Implementing the ESHRE ‘poor responder’ criteria in research studies: methodological implications

Athanasios Papathanasiou*

Institute of Applied Health Sciences, University of Aberdeen, Aberdeen, UK

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

Submitted on January 21, 2014; resubmitted on February 23, 2014; accepted on May 12, 2014

ABSTRACT: The Bologna criteria for defining poor ovarian response (POR) during IVF provide a useful template for new research in this field of assisted conception. However, designing studies around the European Society for Human Reproduction and Embryology POR criteria can be methodologically challenging, as the new definition includes various POR subpopulations with diverse baseline characteristics and unknown clinical prognosis. When designing RCTs, potential result bias may be introduced if women from each subpopulation are not evenly allocated between intervention groups. In the case of small or moderate-size RCTs, a single-sequence randomization method may not ensure balanced allocation between groups. Stratified randomization methods provide an alternative methodological approach. Depending on the chosen methodology, patient characteristics and outcomes within each intervention group may be better reported according to relevant subpopulations.

Key words: poor responder / IVF / criteria / RCT / poor ovarian response

Introduction

The publication of the European Society for Human Reproduction and Embryology (ESHRE) consensus criteria for defining poor ovarian response (POR) in IVF has provided a useful template for new research within this field of assisted conception (Ferraretti et al., 2011). Earlier research on poor responders had been characterized by substantial clinical heterogeneity, particularly related to the varying definitions used for POR by individual authors. It is anticipated that, by applying the ESHRE definition to studies involving poor responders, new research will focus on women that represent the ‘true’ poor responder population. Hence, the introduction of a uniform definition has been a welcome addition.

However, although the ESHRE consensus has set the minimum qualifying criteria for POR, it still allows for clinical heterogeneity within the newly defined population. Clinical heterogeneity, when present, has implications for the design of research studies (Rothwell, 2005). The aim of this paper is to explore clinical heterogeneity within the ESHRE POR population as well as the methodological implications for conducting controlled studies that draw subjects from this particular population.

Clinical heterogeneity within the ESHRE POR population

To further explore clinical heterogeneity within the ESHRE POR population, it is worth considering the various distinct subpopulations that are derived from possible combinations of the ESHRE entry criteria (presence or absence of risk factors, results of ovarian reserve tests and the number of previous IVF attempts with associated POR; Table I). Clinical heterogeneity can be ascertained by exploring two cardinal elements of these subpopulations, namely their baseline characteristics and their clinical prognosis.

The ESHRE criteria are inherently based on the clinical characteristics of the poor responder. Each subpopulation is defined by a unique combination of these criteria. As a result, variation in baseline characteristics between subpopulations is to be expected. For example, if age is used as the primary risk factor in the ESHRE criteria, we expect that some subpopulations (1, 2a, 2c, 3b and 3d: see Table I) include older women than others. Likewise, various subpopulations include women with pathological ovarian reserve tests (1, 2b, 2c, 3c and 3d), while others include women with normal ovarian reserve tests (2a, 3a and 3b). As a consequence, unless researchers opt to study a single subpopulation, e.g. predicted poor responders, the population sample will inevitably include women with diverse baseline characteristics.

The prognostic potential of these newly defined groups remains largely unknown, which has been highlighted by the authors of the Bologna publication (Ferraretti et al., 2011). More importantly, no head-to-head comparisons have been performed. The literature antedating the Bologna publication is likely to provide evidence on the clinical potential of these subpopulations, unless the authors have circumstance used POR criteria that closely relate to the ESHRE POR...
criteria. A retrospective study by Klinkert et al. (2004) compared the success of consecutive IVF treatment cycles in expected poor responders (age >40 years and/or high FSH) with unexpected poor responders (age ≤40 years and normal FSH). After two unsuccessful IVF attempts associated with POR (<3 oocytes retrieved), unexpected poor responders (corresponding to subpopulation 3a) achieved 15–29% ongoing pregnancy rates, while no pregnancies were achieved in the expected poor responder group (corresponding to subpopulations 3b, 3c and 3d). Although this is only a single study, it highlights the fact that prognostic differences indeed exist between some POR subpopulations.

For subpopulations where there is lack of direct comparative evidence, one can still argue that performance differences may still exist. For example, ‘predicted’ poor responders with risk factors and abnormal reserve tests (subpopulation 1) may afford better prognosis than poor responders at a previous stimulation who also have risk factors and abnormal reserve tests (subpopulation 2c). This can be argued on the basis that, despite the presence of abnormal ovarian reserve tests, a significant minority (10–20%) of women in subpopulation 1 will eventually exhibit normal ovarian response (>3 oocytes retrieved) and, subsequently, afford better prognosis (La Marca et al., 2010). It can also be argued that women belonging to subpopulation 2c represent the proportion of women from subpopulation 1 who were unsuccessful during their first IVF attempt and may, therefore, have a more guarded prognosis for the future. Similarly, subpopulation 2b is likely to perform better than subpopulation 2c, as it includes younger women with otherwise similar baseline characteristics. Although no direct comparisons are available between groups 2b and 2c, there is evidence to suggest that age is an important factor for determining live birth after IVF, not only in the general population but also in poor responders (Oudendijk et al., 2012).

