Human Reproduction Update, Vol.22, No.3 pp. 306–319, 2016
Advanced Access publication on February 2, 2016 doi:10.1093/humupd/dmw001

human reproduction update

Trends in 'poor responder' research: lessons learned from RCTs in assisted conception

Athanasios Papathanasiou[1,*], Belinda J. Searle, Nicole M.A. King[2], and Siladitya Bhattacharya[2]

[1]Bourn-Hall Clinic, The Apex, Gateway 11, Farrier Close, Wymondham, Norwich NR18 0WF, UK [2]Aberdeen University, Foresterhill, Aberdeen, UK

*Correspondence address. E-mail: linktothanos@gmail.com

Submitted on September 29, 2015; resubmitted on January 5, 2016; accepted on January 11, 2016

TABLE OF CONTENTS

BACKGROUND: A substantial minority of women undergoing IVF will under-respond to controlled ovarian hyperstimulation. These women—so-called 'poor responders'—suffer persistently reduced success rates after IVF. Currently, no single intervention is unanimously accepted as beneficial in overcoming poor ovarian response (POR). This has been supported by the available research on POR, which consists mainly of randomized controlled trials (RCTs) with an inherent high-risk of bias. The aim of this review was to critically appraise the available experimental trials on POR and provide guidance towards more useful—less wasteful—future research.

METHODS: A comprehensive review was undertaken of RCTs on 'poor responders' published in the last 15 years. Data on various methodological traits as well as important clinical characteristics were extracted from the included studies and summarized, with a view to identifying deficiencies from which lessons can be learned. Based on this analysis, recommendations were provided for further research in this field of assisted conception.

RESULTS: We selected and analysed 75 RCTs. A valid, 'low-risk' randomization method was reported in three out of four RCTs. An improving trend in reporting concealment of patient allocation was also evident over the 15-year period. In contrast, <1 in 10 RCTs 'blinded' patients and <1 in 5 RCTs 'blinded' staff to the proposed intervention. Only 1 in 10 RCTs 'blinded' ultrasound practitioners to patient allocation, when assessing the outcome of early pregnancy. The majority of trials reported an intention-to-treat analysis for at least one of their outcomes, with an improving trend in the recent years. Substantial variation was noted in the definitions used for 'poor responders', the most popular being 'low ovarian response at previous stimulation'. The preferred cut-off value for defining previous low response has been 'less or equal to three retrieved oocytes'. The most popular tests used for diagnosing diminished ovarian reserve have been antral follicle count and FSH. Although the Bologna criteria for POR were only recently introduced, they are expected to become a popular definition in future 'poor responder' trials. Numerous interventions have been studied on 'poor responders'. Most of these have been applied before/during controlled ovarian hyperstimulation. The antagonist protocol, the microdose flare protocol and the long down-regulation protocol have been among the most popular interventions. The analysis of outcomes revealed a clear improving trend in reporting live birth. In contrast, only 10% of RCTs reported significant improvement in reproductive outcomes among tested interventions. Twelve 'significant' interventions were reported, each supported by a single 'positive' RCT. Finally, trials of higher methodological quality were more likely to have been published in a high-impact journal.

CONCLUSIONS: Overall, the majority of published trials on POR suffer from methodological flaws and are, thus, regarded as being high-risk for bias. The same trials have used a variety of definitions for their poor responders and a variety of interventions for their head-to-head comparisons. Not surprisingly, discrepancies are also evident in the findings of trials comparing similar interventions. Based on the identified deficiencies, this novel type of 'methodology and clinical' review has introduced custom recommendations on how to improve future experimental research in the 'poor responder' population.

Key words: poor responder / IVF / RCT / methodology / review / ovarian stimulation / low responder / low response / blinding / live birth

Introduction

Since the conception of IVF, it has become apparent that a proportion of women respond suboptimally to controlled ovarian hyperstimulation with exogenous gonadotrophins (Tanbo *et al.*, 1990). These women, so-called 'poor responders', may suffer persistently low success rates with IVF (Oudendijk *et al.*, 2012; Busnelli *et al.*, 2015). Despite being a popular theme for research, optimal management of poor ovarian response (POR) remains an unsolved enigma, with no single intervention being widely accepted as beneficial (Pandian *et al.*, 2010).

The aim of this review was to critically appraise the available experimental research on this topic, with a view to identifying deficiencies from which lessons can be learned, thus improving future research.

Literature search

A comprehensive search of MEDLINE and the Cochrane Central Register of Controlled Trials (CENTRAL) was undertaken for randomized controlled trials (RCTs) between January 2000 and December 2014, using the following keywords: (poor adj5 respon$ OR low adj5 respon$ OR suboptimal adj5 respon$ OR inadequate adj5 respon$) AND (in-vitro fertili?ation OR IVF OR intracytoplasmic sperm injection OR ICSI OR assisted conception). Only full RCT publications on poor responders were included, from which we were able to collect detailed clinical and methodological data. The references of the included studies were also hand-searched for relevant RCTs potentially missed by the electronic search. In addition, we hand-searched the references of any systematic reviews on 'poor responders' that were published in the last 10 years. One of the authors (A.P.) has his own repository of 'poor responder' trials, having conducted similar searches towards other publications; this was also checked for relevant trials. In view of the large number of retrieved articles through this combined approach (687 unique articles), two reviewers (A.P. and B.J.S.) independently screened these articles to ensure that no relevant studies were missed. Finally, a total of 75 unique RCTs were included in this review (Supplementary data, Table SI).

Methods

This review was conducted on two levels: methodological and clinical.

Methodologically, we evaluated a series of standards that are known to affect the validity of RCTs.

In detail, we assessed:

(i) If individual studies reported the preferred method of randomization or allocation concealment.
(ii) If 'blinding' of patients or personnel was performed and how this was achieved.
(iii) If the assessors of the outcome were also 'blinded' and how this was achieved.
(iv) If a CONSORT flowchart was included.
(v) If the outcomes were reported on an intention-to-treat (ITT) basis. An ITT analysis typically is performed 'per randomized woman'.
(vi) Which reproductive outcomes were reported. We focused on live birth, as it is the most critical reproductive outcome, and miscarriage as the main adverse outcome in fertility.
(vii) Selective reporting of outcomes. A common issue with selective reporting is when only significant outcomes are reported but non-significant outcomes from the same study are omitted. Therefore, we looked at the protocols of studies with statistically significant pregnancy outcomes, aiming to ascertain if any non-significant outcomes were potentially omitted by the final publication.
(viii) How many studies stated a primary outcome or performed sample size calculations. We also collected data on the actual sample size of the included RCTs.

We also aimed to evaluate a series of important clinical characteristics, such as:

(i) 'Poor responder' definitions used in the literature. We expected that substantial variations exist, which may hinder interpretation of results across studies, thus contributing to clinical heterogeneity.

(ii) Interventions and controls (comparators). We expected that combined interventions or various choices of controls were used in POR research, further contributing to heterogeneity.
(iii) Trials with significant results for their reproductive outcomes. These studies may highlight interventions that warrant further research.

In order to identify time trends, we stratified the aforementioned data according to 5-year intervals (2000–2004, 2005–2009 and 2010–2014).

Methodological trends

Randomization and allocation concealment

Seventy-seven per cent (58/75) of trials reported their preferred randomization technique. Of these, 54 RCTs (72% of trials) reported a valid 'low-risk' randomization technique, with 'computer-generated random number sequence' being the preferred method. Reporting the randomization method appears to have been a steady research priority over the last 15 years, being more frequently reported than any other methodological standard (Fig. 1). An upward trend in reporting concealment of patient allocation is also evident during the period concerned (27% of RCTs in 2000–2004, 44% in 2005–2009, 56% in 2010–2014) (Fig. 1). The preferred method for concealing allocation has been the use of sealed envelopes. Twenty-one per cent of trials (16/75) reported a low-risk concealment technique, the most popular being 'sealed, opaque envelopes'. Nevertheless, approximately one in four RCTs still failed to report an adequate method of randomization and four in five RCTs did not perform or report a low-risk method or concealment of allocation.

When planning randomization, one should consider three important questions: who/what will be randomized, how this will take place and when. It is sensible to randomize the unit receiving the intervention; in assisted conception trials, this could be the patient, the cycle, the oocyte or the embryo. The choice of unit, however, may have substantial implications for the reporting of certain outcomes. For example, when randomizing cycles, it may not be straightforward to report patient-based outcomes, unless each patient has undergone a single cycle only. Otherwise, adjustments need to be made for the clustering effect (multiple cycles in the same patient are correlated). It is fortunate that the majority of proposed interventions in POR trials have been applied on patients who undergo a single IVF cycle (93%), as this facilitates the reporting of patient-friendly outcomes.

When considering the choice of randomization technique, random number sequences are a prerequisite for genuine randomization. In the context of RCT planning, pseudorandom sequences—generated through a computer or a list—are ideal, as the number sequence can be back-traced, if required (Machin *et al.*, 2007). Although it does not guarantee similarity in the baseline characteristics between groups, genuine randomization ensures that each participant has an equal chance of being allocated to any one of the study groups, thereby eliminating selection bias. Allocation based on the day of week, date of birth, etc. is not truly randomized, as the researchers are able to guess the next patient allocation. While there is not one best time to randomize patients, it is desirable that randomization takes place close to the onset of the intervention, as this will minimize the number of early drop-outs.

Allocation concealment guarantees that each participant receives the intervention allocated by the randomization sequence. Strictly speaking, sealed envelopes—the most popular method in POR trials—may still be prone to manipulation. However, sealed and opaque envelopes can effectively conceal allocation, even if one was to attempt reading the contents of the envelope in front of a strong light source. Sequentially numbered envelopes are also effective in concealing allocation, as long as the envelope order is maintained. The use of internet randomization is an alternative option that can achieve randomization and allocation concealment by use of a single online system.

