Blastocyst vs cleavage-stage embryo transfer: systematic review and meta-analysis of reproductive outcomes. *Ultrasound Obstet Gynecol* 2017;**49**:583–591.

Mendoza N, Perez L, Simoncini T, Genazzani A. Inositol supplementation in women with polycystic ovary syndrome undergoing intracytoplasmic sperm injection: a systematic review and meta-analysis of randomized controlled trials. *Reprod Biomed Online* 2017;**35**: 529–535.

Nagels HE, Rishworth JR, Siristatidis CS, Kroon B. Androgens (dehydroepiandrosterone or testosterone) for women undergoing assisted reproduction. *Cochrane Database Syst Rev* 2015;**11**:CD009749.

Page MJ, McKenzie JE, Kirkham J, Dwan K, Kramer S, Green S, Forbes A. Bias due to selective inclusion and reporting of outcomes and analyses in systematic reviews of randomised trials of healthcare interventions. In: Cochrane Methodology Review Group (ed). *Cochrane Database Syst Rev* 2014. http://doi.wiley.com/10.1002/14651858.MR000035.pub2 (30 May 2018, date last accessed).

Phillips B, Ball C, Badenoch D, Straus S, Haynes B. Dawes. Oxford Centre for Evidence-based Medicine—Levels of Evidence (March 2009). *CEBM* 2009. https://www.cebm.net/2009/06/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/ (29 May 2018, date last accessed).

Pinheiro LMA, Candido P da S, Moreto TC, Almeida WGD, Castro EC. Estradiol use in the luteal phase and its effects on pregnancy rates in IVF cycles with GnRH antagonist: a systematic review. *JBRA Assist Reprod* 2017;**21**:247–250.

Pribenszky C, Nilselid A-M, Montag M. Time-lapse culture with morphokinetic embryo selection improves pregnancy and live birth chances and reduces early pregnancy loss: a meta-analysis. *Reprod Biomed Online* 2017;**35**:511–520.

Qian Y, Meng Y. Therapeutic effect of acupuncture on the outcomes of in vitro fertilization: evaluated by an updated systematic review and metaanalysis. *Hum Reprod* 2016;**31**:i374.

Qin JC, Fan L, Qin AP. The effect of dehydroepiandrosterone (DHEA) supplementation on women with diminished ovarian reserve (DOR) in IVF cycle: evidence from a meta-analysis. *J Gynecol Obstet Hum Reprod* 2017;**46**:1–7.

Racowsky C, Kovacs P, Martins WP. A critical appraisal of time-lapse imaging for embryo selection: where are we and where do we need to go? *J Assist Reprod Genet* 2015;**32**:1025–1030.

Rawat S, Meena S. Publish or perish: where are we heading? *J Res Med Sci* 2014;**19**:87–89.

Ryan R, Synnot A, Hill S. Describing results. Cochrane Consumers and Communication Group. 2016. http://cccrg.cochrane.org/author-resources (1 May 2020, date last accessed).

Sar-Shalom Nahshon C, Sagi-Dain L, Wiener-Megnazi Z, Dirnfeld M. The impact of intentional endometrial injury on reproductive outcomes: a systematic review and meta-analysis. *Hum Reprod Update* 2019;**25**:95–113.

Shen C, Wu M, Shu D, Zhao X, Gao Y.The role of acupuncture in in vitro fertilization: a systematic review and meta-analysis. *Gynecol Obstet Invest* 2015;**79**:1–12.

Showell MG, Mackenzie-Proctor R, Jordan V, Hart RJ. Antioxidants for female subfertility. Cochrane Database Syst Rev 2017;**7**:CD007807. (

Shrier I, Boivin J-F, Platt RW, Steele RJ, Brophy JM, Carnevale F, Eisenberg MJ, Furlan A, Kakuma R, Macdonald ME *et al*. The interpretation of systematic reviews with meta-analyses: an objective or subjective process? *BMC Med Inform Decis Mak* 2008;**8**:19.