In summary, the very definition of POR according to the Bologna criteria implies an element of clinical heterogeneity, which is duly reflected in the variable baseline characteristics of the respective subpopulations. Furthermore, there is indirect evidence to suggest that at least some of these subpopulations perform differently during IVF than others. From a methodological viewpoint, these considerations are enough to raise doubts regarding the clinical homogeneity of the ESHRE POR population. Researchers should be aware of the potential for population heterogeneity when designing studies on poor responders according to the ESHRE POR criteria.

### The importance of balanced patient allocation into study groups

An important methodological challenge in the design of high-quality intervention studies is to achieve even allocation of patients with similar characteristics into study groups. By doing so, the investigators can confident explain potential differences in their reported outcomes as resulting from the intervention under study. If balanced patient allocation has not been achieved, one cannot exclude the possibility of bias introduced by studying groups with different baseline characteristics and, potentially, different prognosis. Although this bias can be further explored by use of post hoc statistical analysis, it is still preferable that researchers make every effort to a priori address the issue of patient allocation (Kang et al., 2008).

In poor responder studies that draw subjects from clinically heterogeneous subpopulations, overall balanced allocation between study groups is best achieved by promoting even allocation for every POR subpopulation. This can be methodologically challenging.

### How to achieve balanced patient allocation into study groups

In the case of RCTs, randomization aims to achieve balanced patient allocation into treatment groups. Simple randomization, where patients are allocated to groups by use of a single random number sequence, is the simplest form of randomization. Simple randomization has various advantages; mainly it is straightforward to apply but also protects against selection bias (Lachin, 1988). It comes as no surprise that it is the most popular method applied in RCTs for poor responders (Pandian et al., 2010). However, it must be noted that, during simple randomization, the probability of uneven allocation is influenced by the sample size—larger samples are less likely to suffer from group imbalances (Lachin, 1988). It is estimated that for a given population, an uneven allocation of 60–40% between a treatment and a control group has a 5% probability of occurring for a sample size of 100 patients

---

**Table I** A detailed presentation of distinct poor responder subpopulations, based on possible combinations of risk factors, ovarian reserve test results and number of IVF attempts.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Subpopulations</th>
<th>ESHRE criteria fulfilled</th>
</tr>
</thead>
<tbody>
<tr>
<td>No previous IVF attempts (predicted poor responders)</td>
<td>Risk factors for POR and abnormal ovarian reserve tests (1)</td>
<td>Criteria I and III</td>
</tr>
<tr>
<td>One previous unsuccessful IVF attempt where POR was demonstrated</td>
<td>Risk factors for POR but normal ovarian reserve tests (2a)</td>
<td>Criteria I and II</td>
</tr>
<tr>
<td></td>
<td>Abnormal ovarian reserve tests but no risk factors for POR (2b)</td>
<td>Criteria II and III</td>
</tr>
<tr>
<td></td>
<td>Risk factors for POR and abnormal ovarian reserve tests (2c)</td>
<td>Criteria I, II and III</td>
</tr>
<tr>
<td>Two or more previous unsuccessful IVF attempts where POR was demonstrated</td>
<td>No risk factors for POR and normal ovarian reserve tests (3a)</td>
<td>Supplemental criterion IV</td>
</tr>
<tr>
<td></td>
<td>Risk factors for POR but normal ovarian reserve tests (3b)</td>
<td>Criterion I and supplemental criterion IV</td>
</tr>
<tr>
<td></td>
<td>Abnormal ovarian reserve tests but no risk factors for POR (3c)</td>
<td>Criterion III and supplemental criterion IV</td>
</tr>
<tr>
<td></td>
<td>Risk factors for POR and abnormal ovarian reserve tests (3d)</td>
<td>Criteria I, III and supplemental criterion IV</td>
</tr>
</tbody>
</table>

All subpopulations fulfill the ESHRE consensus criteria for POR. Numbers in parentheses indicate subpopulations within categories.
This probability becomes negligible for a sample size of 300 patients.

When applying simple randomization for RCTs on ESHRE poor responders, the appropriate sample size to promote even allocation will depend on the number of POR subpopulations that the researchers decide to study. If researchers focus on ‘predicted’ poor responders (Category 1, single subpopulation), a simple randomization technique will work as above. However, if they choose to study ‘proved’ poor responders based on a previous IVF attempt with associated POR (Category 2, three subpopulations) or other category combinations, the ‘safe’ minimum sample size to achieve balanced allocation would have to be significantly larger. In this scenario, the desired sample size needs to be large enough to achieve even allocation for the smallest subpopulation under study.