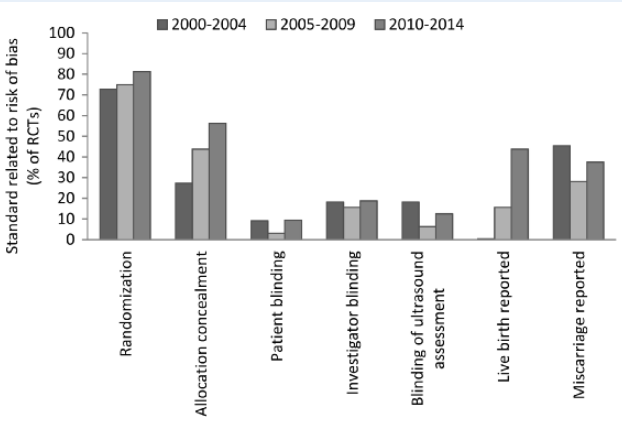

Figure 1 Frequency of reported standards that are relevant to the risk of bias (presented in three 5-year intervals to display trends).

In the absence of these standards, bias may be introduced when allocating patients to study groups (selection bias) which could, in turn, affect the size of the measured effect of the treatment (Schulz *et al.*, 1995). RCTs that have not secured concealment of allocation may report up to 40% exaggerated estimates of the effect, compared with trials that have (Schulz *et al.*, 1995). This highlights the importance of addressing these two methodological principles early in the design phase of POR RCTs.

'Blinding' of patients and personnel

Only 7% of studies (5/75) reported 'blinding' patients to the proposed intervention. Patient 'blinding' was achieved by use of a 'placebo' preparation in all studies except for one (in this single RCT, follicular flushing was performed at the time of oocyte recovery without the patients being aware). Overall, investigators were more frequently 'blinded' than patients, without any appreciable changes during the 15-year period (18% in 2000–2004, 16% in 2005–2009, 19% in 2010–2014) (Fig. 1). Even so, a substantial majority of RCTs (around 80%) did not report or apply any 'blinding' technique for the staff who managed the patients' IVF cycles. The ultrasound practitioners who assessed the outcome (clinical/ongoing pregnancy, miscarriage) were adequately blinded in ~11% (8/75) of studies (Fig. 1). In half of these, this was achieved through patient 'placebo' administration.

The ideal scenario would be that both patients and staff are 'blinded' to the received intervention. Otherwise, patient expectations could be raised if they have been allocated to the 'intervention' or the 'novel' group or, inversely, lowered if they have been allocated to the 'control' or 'conventional' group (Machin *et al.*, 2007). More importantly without blinding, the staff managing patients' cycles could be prone to decision bias with respect to aspects of the treatment other than the intervention under study, e.g. when deciding how to manage the stimulation phase (changes of FSH dose and duration of stimulation). Although the

accurate diagnosis of early pregnancy is greatly facilitated with the use of modern, high-resolution ultrasound devices, it is still desirable to 'blind' the ultrasound practitioners at the time of pregnancy scan, as this will minimize the risk of detection bias (Wood *et al.*, 2008).

It is worth discussing this infrequent use of 'blinding'. It is evident that effective 'blinding' techniques could have been applied more frequently in POR research. In our series, 17 RCTs investigating single add-on interventions that could have used a placebo for their control groups, opted not to do so. For trials comparing various interventions, such as stimulation protocols, introduction of placebos is inevitably a more complex but still not an impossible task. Since IVF stimulation protocols typically involve the administration of two types of drugs—a gonadotrophin and a GnRH analogue—most protocol comparisons could potentially be masked by use of a 'double-dummy' approach (Marusic and Ferencic, 2013).

Even in the absence of placebo, it is still possible to persevere with participant or investigator 'blinding'. When recruiting participants, it is important to describe the received interventions in a factual way, but also refrain from giving out detailed information on study design or revealing if they belong to the experimental or control group (Page and Persch, 2013). It is possible to have a third person deliver certain interventions, such as drug administration, with attention to concealing the name and dose of these drugs from the patients; obviously, one should account for the extra burden on staff resources before applying this modality. Participants can also be instructed not to reveal their allocated group to any of the staff who are involved in their care. Moreover, staff involved in routine IVF care should remain unaware of basic trial characteristics, such as inclusion criteria, hypotheses and outcome measures. Likewise, the assessors of outcomes are more likely to remain 'blinded' if they are unaware of the particulars of the trial and have restricted access to the collected data (Minns Lowe *et al.*, 2011). They should not be the same staff that provided care during delivery of the intervention, as this will minimize the opportunities of finding out about individual allocations. Formal training of healthcare staff on how to facilitate the proper conduction of the trial should be offered, as it may also minimize the chance of breaking 'blinding' (Johnson and Remien, 2003). Obviously, transparent reporting of all the steps undertaken to secure 'blinding' is recommended. Of note is that this has not been adequately addressed by POR trials so far.

Reporting of live birth, miscarriage and surrogate markers

The earlier 'poor responder' RCTs did not report live birth, which is in line with the low reporting frequency of the same outcome in subfertility trials of the same era (Dias *et al.*, 2006). This substantially improved in the following years (32% in 2005–2009, up to 44% in 2010–2014), reflecting the gradually increasing awareness of the importance of live birth as a reproductive outcome (Fig. 1). During the last 5 years, live birth has been the most popular critical outcome reported in POR RCTs, superseding its old-time rival, clinical pregnancy. Miscarriage rates were rather consistently reported only in 35% of RCTs during the study period (Fig. 1). Most studies reported miscarriage per clinical pregnancy. However, the precise definition of miscarriage was usually not clear.

There are arguments supporting the shift towards reporting live birth in assisted conception studies. Live birth is the end-point of the fertility journey and the desired outcome for couples undergoing treatment (the consumers) (Chetkowski, 2014). While women welcome a positive pregnancy result, they ultimately judge success by whether or not they have had a baby through the treatment. Likewise, state funders for IVF treatment also consider live birth as the end-point, typically funding couples up to the point of having their first baby (Lindstrom and Waldau, 2008). In the case of poor responders, the older subgroup may be more prone to pregnancy loss (Slovis and Check, 2013). As a result, reporting only early pregnancy outcomes may not be adequate to capture all miscarriage cases. However, these cases would be accounted for, if live birth was invariably reported. There are further methodological advantages stemming from the reporting of live birth. As an outcome, live birth is an objective and indisputable outcome and, therefore, is not liable to detection bias. For the same reason, it may also be less prone to bias from inadequate 'blinding'. In contrast, early pregnancy outcomes that are diagnosed by ultrasound are exposed to an element of subjectivity depending on the practitioner performing the assessment. Various definitions for clinical and ongoing pregnancy have been used in previous research, which, in contrast to live birth, give rise to challenges in outcome interpretation and meta-analytic attempts.

Despite the improving trend in reporting reproductive outcomes, surrogate markers, such as the number of retrieved oocytes, still remain popular primary outcomes (43% of studies, 32/75). There are several reasons why this happens. From a methodological viewpoint, demonstrating clinical improvement in oocyte numbers rather than in pregnancy rates requires a smaller sample size, translating into RCTs which are more likely to come to completion (Machin *et al.*, 2007). On the clinical side, large observational series have correlated lower numbers of retrieved oocytes with low live birth rates, leading to the erroneous assumption that increasing the oocyte yield will improve reproductive outcomes in the poor responder group (Steward *et al.*, 2014). Many clinicians tend to favour proposed interventions that could improve oocyte yield, even in the absence of higher pregnancy rates. A typical example has been to opt for the antagonist protocol as the default protocol for poor responders, based on the meta-analysis of a limited number of RCTs that suggests higher oocyte numbers with this protocol (Pandian *et al.*, 2010; Patrizio *et al.*, 2015). Within our series, most intervention trials on POR (>80%) that have indeed demonstrated increased oocyte yield, could not prove a concurrent increase in pregnancy rates. Another false assumption is that the potential benefit of an intervention during the stimulation stage is always delivered via improvement in oocyte yield. As confirmed in our series, four trials of interventions that were implemented during the stimulation phase led to improved reproductive outcomes without associated improvement in the oocyte yield (antagonist versus microdose protocol, added LH versus increased FSH dose, 300 IU FSH dose versus 150 IU FSH dose, Day 4 FSH start versus Day 1 FSH start). This implies that POR research should not just focus on surrogate markers, but invest more on reproductive outcomes, such as live birth.

CONSORT flowchart

A common challenge in the estimation of the size of the effect of a given intervention is accounting for cases with unknown outcomes or withdrawals. Such cases, if unaccounted, may contribute to outcome bias (attrition bias) (Tierney and Stewart, 2005).

The CONSORT statement has been a major contributor towards improving the reporting of RCTs, since being widely publicized in 2001

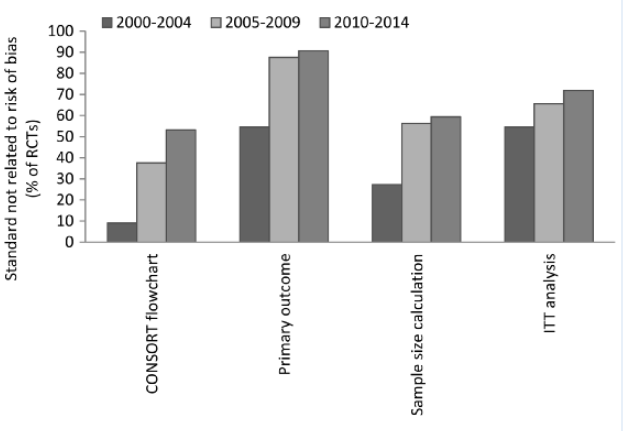

Figure 2 Frequency of reported standards not related to the risk of bias (presented in three 5-year intervals to display trends). ITT, intention-to-treat.

(Moher *et al.*, 2001). An integral part of this statement is the CONSORT flowchart, which details the patients' pathway during the conduction of a trial. An increasing number of publications on poor responders included a CONSORT flowchart (9% in 2000–2004, 38% in 2005–2009, 53% in 2010–2014) (Fig. 2). This improving trend likely follows the decision taken by selected reproductive medicine journals to ask for a CONSORT checklist to accompany every submitted RCT manuscript. This is, however, not mandatory for the majority of journals that have been publishing assisted conception trials.

ITT analysis

In the studied period, an increasing number of trials reported an ITT analysis for at least one of their outcomes (55% in 2000–2004, 66% in 2005–2009, 72% in 2010–2014) (Fig. 2). It is worth mentioning that the majority of trials in this review reported outcomes 'per cycle' rather than 'per randomized woman'. However, they were still considered as having performed an ITT analysis, after confirming that the two denominators were essentially the same (all randomized women in these studies also received a single cycle of the intervention).