Simon C, Bellver J. Scratching beneath "The Scratching Case": systematic reviews and meta-analyses, the back door for evidence-based medicine. *Hum Reprod* 2014;**29**:1618–1621.

Tugwell P, Welch VA, Karunananthan S, Maxwell LJ, Akl EA, Avey MT, Bhutta ZA, Brouwers MC, Clark JP, Cook S *et al*. When to replicate systematic reviews of interventions: consensus checklist. *BMJ* 2020;**370**:m2864.

van Hoogenhuijze NE, Kasius JC, Broekmans FJM, Bosteels J, Torrance HL. Endometrial scratching prior to IVF; does it help and for whom? A systematic review and meta-analysis. *Hum Reprod Open* 2019;**2019**:hoy025.

Vitagliano A, Andrisani A, Alviggi C, Vitale SG, Valenti G, Sapia F, Favilli A, Martins WP, Raine-Ferring N, Polanski L *et al*. Endometrial scratching for infertile women undergoing a first embryo transfer: a systematic review and meta-analysis of published and unpublished data from randomized controlled trials. *Fertil Steril* 2019;**111**:734–746.e2.

Vitagliano A, Di Spiezio Sardo A, Saccone G, Valenti G, Sapia F, Kamath MS, Blaganje M, Andrisani A, Ambrosini G. Endometrial scratch injury for women with one or more previous failed embryo transfers: a systematic review and meta-analysis of randomized controlled trials. *Fertil Steril* 2018;**110**:687–702.e2.

Wilkinson J, Bhattacharya S, Duffy JMN, Kamath MS, Marjoribanks J, Repping S, Vail A, Wely M, Farquhar CM. Reproductive medicine: still more ART than science? *BJOG* 2019;**126**:138–141.

Xie Y, Zhang T, Tian Z, Zhang J, Wang W, Zhang H, Zeng Y, Ou J, Yang Y. Efficacy of intrauterine perfusion of granulocyte colony-stimulating factor (G-CSF) for infertile women with thin endometrium: a systematic review and meta-analysis. *Am J Reprod Immunol* 2017;**78**:e12701. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=ovfts&NEWS=N&AN=00001943-201708000-00003 (1 January 1AD, date last accessed).

Yu X, Ruan J, He L-P, Hu W, Xu Q, Tang J, Jiang J, Han J, Peng Y-F. Efficacy of growth hormone supplementation with gonadotrophins *in vitro* fertilization for poor ovarian responders: an updated meta-analysis. *Int J Clin Exp Med* 2015;**8**:4954–4967.

Zhang B, Cui L, Tang R, Ding L, Yan L, Chen Z-J. Reduced ectopic pregnancy rate on day 5 embryo transfer compared with day 3: a meta-analysis. *PLoS One* 2017;**12**:e0169837.

Zhang M, Niu W, Wang Y, Xu J, Bao X, Wang L, Du L, Sun Y. Dehydroepiandrosterone treatment in women with poor ovarian response undergoing IVF or ICSI: a systematic review and meta-analysis. *J Assist Reprod Genet* 2016;**33**:981–991.

Zhang X-M, Lv F, Wang P, Huang X-M, Liu K-F, Pan Y, Dong N-J, Ji Y-R, She H, Hu R. Estrogen supplementation to progesterone as luteal phase support in patients undergoing *in vitro* fertilization: systematic review and meta-analysis. *Medicine (Baltimore)* 2015;**94**:e459.

Zhao J, Xu B, Xie S, Zhang Q, Li YP. Whether G-CSF administration has beneficial effect on the outcome after assisted reproductive technology? A systematic review and meta-analysis. *Reprod Biol Endocrinol* 2016;**14**:62.

Zheng X, Lin D, Zhang Y, Lin Y, Song J, Li S, Sun Y. Inositol supplement improves clinical pregnancy rate in infertile women undergoing ovulation induction for ICSI or IVF-ET. *Medicine (Baltimore)* 2017;**96**:e8842.