An alternative approach to randomizing groups comprising multiple subpopulations is the stratified randomization method. This technique considers the presence of POR subpopulations as a potential covariate and attempts to account for it by assigning patients to their corresponding subpopulation (covariate) group before randomization takes place. A ‘block’ clustering technique is utilized to subsequently randomize patients within each subpopulation. Details of the stratified method have been published elsewhere (Altman and Bland, 1999). Stratified randomization can achieve balanced numbers of women from each subpopulation between the intervention and control groups irrespective of sample size. This is the greatest strength of this type of randomization, which also improves the power of the study. A weakness of this technique is the potential for selection bias—the researchers may be able to guess the allocation of the last woman in every block (Lachin et al., 1988). This bias is eliminated if randomization takes place only after assigning all patients to a given block. However, this is rarely practical in real-life situations where women are recruited into studies on a first-come, first-serve basis. Another way to address this bias is by concealing allocation of the patients, followed by blinding of the investigators for the remainder of the intervention (Lachin et al., 1988). Stratified randomization is operationally more complex than simple randomization, but still more straightforward than other randomization methods (Schulz and Grimes, 2002; Kang et al., 2008). A comparison of the two randomization methods is detailed in Table II.

### Table II A comparison of the methodological properties of simple versus stratified randomization.

<table>
<thead>
<tr>
<th>Simple randomization</th>
<th>Stratified randomization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Randomizes individual patients</td>
<td>Randomizes ‘blocks’ of patients</td>
</tr>
<tr>
<td>Balanced patient allocation into treatment groups is influenced by sample size</td>
<td>Balanced patient allocation into treatment groups is guaranteed, regardless of sample size</td>
</tr>
<tr>
<td>Performs better with large sample size</td>
<td>Not influenced by sample size</td>
</tr>
<tr>
<td>Small negative impact on the power of the study if balanced patient allocation is not achieved</td>
<td>Optimal power of the study</td>
</tr>
<tr>
<td>Free of selection bias</td>
<td>May be prone to selection bias</td>
</tr>
<tr>
<td>Operationally straightforward</td>
<td>Operationally more complex</td>
</tr>
</tbody>
</table>

It is important to note that most of the published intervention RCTs for poor responders are limited by their small sample size (Pandian et al., 2010). This is to be expected from single-centre studies but may also reflect a genuine difficulty in recruiting poor responders into RCTs. Based on the experience from prior research, it may be unrealistic to expect future RCTs on POR to achieve large sample sizes, which highlights the importance of choosing the most appropriate randomization technique.

When designing non-RCTs, achieving even patient allocation between groups is equally important. As the ESHRE POR criteria allow for various baseline characteristics and combinations thereof, it may be impractical to apply matching techniques based on these characteristics. A more practical approach would be to stratify patients according to their respective subpopulation. Obviously, individual data on risk factors, ovarian reserve tests and previous IVF treatment history are required, so as to enable classification of patients into subpopulations.

### Reporting tables of group characteristics

It is highly recommended that researchers produce tables of group characteristics with between-group comparisons, so that the readers can make their own judgments on the outcome of patient allocation.

When studying poor responders who comprise of multiple subpopulations (such as women who belong to categories 2 and 3), the researchers have two options for presenting their tables of group characteristics. They may use the traditional format of reporting baseline parameters, such as age, ovarian reserve test results, and number of previous IVF attempts. If balanced allocation has been achieved, one expects that these parameters will not be different between groups. It must be noted though that the presence of overall balanced baseline characteristics does not guarantee that even allocation has taken place for every subpopulation. Alternatively, researchers may present case frequencies (number of patients in each group) stratified by subpopulation. This is more desirable if subpopulations were already utilized at the patient allocation stage of the study. Even so, it may still be appropriate to report baseline characteristics, so as to confirm similarity of the study groups. However, it should be noted that reporting subpopulations remains a novel approach that has not yet been implemented in POR research. Therefore, it is difficult to predict how this will be received by reviewers and the general readership.

### Reporting outcome measures

Presentation of outcome measures based on total group comparisons should be the main focus of the results section. When subpopulations have been used as part of the patient allocation process, for example during stratified randomization, it is reasonable to also report outcomes per subpopulation. Reporting subpopulation outcomes is likely to facilitate any subsequent meta-analyses looking at the differential prognostic potential of these subpopulations when exposed to a particular intervention. However, one should refrain from performing subgroup statistical analysis unless this had been agreed at the stage of study design and only if it addresses the research question of interest (Rothwell, 2005).
Conclusion

In conclusion, the newly introduced ESHRE definition for POR has moderated but not eliminated clinical heterogeneity within the poor responder population. When designing intervention studies on poor responder populations, accounting for such heterogeneity is expected to improve aspects of the design of the study and, by extension, enhance the overall quality of the research findings.

Funding

No funding was provided for this publication.

Conflict of interest

None declared.

References