POR research consists mostly of 'superiority' trials, which aim to assess if a novel intervention is better than conventional care. For these trials, an ITT analysis is considered the least biased way of estimating the true effect of the intervention under study. Some aspects of ITT analysis may appear counter-intuitive to clinicians. For example, cases randomized to one intervention but ending up receiving another would still be included in their originally assigned group according to ITT analysis. Clinically, it would make sense to count these cases under the arm of the intervention actually received; however, this would result in 'breaking' the randomization order. In the case of withdrawals, exclusion from analysis would introduce attrition bias; however, this can be accounted for by performing an ITT analysis.

Research in assisted conception is well suited for promoting ITT analysis. In the majority of IVF studies—including POR trials—patients tend to receive the interventions assigned to them by randomization. For single cycle trials, a 'per protocol' or 'per started cycle' analysis would then be identical to an ITT analysis. Moreover, it is not common for women undergoing IVF to withdraw or go missing, particularly when the study is completed within a short time frame, e.g. when studying single cycles. Even for later occurring outcomes, such as live birth, most IVF units have a robust follow-up system to ensure proper outcome documentation.

Few trials in POR research are 'equivalence' trials, where a novel intervention is assessed to determine whether it is as efficacious as standard care. A typical example involves trials utilizing milder forms of ovarian stimulation on 'poor responders'. The rationale behind investigating milder stimulation is 2-fold; it may be as efficacious as conventional stimulation in recruiting the limited number of available follicles, and, if so, it may be cost-effective in view of the low doses of administered drugs (Ragni *et al.*, 2012). Such trials tend to be more 'pragmatic' than superiority trials, with higher rates of withdrawals, since even women may find it counter-intuitive to address the problem of low oocyte yield with milder stimulation (Morgia *et al.*, 2004). In 'equivalence' trials, an ITT analysis is less likely to detect a difference, inflating the likelihood of equivalence. A 'per protocol' analysis may be more appropriate in this case.

Other denominators, such as 'per egg collection' or 'per embryo transfer' are popular in POR literature. Indicators of performance, e.g. live birth per embryo transfer, implantation rates, number of supernumerary embryos, etc., can still be useful as indirect measures of oocyte quality and may help to further describe the underlying mechanisms of given interventions. However, such analyses may be biased, not only because they have 'broken' the randomization order, but also because they often produce inflated measures. This is caused by the rather high proportion of poor responders that drop-out before egg collection, resulting in smaller denominators within the analysis. Consequently, they should be reported as secondary analyses.

Selective reporting of outcomes

Twelve RCTs reported statistically significant reproductive outcomes. Of these, only three studies had published 'a priori' study protocols (Lainas *et al.*, 2008; Baerwald, 2012; Mok-Lin, 2013). In these studies, there was complete agreement between the stated outcomes of the protocol and the full study.

Selective outcome reporting is one the most important sources of bias in RCTs. A common occurrence is the preferential reporting of only positive, statistically significant outcomes. This raises various concerns; firstly, it may misleadingly overestimate the effectiveness of a given intervention. In the absence of reporting adverse outcomes, e.g. miscarriage, it is not possible to make a sound judgment of benefits over risks. In the absence of reporting non-significant outcomes, e.g. live birth, a proposed intervention may appear more beneficial than it actually is. Selective reporting from individual trials can also introduce bias to subsequent meta-analyses, where, by combining primarily statistically significant results, an exaggerated effect is likely to be reported for the outcome of interest (Page *et al.*, 2014).

A most effective way to minimize the risk of reporting selected outcomes is by registering the study after the design has been agreed but before recruitment commences. Registration involves publishing a research protocol on the internet (www.clinicaltrials.gov) or an academic journal, which, among others, also details the planned outcomes for the study of interest. Trial registration promotes transparency, by allowing detection of discrepancies between the protocol and final publication. Our own limited cohort clearly supports this statement. For this reason, major reproductive medicine journals require that a trial has been registered before being considered for publication.

Primary outcome, sample size calculation and actual sample size

During the 15-year period, 84% (63/75) of 'poor responder' trials clearly stated their primary outcome, the most popular being the 'number of retrieved oocytes' (Fig. 2). Only 53% (40/75) reported sample size calculations. Twenty-one studies performed sample size calculations based on the number of oocytes, 10 on reproductive outcomes (pregnancy or live birth) and six on other primary outcomes, while three RCTs did not clearly report which outcome they used. The actual sample size ranged widely (7–355 participants per group). In the 5-year period between 2000 and 2004, 91% of trials were small (<50 patients per group), with no trials having more than 100 patients per group. From 2010 onwards, 62% of studies were still small, with almost one in five RCTs having more than 100 patients per group.

The choice of primary outcome has substantial methodological implications for RCTs. The potential benefit of a given intervention is truly judged by its effect on the primary outcome. Moreover, defining a primary outcome is critical for 'a priori' sample size calculations (Zlowodzki and Bhandari, 2009). Methodologically speaking, performing sample size calculations based on oocyte numbers will lead to lower numbers of participants required for the trial, when compared with calculations based on reproductive outcomes. Some authors have even consciously decided to use oocyte numbers as their primary outcome for sample size calculations, as the only way of achieving adequate numbers for recruitment within the setting of their practice (Weissman *et al.*, 2003). However, this should not justify the decision to use a surrogate marker as a primary outcome. Sample size calculations are paramount for the design of a high-quality RCT, for they minimize the incidence of a Type II error (the probability of the study not showing a statistical difference in outcomes although such a difference truly exists) (Maggard *et al.*, 2003). With more than 90% of POR RCTs showing no statistical differences in their reported outcomes and almost 50% of RCTs not having performed sample size calculations, one expects that a substantial proportion of these studies will be exposed to a Type II error. Such research is not as useful, since it lacks the statistical power to infer equivalence between compared interventions (Dimick *et al.*, 2001).

How large should a RCT be in order to detect clinically important differences in reproductive outcomes in the field of POR? Although reported pregnancy rates do vary, our observations (from the control groups) suggest that poor responders achieve an average clinical pregnancy rate of 15.9% with IVF treatment [95% confidence interval (CI) 13.8–17.9%, 63 RCTs, $I^2 = 61\%$, binary random effects]. For an anticipated doubling in success with a given intervention (relative improvement 100%), a trial with 90% power and Type I error of 0.05 would require at least 149 participants per study group (http://clincalc.com). Only four of our RCTs used high enough numbers of participants in line with the above calculations, two originating from the same research group (Lainas *et al.*, 2008; Revelli *et al.*, 2012, 2014; Prapas *et al.*, 2013). This clearly implies that it is challenging to deliver large-scale RCTs on poor responders. An obvious reason for the difficulty is that poor responders only represent a minority of IVF patients. Eligibility for inclusion is also influenced by the 'strictness' of the POR criteria used by individual studies. One should also take into account that poor responders with a history of repeated IVF failures may be less willing to accept randomization or blinding, particularly if they are funding their own treatment (Hemminki *et al.*, 2004; Page and Persch, 2013).

Table I Key methodological facts and trends in 'poor responder' research in assisted conception.

- The most frequently reported 'risk of bias' standard in RCTs is the preferred randomization method
- A valid, 'low-risk' randomization method was reported in three out of four RCTs
- Recent RCTs are more likely to have reported concealment of patient allocation
- Less than 1 in 10 RCTs 'blinded' patients and <1 in 5 RCTs 'blinded' staff to the proposed intervention
- One in 10 RCTs 'blinded' ultrasound practitioners to patient allocation when assessing the outcome of early pregnancy
- Recently published RCTs are more likely to include a CONSORT flowchart
- The majority of RCTs reported an ITT analysis for at least one of their outcomes, with an improving trend in the recent years
- A clear trend of improvement in reporting live birth is evident in the more recent studies
- Only one in four RCTs with statistically significant results had previously submitted a research protocol
- The most commonly reported primary outcome is the number of retrieved oocytes
- Approximately half of RCTs performed a sample size calculation
- A trend towards larger RCTs is evident in the more recent years
- RCTs with more high-quality methodological features were more likely to have been published in a high-impact journal

Publishing in high-impact journals

In summary, 23 RCTs were published in the *Fertility and Sterility* journal, followed by 14 in *Human Reproduction*, 6 in *Assisted Reproduction and Genetics*, 5 in *Gynecological Endocrinology*, 4 in *Reproductive Biomedicine Online* and 4 in *Archives of Gynecology and Obstetrics*. The remaining 19 RCTs were published in various journals specializing in gynaecology or reproductive medicine.

We assessed whether trials with a higher number of favourable methodological traits were more likely to have been published in high-impact journals (impact factor of 2 or more). The following 10 'positive' traits were accounted for: reporting of live birth, ITT analysis, sample size calculation, low-risk randomization technique, allocation concealment, blinding of patients, blinding of staff, blinding of outcome assessors, CONSORT flowchart, sample size more than 100 participants per group. Trials with five or more positive traits were approximately three times more likely to have been published in high-impact journals (odds ratio 3.45, 95% CI 1.11–10.75).

The key methodological facts and trends in POR research are summarized in Table I.

Clinical trends

Heterogeneity of the POR criteria

A variety of criteria have been used to characterize poor responders in the trials in our series. These are broadly categorized in five groups (Fig. 3):

(i) women with previous low ovarian response to IVF ovulation induction ('proven' poor responders),
(ii) women expected to exhibit low ovarian response to planned IVF ovulation induction ('expected' poor responders),

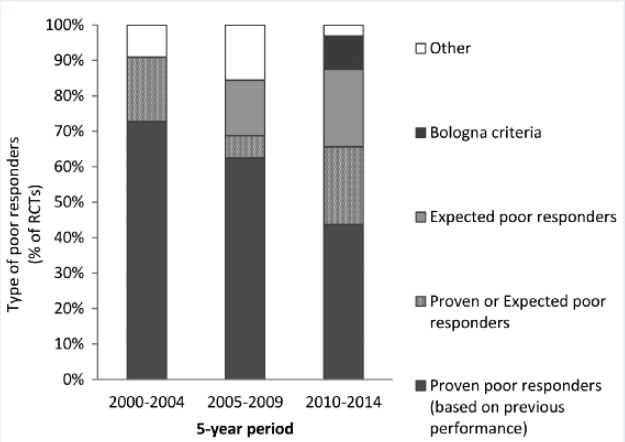

Figure 3 Frequency of various 'poor responder' definitions appearing in randomized trials (presented in three 5-year intervals to display trends).

(iii) either 'proven' or 'expected' poor responders,
(iv) women fulfilling the recently published Bologna criteria (two out of three parameters: advanced age/risk factors, abnormal ovarian reserve tests, less or equal to three retrieved oocytes at previous stimulation or a suboptimal response with maximal stimulation on at least two occasions) and
(v) other definitions.

Previous low response to ovarian hyperstimulation has been the most popular definition, utilized by 56% of RCTs (Fig. 3). Even within this definition though, substantial heterogeneity is present, from the various standards used to define previous POR (number of retrieved oocytes, number of recruited follicles, estradiol levels on the day of HCG or combinations of the above) to the various cut-offs used for each standard. The majority of trials in this review used '$\leq$3 retrieved oocytes' (29%) or '$\leq$4 retrieved oocytes' (25%) in order to define 'proven' POR.

'Expected' poor responders are also at risk of POR, based on evidence of their diminished ovarian reserve. Substantial heterogeneity exists here as well, through the use of various reserve tests with different cut-offs. While FSH was the dominant criterion in earlier research, it has been recently superseded by antral follicle count (AFC). The most popular antral follicle cut-off has been AFC $<$ 5 (33%) and AFC $<$ 6 (27%). Anti-Mullerian hormone (AMH), a rather recent addition to the family of ovarian reserve tests, has not been as popular (only two RCTS used AMH in 2010–2014), which could be attributed to limitations in the availability or standardization of the older AMH assays in particular (Iliodromiti *et al.*, 2014a, b; Nelson *et al.*, 2015).

Using a single criterion to define POR may not be satisfactory, as no single test is able to effectively discriminate between women who have a reasonable and a low chance of success with IVF treatment. Indeed, more than half of women with an initial low ovarian response will have a normal response at a subsequent cycle, suggesting that a substantial proportion of women may be 'understimulated' during their first IVF cycle (Veleva *et al.*, 2005). This implies that the intensity of stimulation is also relevant for defining POR, being more likely when few oocytes are retrieved after consumption of high gonadotrophin doses. Even AFC and AMH, which enjoy a high discriminatory ability between low and normal ovarian response, are of limited value in predicting women who will suffer reduced success after IVF (Iliodromiti *et al.*, 2014a, b; Hamdine *et al.*, 2015). The Bologna criteria were introduced to minimize such heterogeneity, by requiring the combined presence of at least two adverse factors for qualifying (Ferraretti *et al.*, 2011). Advanced female age has been incorporated into these criteria, as it was considered an independent risk factor for POR. This is supported by previous research, where younger 'poor responders' have been shown to perform better than older ones, indicating that the woman's age may be related to oocyte quality, in the same way that ovarian reserve correlates with oocyte availability (Oudendijk *et al.*, 2012). Early observational data indicate that different combinations of the Bologna criteria are associated with similar reproductive outcomes, supporting the hypothesis that these criteria are indeed doing what they were designed for—providing a more homogeneous POR population for research purposes (La Marca *et al.*, 2015).

Although only three RCTs have been published according to the Bologna criteria by the end of 2014, it is expected that many more will follow. At the time of writing this review, out of the 33 registered, open RCTs on POR, 12 RCTs (36%) have used the Bologna criteria and a further five RCTs (15%) have used similar to the Bologna criteria (www.clinicaltrials.gov). This implies that these criteria are already gaining ground in matters of acceptance by researchers. One would expect that they will eventually become the default inclusion criteria in future research. If so, researchers would have to justify any decision not to utilize these criteria for their upcoming RCTs.

It must be stressed though that the Bologna criteria may have not yet provided all the answers. Concerns have been raised that a degree of heterogeneity remains, which could impact on the methodological quality of future studies (Papathanasiou, 2014). Although the criteria acknowledge the importance of 'risk factors' towards the definition of POR, the prognostic impact of individual factors is still unclear, particularly for the young poor responder group (Younis, 2012). Environmental (after surgery, irradiation, etc.) or genetic causes (such as carrying an unfavourable single mononucleotide polymorphism) may contribute variably to POR in the young patient, even though they are both included under the same criteria (La Marca *et al.*, 2013). In addition, gamete quality issues are not addressed within the Bologna definition and could, in theory, affect the prognosis of a young poor responder and the decision to further persevere with treatment or not. Future research should aim to clarify these uncertainties and, if appropriate, lead to revision of the current criteria.

Uncertainty also exists how to meta-analyse studies using the Bologna criteria, in the light of older studies that have used other definitions. Methodologically, it may be unwise to include all studies in the same meta-analysis, in view of problems with heterogeneity. Running separate analyses is a safer approach, but it may take time for enough suitable trials to become available, before meaningful conclusions are reached.

There is emerging evidence of a group of 'suboptimal' responders, whose ovaries do not underperform enough to qualify as Bologna 'poor responders', but may still suffer from suboptimal reproductive outcomes (Polyzos and Sunkara, 2015). The notion of an intermediate 'poor responder' group is in agreement with our current understanding of ovarian physiology, indicating that the decline in ovarian reserve is a gradual, non-reversible process, not an all-or-nothing phenomenon (te Velde and Pearson, 2002; Broekmans *et al.*, 2009).

Although the hypothesis of 'suboptimal' responders remains to be explored, researchers may wish to take this new trend into account when designing their POR populations. There are three reasons for doing so; firstly, women with a mediocre but not low ovarian performance are plenty in everyday practice of IVF; secondly, only by including an intermediate group in future research, we will ever gain insight into its significance and potential. More importantly in the context of RCTs, it may well be the case that a 'suboptimal' responder group will benefit more from proposed interventions, simply because it represents the mild end of the spectrum of POR. In contrast, it is possible that future POR RCTs will confirm that the Bologna group, being associated with particularly low prognosis, is more resistant to treatment than any other POR group. How such an intermediate group should be defined for the purposes of future research remains to be confirmed, although the authors of the original publication proposed 'four to nine retrieved oocytes' as their preferred cut-off (Polyzos and Sunkara, 2015). This definition may not be specific enough, particularly when considering the recent trend towards more individualized stimulation regimens. As in the case of the Bologna criteria for poor responders, several markers alluding to suboptimal performance (tests suggesting diminishing ovarian reserve, moderate oocyte numbers in the presence of intense stimulation) may need to be combined, in order to better characterize this novel group of suboptimal responders. For future research, we would recommend that any intermediate group should be an add-on, not a replacement of the Bologna criteria.

In search of a universal comparator

The choice of the control group (comparator) is an important consideration for researchers. In pragmatic trials, the comparator typically comprises what is considered 'standard care' in the field of interest. There are arguments in support of using 'standard care' to ascertain the merits of novel interventions. By doing so, the trial findings can be more confidently generalized to the everyday clinical setting (where standard care is delivered). It also allows the interpretation of findings across different RCTs. Finally, in conjunction with more uniform population criteria, e.g. the Bologna criteria, a common comparator may promote more meaningful meta-analyses.

Often the choice of a control is straightforward; either no additional intervention (standard care only) or placebo use. However, in the field of assisted conception and in POR research in particular, it is common that various interventions, such as stimulation protocols, are compared against each other. Within our series, the most popular comparison has been between the antagonist protocol and the microdose flare protocol (10 RCTs in total), followed by the comparison between the antagonist protocol and the long down-regulation (long) protocol (8 RCTs). Expectedly, protocols other than the long protocol appear at least as prominently in POR trials (Fig. 4); this may be explained by the fact that, to date, there is no agreed best protocol to represent 'standard care' in POR research. Appraisal of the research evidence indicates that no single stimulation protocol is superior to another for poor responders. Our series supports this statement; of the eight RCTs comparing the long with the antagonist protocol, the results were inconclusive, with only one trial reporting improved ongoing pregnancy rates with the long protocol (Prapas *et al.*, 2013). The long protocol was compared with the microdose flare in a single RCT, without any differences in oocyte numbers or reproductive outcomes (Chatillon-Boissier *et al.*, 2012). A Cochrane

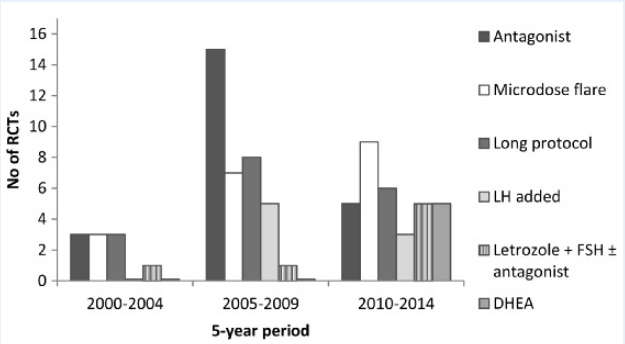

Figure 4 Most popular interventions in 'poor responder' randomized trials (presented in three 5-year intervals to display trends). DHEA, dehydroepiandrosterone.

review suggested that the antagonist protocol may lead to a higher oocyte yield compared with the long protocol (Pandian *et al.*, 2010). However, the findings were based on low-quality evidence which is likely to change with future research. Even so, neither protocol is superior in terms of the more important pregnancy outcomes. A more recent meta-analysis did not detect any differences in oocyte yields or pregnancy outcomes between agonist and antagonist protocols (Pu *et al.*, 2011).

Clinical practice also varies, with the majority of clinicians (67%) not opting for a down-regulation protocol for their poor responders (Tur-Kaspa and Fauser). This is in contrast to the unselected IVF population, where the long down-regulation protocol is still considered the first-line treatment option worldwide (Tur-Kaspa and Fauser). It is not clear why certain protocols have been so popular for poor responders. They could have become physician favourites, based on their favourable pharmacokinetics (avoiding down-regulation, flare effect) or the available low-quality evidence of improvement in surrogate markers (Pandian *et al.*, 2010). For the antagonist stimulation protocol in particular, experience is rapidly growing worldwide. Consequently, it is currently acknowledged to be as efficacious as the long protocol, and a recognized modality in conventional stimulation (Nargund *et al.*, 2007).

When considering the optimal strength of the stimulation regimen, the Bologna criteria state that at least 150 IU of FSH should qualify as conventional stimulation (Ferraretti *et al.*, 2011). However, many women diagnosed with POR will have already been subjected to stimulation with average FSH doses. For these women, it may be perfectly acceptable that, in the context of an RCT, they receive high strength stimulation (300 IU–450 IU/daily).

In summarizing the above, a strong argument can be made for using the long down-regulation or the antagonist protocol as 'universal comparators' in POR research.

Multiple interventions

IVF qualifies as a complex intervention, as it includes a multitude of processes, it is being delivered by a multidisciplinary team and it has a variety of outcomes which are of interest to consumers, clinicians and funders. Furthermore, IVF allows for substantial flexibility in the planning and delivery of the treatment, with a clear trend of becoming even more individualized in the recent years (La Marca and Sunkara, 2014). Therefore,

designing and executing clinical trials on IVF patients requires careful consideration of the included processes and/or behaviours, potential interactions, as well as knowledge of the basic science in support of these interactions (Moore et al., 2015).

A common occurrence in POR research is the use of multiple interventions in the same study group with the potential to interact with each other. We are referring to the deliberate use of extra interventions as part of the original study design, not the inadvertent subtle differences in managing the compared groups which result from lack of 'blinding'. An example is the use of pretreatment, such as the combined oral contraceptive pill (COCP), a progestogen or estradiol tablets, as priming agents before the onset of down-regulation or stimulation (Smulders et al., 2010). Six RCTs, within our series, used pretreatment regimens preferentially for one of their study groups, in addition to a different stimulation protocol. Another common scenario is the use of different stimulation drug regimens in addition to different protocols for group comparisons (eight RCTs in our series).

Multiple interventions may interfere with the interpretation and generalizability of the findings of a given trial. If a significant difference in outcomes is detected between groups exposed to multiple interventions, it is impossible to confirm which intervention—or interventions—would be responsible. Even if one would refer to previous research for evidence of efficacy of these interventions, this is still methodologically inferior to a one-to-one comparison in a RCT setting. Consequently, this research may only be applicable to clinical settings where the same interventions are faithfully applied. Trials comparing multiple interventions may also introduce heterogeneity and bias in subsequent meta-analysis, when combined with trials of single interventions.

An example of a multiple intervention in POR research with high potential for bias is when one of the groups preferentially receives the COCP before an antagonist protocol. There is accumulating evidence that such pretreatment may 'dampen' ovarian response to stimulation and, thus, reduce success of the antagonist protocol (Smulders et al., 2010; Griesinger et al., 2015). Consequently, if this combined protocol proves inferior to another during a head-to-head comparison, it is not possible to generalize these findings to the original antagonist protocol (without pretreatment). Furthermore, it may be inappropriate to 'pool' together antagonist trials with and without COCP pretreatment, as this will inevitably introduce bias in the size or even the direction of the measured effect.

In summary, we would recommend that researchers make every effort to avoid comparing multiple interventions. For the common practice of pretreatment before IVF stimulation, they should aim to use the same for all groups under study or, even better, none at all. The alternative of including two groups for each protocol under study—one with and one without pretreatment—is feasible but will substantially inflate the required sample size of the planned study.

'Significant' interventions

A total of 33 interventions have been investigated against POR within our series (Table II). Of these, only 12 interventions have been reported as beneficial for reproductive outcomes in isolated trials (Table III).

Reviewing the literature is one of the principal tasks of a researcher, with a view to identifying 'gaps' that merit further research. However, deciding which intervention to investigate depends heavily on personal judgement, specialist interests and the available resources. We comment below on interventions which have been reported as potentially beneficial at least once in POR clinical trials. For such 'promising' interventions, further research is usually warranted. Nonetheless, it is not the intention of this review to scrutinize the methodological quality of separate trials; this should be part of future RCTs or systematic reviews.

Conventional protocol comparisons have been evaluated on poor responders more than any other intervention. The available trials, more often than not, have reported non-significant differences in pregnancy outcomes. The presence of substantial heterogeneity and variation in methodological quality between studies have also contributed to less than conclusive meta-analyses with a substantial risk of bias (Sunkara et al., 2007; Al-Inany et al., 2011). Nevertheless, since the two largest trials on stimulation protocols were the ones that showed significant results, it may be worth persevering with further research on the potential benefit of these protocols.

Table II Interventions investigated by RCTs in 'poor responders' (most popular intervention first).

Antagonist
Microdose flare
Long protocol
LH added
Letrozole + FSH ± antagonist
DHEA
Short protocol
Transdermal testosterone
Growth hormone
HCG added at stimulation
Increase of FSH dose
Clomiphene citrate + FSH/HMG + -antagonist
Luteal FSH start
Estrogen for luteal support
Follicular flushing
Long-stop protocol
FSH/HMG only (no agonist or antagonist)
FSH dose 300 IU
Late FSH start
Metformin
Ultrashort agonist-antagonist
Modified flare
Low-dose aspirin
Natural cycle
Mini-long protocol
Step-down of FSH dose
Luteal phase antagonist
Gamete intrauterine transfer
Day of embryo transfer
Early (Day 1) FSH start
FSH dose 450 IU
FSH dose 600 IU
Clomiphene citrate only

DHEA, dehydroepiandrosterone.

Table III Interventions with at least one RCT indicating benefit in reproductive outcomes.

Intervention	Significant outcome	Number of RCTs showing benefit	Number of RCTs showing no benefit
Estrogen add-back for luteal support	Live birth	**1 RCT** Kutlusoy *et al.* (2014)	**1 RCT** Aghahosseini *et al.* (2011)
rLH 4-day treatment followed by rFSH treatment during long protocol	Live birth	**1 RCT** Ferraretti *et al.* (2014)	**None**
DHEA supplementation	Ongoing pregnancy	**1 RCT** Moawad and Shaeer (2012)	**4 RCTs** Wiser *et al.* (2010) Artini *et al.* (2012) Kara *et al.* (2014) Yeung *et al.* (2014)
Antagonist flexible protocol (compared with microdose flare protocol)	Ongoing pregnancy	**1 RCT** Lainas *et al.* (2008)	**8 RCTs** Akman *et al.* (2001) Martinez *et al.* (2003) Malmusi *et al.* (2005) Schmidt *et al.* (2005) De Placido *et al.* (2006) Demirol and Gurgan (2009) Kahraman *et al.* (2009) Davar *et al.* (2013)
Day 2 embryo transfer (compared with Day 3)	Ongoing pregnancy	**1 RCT** Bahceci *et al.* (2006)	**None**
Long protocol (compared with antagonist protocol)	Clinical pregnancy	**1 RCT** Prapas *et al.* (2013)	**7 RCTs** Cheung *et al.* (2005) Marci *et al.* (2005) Tazegul *et al.* (2008) Kim *et al.* (2009) Shahrokh Tehrani Nejad *et al.* (2008) Kim *et al.* (2011) Sunkara *et al.* (2014)
Follicular flushing	Clinical pregnancy	**1 RCT** Mok-Lin *et al.* (2013)	**1 RCT** Levens *et al.* (2009)
Day 4 FSH start (compared with Day 1 FSH start) during antagonist protocol	Clinical pregnancy	**1RCT** Baerwald *et al.* (2012)	**None**
Transdermal testosterone	Clinical pregnancy	**1 RCT** Kim *et al.* (2011)	**2 RCTs** Massin *et al.* (2006) Fabregues *et al.* (2009)
Luteal phase FSH start	Clinical pregnancy	**1 RCT** Kucuk *et al.* (2008)	**2 RCTs** Kucuk and Sozen (2007) Kansal Kalra *et al.* (2008)
Addition of rLH mid-stimulation (compared with FSH dose increase)	Clinical pregnancy	**1 RCT** Ruvolo *et al.* (2007)	**2 RCTs** De Placido *et al.* (2001) De Placido *et al.* (2005)
High FSH dose (300 IU/day) (compared with 150 IU/day)	Clinical pregnancy	**1 RCT** Klinkert *et al.* (2005)	**None**

rLH/rFSH, recombinant LH/FSH.

Novel stimulation protocols have been sporadically investigated in the research context with promising results. A particularly interesting modality involves the use of FSH or LH during the luteal phase of a long down-regulation protocol, in an effort to improve follicular recruitment (Kucuk and Sozen, 2007; Ferraretti *et al.*, 2014). Likewise, limited but promising experimental research is available on interventions applied past the stage of ovarian stimulation, exploring other factors unrelated to oocyte quantity or quality. These have focused on the optimal timing of embryo transfer or the optimal type of luteal support (Bahceci *et al.*, 2006; Kutlusoy *et al.*, 2014). However, these findings need to be replicated in future trials before becoming part of routine clinical practice.

A degree of enthusiasm currently revolves around the use of androgen supplements as adjuvants for poor responders (Sunkara *et al.*, 2012). Androgen supplementation may have a direct beneficial effect on the 'poor responder' ovary, through an increase in the number of antral follicles or up-regulation of FSH receptors (Weil *et al.*, 1998; Nielsen *et al.*, 2011). Our review has indeed identified such supplements—transdermal testosterone and dehydroepiandrosterone (DHEA)—as of some promise,

Table IV Key clinical facts and trends in 'poor responder' research.

- The most popular criterion for defining 'poor responders' in RCTs has been low ovarian response at previous stimulation
- The most popular cut-off value for defining previous low response is 'less or equal to three retrieved oocytes'
- The most popular tests used in RCTs to define diminished ovarian reserve are AFC and FSH, followed by age and AMH
- Most research interventions were applied before/during controlled ovarian hyperstimulation
- The most popular stimulation protocols investigated in 'poor responder' research are the antagonist protocol, the microdose flare protocol and the long down-regulation protocol
- RCTs on popular protocols for poor responders have reported conflicting results with regard to oocyte yields and reproductive outcomes
- Only 1 in 10 RCTs has reported statistically significant differences in reproductive outcomes
- No 'positive' intervention is supported by more than one 'positive' RCT

based on a single 'favourable' RCT for each intervention (Kim *et al.*, 2011; Moawad and Shaeer, 2012). Nevertheless, other similar RCTs (two for testosterone and four for DHEA) have not confirmed these findings, making a strong case for further research (Table III).

A renewed interest in milder forms of stimulation on the POR population has also been documented, with six RCTs having used letrozole/gonadotrophins (Fig. 4) and three RCTs having used clomiphene citrate with or without gonadotrophins for their study groups. The rationale for a milder approach is that of non-inferiority, aiming to test if milder stimulation, in spite of potentially recruiting fewer oocytes, is as effective as conventional stimulation. Indeed, although four studies demonstrated a lower oocyte yield with these custom protocols, no differences in reproductive outcomes were reported. Further trials on the effect of milder stimulation on the Bologna responders are required in order to substantiate the role of such protocols.

The key clinical facts and trends in POR research are summarized in Table IV.

Designing future trials on poor responders

Based on this review's findings, we provide a series of recommendations for designing future trials for poor responders:

Population: Choosing to define the population based on the Bologna criteria appears to be a 'safe' approach, in view of the increasing popularity and acceptance of these criteria by the scientific community (Ferraretti *et al.*, 2011). If the researchers opt for more 'relaxed' definitions—including a group of 'suboptimal' responders'—they should be prepared to justify this decision.

Intervention: Since no single intervention stands out as a clearly beneficial one, any novel or already tested intervention can be a candidate for future trials. It is sensible to focus on interventions already shown as beneficial in at least one previous RCT. In addition, clinical research should closely follow more basic research that explores the underlying mechanisms through which these interventions may alleviate POR. Detailing these mechanisms should be an integral part of the final publication on novel POR interventions.

Comparator: For investigating novel stimulation protocols, one could opt to use the long agonist down-regulation or the antagonist protocol for one of the control groups.

Outcomes: It is important that reporting live birth becomes a priority for future POR research. Studies may continue to report ongoing/clinical pregnancy, as it is the imminent outcome reflecting success or failure of treatment and, also, a favourite among clinicians (Braakhekke *et al.*, 2014). Reporting miscarriage rates is also recommended, as it is the main adverse event after IVF treatment.

Randomization and allocation concealment: It is important to opt for a genuine randomization method, such as a computer-generated pseudo-random number sequence. For concealing allocation, one could opt for an internet-based randomization system or the old-fashioned sealed, opaque, sequentially numbered envelopes.

Blinding of patients or staff: Although a well-designed placebo is methodologically the preferred 'blinding' tool for single add-on interventions, researchers should consider the practicalities of designing a placebo, as well as patient acceptability for using one. For protocol comparisons, designing a proper placebo may be even harder. Even so, the investigators should aim to secure 'blinding' through a variety of modalities that aim to conceal sensitive methodological information from patients, investigators and assessors. In that respect, staff and patients could be trained with regard to their expected roles and conduct during the course of the trial.

Minimization of attrition bias: Publishing a CONSORT flowchart that details the patients' pathway through treatment, including withdrawals, results in less biased analyses and is strongly recommended. It is also a prerequisite for publication in select high-impact journals. ITT analysis also accounts for attrition bias and should be reported, at least in the context of 'superiority' trials.

Minimization of selective reporting bias: Researchers should aim to publish a study protocol. High-impact journals require that such a protocol has been published 'a priori' before even considering an RCT for publication. Reporting live birth and miscarriage rates also acts as a safeguard against selective outcome reporting.

Primary outcome: Studies should clearly state their primary outcome. Preference should be given to pregnancy outcomes, with particular emphasis on live birth.

Sample size: Performing sample size calculations at the design stage ensures adequate study power and minimizes the risk of a Type II error. It is likely that an adequately powered RCT on POR will require at least 150 participants per group. To recruit these high numbers, trials may have to be multi-centred.

Cost-effectiveness: Although this review has not looked specifically at how often POR RCTs have addressed cost-effectiveness issues, it appears this has not been a priority so far. Cost-effectiveness is particularly relevant to non-inferiority POR studies which focus on milder—and frequently less costly—stimulation protocols. However, in this particular field of assisted conception, even 'superiority' trials may benefit from including cost-effectiveness analyses. This is because the likelihood of demonstrating similar efficacy between compared interventions is high in POR trials. In case of two interventions being reported as equally clinically effective, it is sensible to opt for the one which is clearly more cost-effective.

Limitations

An obvious limitation of the review is that, in spite of having applied a multi-source approach for searching the scientific literature, some 'poor responder' RCTs may still have been missed. Nevertheless, we expect that, in view of the large number of included studies (75 RCTs), a small number of missing trials should not affect the overall interpretation of our findings.

We also decided to investigate a particular set of methodological standards, out of the many which have been reported by other methodology reviews in healthcare (Dechartres *et al.*, 2011). We focused on these standards because we felt that they relate to the more practical qualitative aspects of study design for the benefit of researchers. We were also conscious not to over-expand this part of the review to the detriment of the latter part, which evaluated the equally important clinical aspects of 'poor responder' trials.

Conclusions

Our observations are in support of a continuing interest being vested in POR research. This is not unexpected, if one considers that POR is common in everyday IVF practice, with evidence of increasing prevalence (Devine *et al.*, 2015).

Despite the presence of an improving trend in certain aspects of study design, the majority of published trials on POR still suffer from methodological flaws and are, thus, regarded as being high-risk of bias. The same trials have used a variety of definitions for their poor responders and a variety of interventions for their head-to-head comparisons. As a consequence, no real progress has been made in identifying clearly beneficial interventions. Moreover, the presence of clinical heterogeneity precludes the conducting of meaningful and conclusive meta-analysis.

It is recognized that ~85% of research is wasted, usually because it asks the wrong questions, is badly designed, not published or poorly reported (Chalmers and Glasziou, 2009). As a result, an initiative has recently been launched to coordinate the efforts towards reducing research waste by promoting more useful research (http://researchwaste.net). In line with this initiative, this novel type of 'methodology and clinical' review has provided a comprehensive overview of POR research, with particular focus on quality aspects of the relevant randomized trials. Based on this analysis, it has provided custom research recommendations; it has extended beyond the particulars of how to design a good trial, by also exploring which 'poor responders' and what interventions should be given research priority. By applying the lessons learned from past research, it is hoped that future trials will be of sufficient high quality to provide clear and conclusive answers on how to best manage this challenging group of women.

Supplementary data

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

Authors' roles

A.P. conceived the idea for this review, submitted the proposal, performed the literature search, data extraction and analysis and manuscript writing. B.J.S. performed data extraction and analysis and manuscript revision. N.M.A.K. performed the literature search and contributed to manuscript writing and revision. S.B. contributed to manuscript writing and revision.

Funding

No funding was received for this work.

Conflict of Interest

None declared.

References

Aghahosseini M, Aleyassin A, Khodaverdi S, Esfahani F, Mohammadbeigi R, Movahedi S, Kord Valeshabad A, Mahdavi A, Fallahi P, Shabani P *et al.* Estradiol supplementation during the luteal phase in poor responder patients undergoing in vitro fertilization: a randomized clinical trial. *J Assist Reprod Genet* 2011;**9**:785–790.

Akman MA, Erden HF, Tosun SB, Bayazit N, Aksoy E, Bahceci M. Comparison of agonistic flare-up-protocol and antagonistic multiple dose protocol in ovarian stimulation of poor responders: results of a prospective randomized trial. *Hum Reprod* 2001;**5**:868–870.

Al-Inany HG, Youssef MA, Aboulghar M, Broekmans F, Sterrenburg M, Smit J, Abou-Setta AM. Gonadotrophin-releasing hormone antagonists for assisted reproductive technology. *Cochrane Database Syst Rev* 2011;**5**:CD001750.

Artini PG, Simi G, Ruggiero M, Pinelli S, Di Berardino OM, Papini F, Papini S, Monteleone P, Cela V. DHEA supplementation improves follicular microenvironment in poor responder patients. *Gynecol Endocrinol* 2012;**9**: 669–673.

Baerwald A, Anderson P, Yuzpe A, Case A, Fluker M. Synchronization of ovarian stimulation with follicle wave emergence in patients undergoing in vitro fertilization with a prior suboptimal response: a randomized, controlled trial. *Fertil Steril* 2012;**4**:881–882.

Bahceci M, Ulug U, Ciray HN, Akman MA, Erden HF. Efficiency of changing the embryo transfer time from day 3 to day 2 among women with poor ovarian response: a prospective randomized trial. *Fertil Steril* 2006;**1**:81–85.

Braakhekke M, Kamphuis EI, Dancet EA, Mol F, van der Veen F, Mol BW. Ongoing pregnancy qualifies best as the primary outcome measure of choice in trials in reproductive medicine: an opinion paper. *Fertil Steril* 2014;**5**:1203–1204.

Broekmans FJ, Soules MR, Fauser BC. Ovarian aging: mechanisms and clinical consequences. *Endocr Rev* 2009;**5**:465–493.

Busnelli A, Papaleo E, Del Prato D, La Vecchia I, Iachini E, Paffoni A, Candiani M, Somigliana E. A retrospective evaluation of prognosis and cost-effectiveness of IVF in poor responders according to the Bologna criteria. *Hum Reprod* 2015;**2**:315–322.

Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet* 2009;**9683**:86–89.

Chatillon-Boissier K, Genod A, Denis-Belicard E, Felloni B, Chene G, Seffert P, Chauleur C. Prospective randomised study of long versus short agonist protocol with poor responder patients during in vitro fertilization. *Gynecol Obstet Fertil* 2012;**11**:652–657.

Chetkowski RJ. Consumer-friendly reporting of in vitro fertilization outcomes. *Fertil Steril* 2014;**1**:e7.

Cheung L, Lam P, Lok IH, Chiu TT, Yeung S, Tjer C, Haines CJ. GnRH antagonist versus long GnRH agonist protocol in poor responders undergoing IVF: a randomized controlled trial. *Hum Reprod* 2005;**3**:616–621.

Davar R, Rahsepar M, Rahmani E. A comparative study of luteal estradiol pre-treatment in GnRH antagonist protocols and in micro dose flare protocols for poor-responding patients. *Arch Gynecol Obstet* 2013;**1**:149–153.

Dechartres A, Charles P, Hopewell S, Ravaud P, Altman DG. Reviews assessing the quality or the reporting of randomized controlled trials are increasing over time but raised questions about how quality is assessed. *J Clin Epidemiol* 2011; **64**:136–144.

Demirol A, Gurgan T. Comparison of microdose flare-up and antagonist multiple-dose protocols for poor-responder patients: a randomized study. *Fertil Steril* 2009; **2**:481–485.

De Placido G, Mollo A, Alviggi C, Strina I, Varricchio MT, Ranieri A, Colacurci N, Tolino A, Wilding M. Rescue of IVF cycles by HMG in pituitary down-regulated

normogonadotrophic young women characterized by a poor initial response to recombinant FSH. *Hum Reprod* 2001;**9**:1875–1879.

De Placido G, Alviggi C, Perino A, Strina I, Lisi F, Fasolino A, De Palo R, Ranieri A, Colacurci N, Mollo A *et al.* Recombinant human LH supplementation versus recombinant human FSH (rFSH) step-up protocol during controlled ovarian stimulation in normogonadotrophic women with initial inadequate ovarian response to rFSH. A multicentre, prospective, randomized controlled. *Hum Reprod* 2005;**2**:390–396.

De Placido G, Mollo A, Clarizia R, Strina I, Conforti S, Alviggi C. Gonadotropin-releasing hormone (GnRH) antagonist plus recombinant luteinizing hormone vs. a standard GnRH agonist short protocol in patients at risk for poor ovarian response. *Fertil Steril* 2006;**1**:247–250.

Devine K, Mumford SL, Wu M, DeCherney AH, Hill MJ, Propst A. Diminished ovarian reserve in the United States assisted reproductive technology population: diagnostic trends among 181,536 cycles from the Society for Assisted Reproductive Technology Clinic Outcomes Reporting System. *Fertil Steril* 2015;**104**:612–619.

Dias S, McNamee R, Vail A. Evidence of improving quality of reporting of randomized controlled trials in subfertility. *Hum Reprod* 2006;**10**:2617–2627.

Dimick JB, Diener-West M, Lipsett PA. Negative results of randomized clinical trials published in the surgical literature: equivalency or error? *Arch Surg* 2001;**7**:796–800.

Fabregues F, Penarrubia J, Creus M, Manau D, Casals G, Carmona F, Balasch J. Transdermal testosterone may improve ovarian response to gonadotrophins in low-responder IVF patients: a randomized, clinical trial. *Hum Reprod* 2009;**2**:349–359.

Ferraretti AP, La Marca A, Fauser BC, Tarlatzis B, Nargund G, Gianaroli L, ESHRE working group on Poor Ovarian Response Definition. ESHRE consensus on the definition of 'poor response' to ovarian stimulation for in vitro fertilization: the Bologna criteria. *Hum Reprod* 2011;**7**:1616–1624.

Ferraretti AP, Motrenko T, Feliciani E, Tabanelli C, Magli MC, Gianaroli L. LH pretreatment as a novel strategy for poor responders. *BioMed Res Int* 2014;ID 926172.

Griesinger G, Venetis CA, Tarlatzis B, Kolibianakis EM. To pill or not to pill in GnRH-antagonist cycles: the answer is in the data already! *Reprod Biomed Online* 2015;**1**:6–8.

Hamdine O, Eijkemans MJ, Lentjes EG, Torrance HL, Macklon NS, Fauser BC, Broekmans FJ. Antimullerian hormone: prediction of cumulative live birth in gonadotropin-releasing hormone antagonist treatment for in vitro fertilization. *Fertil Steril* 2015;**104**:891–898.

Hemminki E, Hovi SL, Veerus P, Sevon T, Tuimala R, Rahu M, Hakama M. Blinding decreased recruitment in a prevention trial of postmenopausal hormone therapy. *J Clin Epidemiol* 2004;**12**:1237–1243.

Iliodromiti S, Anderson RA, Nelson SM. Technical and performance characteristics of anti-Mullerian hormone and antral follicle count as biomarkers of ovarian response. *Hum Reprod Update* 2014a;**21**:698–710.

Iliodromiti S, Kelsey TW, Wu O, Anderson RA, Nelson SM. The predictive accuracy of anti-Mullerian hormone for live birth after assisted conception: a systematic review and meta-analysis of the literature. *Hum Reprod Update* 2014b;**4**:560–570.

Johnson MO, Remien RH. Adherence to research protocols in a clinical context: challenges and recommendations from behavioral intervention trials. *Am J Psychother* 2003;**3**:348–360.

Kara M, Aydin T, Aran T, Turktekin N, Ozdemir B. Does dehydroepiandrosterone supplementation really affect IVF-ICSI outcome in women with poor ovarian reserve? *Eur J Obstet Gynecol Reprod Biol* 2014;**173**:63–65.

Kahraman K, Berker B, Atabekoglu CS, Sonmezer M, Cetinkaya E, Aytac R, Satiroglu H. Microdose gonadotropin-releasing hormone agonist flare-up protocol versus multiple dose gonadotropin-releasing hormone antagonist protocol in poor responders undergoing intracytoplasmic sperm injection-embryo transfer cycle. *Fertil Steril* 2009;**6**:2437–2444.

Kansal Kalra S, Ratcliffe S, Gracia CR, Martino L, Coutifaris C, Barnhart KT. Randomized controlled pilot trial of luteal phase recombinant FSH stimulation in poor responders. *Reproductive Biomedicine Online* 2008;**6**:745–750.

Kim C, Jeon G, Cheon Y, Jeon I, Kim S, Chae H, Kang B. Comparison of GnRH antagonist protocol with or without oral contraceptive pill pretreatment and GnRH agonist low-dose long protocol in low responders undergoing IVF/intracytoplasmic sperm injection. *Fertil Steril* 2009;**5**:1758–1760.

Kim C, Howles CM, Lee H. The effect of transdermal testosterone gel pretreatment on controlled ovarian stimulation and IVF outcome in low responders. *Fertil Steril* 2011;**2**:679–683.

Klinkert ER, Broekmans FJM, Looman CWN, Habbema JDF, te Velde ER. Expected poor responders on the basis of an antral follicle count do not benefit from a higher starting dose of gonadotrophins in IVF treatment: a randomized controlled trial. *Hum Reprod* 2005;**3**:611–615.

Kucuk T, Sozen E. Luteal start of exogenous FSH in poor responder women. *J Assist Reprod Genet* 2007;**12**:635–638.

Kucuk T, Goktolga U, Sozen E. Efficiency of follicle-stimulating hormone, commenced in the luteal phase, for overcoming a poor response in assisted reproduction. *J Obstet Gynaecol Res* 2008;**4**:574–577.

Kutlusoy F, Guler I, Erdem M, Erdem A, Bozkurt N, Biberoglu EH, Biberoglu KO. Luteal phase support with estrogen in addition to progesterone increases pregnancy rates in in vitro fertilization cycles with poor response to gonadotropins. *Gynecol Endocrinol* 2014;**5**:363–366.

La Marca A, Sunkara SK. Individualization of controlled ovarian stimulation in IVF using ovarian reserve markers: from theory to practice. *Hum Reprod Update* 2014;**1**:124–140.

La Marca A, Sighinolfi G, Argento C, Grisendi V, Casarini L, Volpe A, Simoni M. Polymorphisms in gonadotropin and gonadotropin receptor genes as markers of ovarian reserve and response in in vitro fertilization. *Fertil Steril* 2013;**4**:970–978.

La Marca A, Grisendi V, Giulini S, Sighinolfi G, Tirelli A, Argento C, Re C, Tagliasacchi D, Marsella T, Sunkara SK. Live birth rates in the different combinations of the Bologna criteria poor ovarian responders: a validation study. *J Assist Reprod Genet* 2015;**6**:931–937.

Lainas TG, Sfontouris IA, Papanikolaou EG, Zorzovilis JZ, Petsas GK, Lainas GT, Kolibianakis EM. Flexible GnRH antagonist versus flare-up GnRH agonist protocol in poor responders treated by IVF: a randomized controlled trial. *Hum Reprod* 2008;**6**:1355–1358.

Levens ED, Whitcomb BW, Payson MD, Larsen FW. Ovarian follicular flushing among low-responding patients undergoing assisted reproductive technology. *Fertil Steril* 2009;**4** Suppl:1381–1384.

Lindstrom H, Waldau S. Ethically acceptable prioritisation of childless couples and treatment rationing: 'accountability for reasonableness'. *Eur J Obstet Gynecol Reprod Biol* 2008;**2**:176–186.

Machin D, Day S, Green S (eds). *Textbook of Clinical Trials*. New Jersey, USA: Wiley, 2007.

Maggard MA, O'Connell JB, Liu JH, Etzioni DA, Ko CY. Sample size calculations in surgery: are they done correctly? *Surgery* 2003;**2**:275–279.

Malmusi S, La Marca A, Giulini S, Xella S, Tagliasacchi D, Marsella T, Volpe A. Comparison of a gonadotropin-releasing hormone (GnRH) antagonist and GnRH agonist flare-up regimen in poor responders undergoing ovarian stimulation. *Fertil Steril* 2005:**2**;402–406.

Marci R, Caserta D, Dolo V, Tatone C, Pavan A, Moscarini M. GnRH antagonist in IVF poor-responder patients: results of a randomized trial. *Reproductive Biomedicine Online* 2005;**2**:189–193.

Marusic A, Ferencic SF. Adoption of the double dummy trial design to reduce observer bias in testing treatments. *J R Soc Med* 2013;**5**:196–198.

Massin N, Cedrin-Durnerin I, Coussieu C, Galey-Fontaine J, Wolf JP, Hugues J. Effects of transdermal testosterone application on the ovarian response to FSH in poor responders undergoing assisted reproduction technique–a prospective, randomized, double-blind study. *Hum Reprod* 2006;**5**:1204–1211.

Martinez F, Coroleu B, Marques L, Parera N, Buxaderas R, Tur R, Barri PN. Comparison of "short protocol" versus "antagosnits" with or without clomiphene citrate for stimulation in IVF of patients with "low response". *Revista Iberoamericana de Fertilidad y Reproduccion Humana* 2003;**6**:355–360.

Minns Lowe CJ, Wilson MS, Sackley CM, Barker KL. Blind outcome assessment: the development and use of procedures to maintain and describe blinding in a pragmatic physiotherapy rehabilitation trial. *Clin Rehabil* 2011;**3**:264–274.

Moawad A, Shaeer M. Long-term androgen priming by use of dehydroepiandrosterone (DHEA) improves IVF outcome in poor-responder patients. A randomized controlled study. *Middle East Fer Soc J* 2012;**17**:268–274.

Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001;**9263**:1191–1194.

Mok-Lin E, Brauer AA, Schattman G, Zaninovic N, Rosenwaks Z, Spandorfer S. Follicular flushing and in vitro fertilization outcomes in the poorest responders: a randomized controlled trial. *Hum Reprod* 2013;**11**:2990–2995.

Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, Moore L, O'Cathain A, Tinati T, Wight D *et al.* Process evaluation of complex interventions: Medical Research Council guidance. *Br Med J* 2015;**350**:h1258.

Morgia F, Sbracia M, Schimberni M, Giallonardo A, Piscitelli C, Giannini P, Aragona C. A controlled trial of natural cycle versus microdose gonadotropin-releasing hormone analog flare cycles in poor responders undergoing in vitro fertilization. *Fertil Steril* 2004;**6**:1542–1547.

Nargund G, Fauser BC, Macklon NS, Ombelet W, Nygren K, Frydman R; Rotterdam ISMAAR Consensus Group on Terminology for Ovarian Stimulation for IVF. The ISMAAR proposal on terminology for ovarian stimulation for IVF. *Hum Reprod* 2007;**11**:2801–2804.

Nelson SM, Pastuszek E, Kloss G, Malinowska I, Liss J, Lukaszuk A, Plociennik L, Lukaszuk K. Two new automated, compared with two enzyme-linked immunosorbent, antimullerian hormone assays. *Fertil Steril* 2015;**104**:1016–1021.

Nielsen ME, Rasmussen IA, Kristensen SG, Christensen ST, Mollgard K, Wreford Andersen E, Byskov AG, Yding Andersen C. In human granulosa cells from small antral follicles, androgen receptor mRNA and androgen levels in follicular fluid correlate with FSH receptor mRNA. *Mol Hum Reprod* 2011;**1**:63–70.

Oudendijk JF, Yarde F, Eijkemans MJ, Broekmans FJ, Broer SL. The poor responder in IVF: is the prognosis always poor? A systematic review. *Hum Reprod Update* 2012; **1**:1–11.

Page SJ, Persch AC. Recruitment, retention, and blinding in clinical trials. *Am J Occup Ther* 2013;**2**:154–161.

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. *Cochrane Database Syst Rev* 2014; **1**:MR000035.

Pandian Z, McTavish AR, Aucott L, Hamilton MP, Bhattacharya S. Interventions for 'poor responders' to controlled ovarian hyper stimulation (COH) in in-vitro fertilisation (IVF). *Cochrane Database Syst Rev* 2010;**1**:CD004379.

Papathanasiou A. Implementing the ESHRE 'poor responder' criteria in research studies: methodological implications. *Hum Reprod* 2014;**9**:1835–1838.

Patrizio P, Vaiarelli A, Levi Setti PE, Tobler KJ, Shoham G, Leong M, Shoham Z. How to define, diagnose and treat poor responders? Responses from a worldwide survey of IVF clinics. *Reprod Biomed Online* 2015;**6**:581–592.

Polyzos NP, Sunkara SK. Sub-optimal responders following controlled ovarian stimulation: an overlooked group? *Hum Reprod* 2015 doi:10.1093/humrep/dev149.

Prapas Y, Petousis S, Dagklis T, Panagiotidis Y, Papatheodorou A, Assunta I, Prapas N. GnRH antagonist versus long GnRH agonist protocol in poor IVF responders: a randomized clinical trial. *Eur J Obstet Gynecol Reprod Biol* 2013;**1**:43–46.

Pu D, Wu J, Liu J. Comparisons of GnRH antagonist versus GnRH agonist protocol in poor ovarian responders undergoing IVF. *Hum Reprod* 2011;**10**:2742–2749.

Ragni G, Levi-Setti P, Fadini R, Brigante C, Scarduelli C, Alagna F, Arfuso V, Mignini-Renzini M, Candiani M, Paffoni A *et al*. Clomiphene citrate versus high doses of gonadotropins for in vitro fertilisation in women with compromised ovarian reserve: a randomised controlled non-inferiority trial. *Reprod Biol Endocrinol* 2012; **10**:114.

Revelli A, Chiado' A, Guidetti D, Bongioanni F, Rovei V, Gennarelli G. Outcome of in vitro fertilization in patients with proven poor ovarian responsiveness after early vs. mid-follicular LH exposure: a prospective, randomized, controlled study. *J Assist Reprod Genet* 2012;**9**:869–875.

Revelli A, Chiado A, Dalmasso P, Stabile V, Evangelista F, Basso G, Benedetto C. 'Mild' vs. 'long' protocol for controlled ovarian hyperstimulation in patients with expected poor ovarian responsiveness undergoing in vitro fertilization (IVF): a large prospective randomized trial. *J Assist Reprod Genet* 2014;**7**:809–815.

Ruvolo G, Bosco L, Pane A, Morici G, Cittadini E, Roccheri MC. Lower apoptosis rate in human cumulus cells after administration of recombinant luteinizing hormone to women undergoing ovarian stimulation for in vitro fertilization procedures. *Fertil Steril* 2007;**3**:542–546.

Schmidt DW, Bremner T, Orris JJ, Maier DB, Benadiva CA, Nulsen JC. A randomized prospective study of microdose leuprolide versus ganirelix in in vitro fertilization cycles for poor responders. *Fertil Steril* 2005;**5**:1568–1571.

Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *J Am Med Assoc* 1995;**5**:408–412.

Shahrokh Tehrani Nejad E, Hoseini Rashidi B, Ramezanzade F, Shariat M, Attar Shakeri B. GnRHa stop protocol versus long protocol in poor responder IVF patients. *Iran J Reprod Med* 2008;**6**:33–37.

Slovis BH, Check JH. Younger women with diminished oocyte reserve are not more prone to meiosis errors leading to spontaneous abortion than their age peers with normal oocyte reserve. *Clin Exp Obstet Gynecol* 2013;**1**:29–32.

Smulders B, van Oirschot SM, Farquhar C, Rombauts L, Kremer JA. Oral contraceptive pill, progestogen or estrogen pre-treatment for ovarian stimulation protocols for women undergoing assisted reproductive techniques. *Cochrane Database Syst Rev* 2010;**1**:CD006109.

Steward RG, Lan L, Shah AA, Yeh JS, Price TM, Goldfarb JM, Muasher SJ. Oocyte number as a predictor for ovarian hyperstimulation syndrome and live birth: an analysis of 256,381 in vitro fertilization cycles. *Fertil Steril* 2014;**4**:967–973.

Sunkara SK, Coomarasamy A, Faris R, Braude P, Khalaf Y. Long gonadotropin-releasing hormone agonist versus short agonist versus antagonist regimens in poor responders undergoing in vitro fertilization: a randomized controlled trial. *Fertil Steril* 2014;**1**:147–153.

Sunkara SK, Tuthill J, Khairy M, El-Toukhy T, Coomarasamy A, Khalaf Y, Braude P. Pituitary suppression regimens in poor responders undergoing IVF treatment: a systematic review and meta-analysis. *Reprod Biomed Online* 2007;**5**:539–546.

Sunkara SK, Coomarasamy A, Arlt W, Bhattacharya S. Should androgen supplementation be used for poor ovarian response in IVF? *Hum Reprod* 2012;**3**:637–640.

Tanbo T, Abyholm T, Bjoro T, Dale PO. Ovarian stimulation in previous failures from in-vitro fertilization: distinction of two groups of poor responders. *Hum Reprod* 1990; **7**:811–815.

Tazegul A, Gorkemli H, Ozdemir S, Aktan TM. Comparison of multiple dose GnRH antagonist and minidose long agonist protocols in poor responders undergoing in vitro fertilization: a randomized controlled trial. *Arch Gynecol Obstet* 2008;**5**:467–472.

te Velde ER, Pearson PL. The variability of female reproductive ageing. *Hum Reprod Update* 2002;**2**:141–154.

Tierney JF, Stewart LA. Investigating patient exclusion bias in meta-analysis. *Int J Epidemiol* 2005;**1**:79–87.

Tur-Kaspa I, Fauser B. The use of GnRH agonist in IVF protocols. 2015. http://www.ivf-worldwide.com/survey/the-use-of-gnrh-agonist-in-ivf-protocols/results-the-use-of-gnrh-agonist-in-ivf-protocols.html. (2 September 2015, date last accessed).

Veleva Z, Jarvela IY, Nuojua-Huttunen S, Martikainen H, Tapanainen JS. An initial low response predicts poor outcome in in vitro fertilization/intracytoplasmic sperm injection despite improved ovarian response in consecutive cycles. *Fertil Steril* 2005;**5**:1384–1390.

Weil SJ, Vendola K, Zhou J, Adesanya OO, Wang J, Okafor J, Bondy CA. Androgen receptor gene expression in the primate ovary: cellular localization, regulation, and functional correlations. *J Clin Endocrinol Metab* 1998;**7**:2479–2485.

Weissman A, Farhi J, Royburt M, Nahum H, Glezerman M, Levran D. Prospective evaluation of two stimulation protocols for low responders who were undergoing in vitro fertilization-embryo transfer. *Fertil Steril* 2003;**4**:886–892.

Wiser A, Gonen O, Ghetler Y, Shavit T, Berkovitz A, Shulman A. Addition of dehydroepiandrosterone (DHEA) for poor-responder patients before and during IVF treatment improves the pregnancy rate: a randomized prospective study. *Hum Reprod* 2010;**10**:2496–2500.

Wood L, Egger M, Gluud LL, Schulz KF, Juni P, Altman DG, Gluud C, Martin RM, Wood AJ, Sterne JA. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *Br Med J* 2008;**7644**:601–605.

Yeung TWY, Chai J, Li RHW, Lee VCY, Ho PC, Ng EHY. A randomized, controlled, pilot trial on the effect of dehydroepiandrosterone on ovarian response markers, ovarian response, and in vitro fertilization outcomes in poor responders. *Fertil Steril* 2014;**1**:108–115.

Younis JS. The Bologna criteria for poor ovarian response; has the job been accomplished? *Hum Reprod* 2012;**6**:1874–1875.

Zlowodzki M, Bhandari M. Outcome measures and implications for sample-size calculations. *J Bone Joint Surg Am* 2009;**91**:35–40.