The Bologna criteria for poor ovarian response: the good, the bad and the way forward

Christos A. Venetis

1Women’s & Children’s Health, St George Hospital, 2nd Level – Prichard Wing, Gray Street, Kogarah P.C. 2217, NSW, Australia 2School of Women’s & Children’s Health, Faculty of Medicine, University of New South Wales, Australia

*Correspondence address. E-mail: venetis@gmail.com; c.venetis@unsw.edu.au

Submitted on April 24, 2014; accepted on May 12, 2014

ABSTRACT: The management of poor ovarian response (POR) remains one of the most significant challenges of in vitro fertilization. Numerous interventions have been proposed, yet few have been shown to be beneficial. The most important problem in evaluating the available evidence has been the lack of a standardized definition of POR. The Bologna criteria for POR have been recently introduced to provide a framework allowing future research in this field to be performed on a relatively homogenous population. However, it has been suggested that the population described by the Bologna criteria might not be sufficiently homogenous and for this reason stratified randomization should be used in relevant randomized controlled trials (RCTs). Stratified randomization, besides its advantages, also has important shortcomings and for this reason it should be used only when there is clear evidence mandating such a design. Currently, there is insufficient data to support such practice in RCTs performed on the population described by the Bologna criteria for POR. Until such evidence becomes available, the scientific community should aim at evaluating interventions for POR according to the Bologna criteria in RCTs of sufficient sample size, with proper allocation concealment and masking.

Key words: ovarian stimulation / randomized controlled trials / poor ovarian response

Introduction

Since the beginning of ovarian stimulation for in vitro fertilization (IVF), the management of poor ovarian response (POR) has been a baffling riddle for clinicians (Tarlatzis et al., 2003). The challenges arising from POR are many. The pathophysiology of POR is frequently not well understood and unexpected cases of POR may occur. Prediction of POR has been one major field of research, yet the results have not been optimal (Broekmans et al., 2006). Many interventions have been proposed as antidotes for POR, yet few have shown much promise (Kolibianakis et al., 2009, 2011; Kyrou et al., 2009; Venetis et al., 2010; Bosdou et al., 2012). Multiple systematic reviews and meta-analyses have been published during recent years, all highlighting the significant heterogeneity in the definition of POR used in the various randomized controlled trials (RCTs) and calling for a universally accepted definition (Surrey and Schoolcraft, 2000; Tarlatzis et al., 2003; Kolibianakis et al., 2009; Kyrou et al., 2009; Pandian et al., 2010; Venetis et al., 2010; Bosdou et al., 2012).

The Bologna criteria: a step in the right direction?

In July 2011, the ESHRE working group on Poor Ovarian Response Definition published a consensus statement aiming to offer the first standardized definition (Ferraretti et al., 2011). This definition is based on a set of minimal criteria that should be fulfilled for a patient to be characterized as a ‘poor responder’ and aims to provide a framework so that future research in this field is performed on a relatively homogenous population.

Such a definition might allow for a better critical assessment of future evidence by reducing sample heterogeneity and also facilitate the statistical pooling of results with the methods of meta-analysis. The latter, given the small size of RCTs on POR that have been published so far, might be particularly important for the extraction of meaningful conclusions. Last but not least, a straightforward, standardized definition of POR will probably enhance the external validity and generalizability of published evidence.

The Bologna criteria: homogeneous population or not?

Although the introduction of the Bologna criteria for POR has marked a long desired change in this field, some researchers have considered that these criteria might be too broad and actually bring together diverse subpopulations with considerably different prognostic classification (Papathanasiou, 2014). For example, a major distinction might exist between those patients that are ‘expected poor responders’ and those...
that have already demonstrated a POR. Evidence from studies published prior to the Bologna criteria has shown that indeed such groups of patients may have diverse clinical behavior in their subsequent IVF cycles (Klinkert et al., 2004; Hendriks et al., 2008; Oudendijk et al., 2012).

These concerns are expressed in the opinion article by Papathanasiou (Papathanasiou, 2014) who argues that, indeed, clinical heterogeneity in the POR population as defined by the Bologna criteria is significant and for this reason the author proposes certain methodological adjustments for future RCTs in this field, i.e. stratified randomization. The article suggests that future RCTs on POR using the Bologna criteria should consider randomizing patients in eight different strata, which are defined by the possible combinations of the Bologna criteria that need to be fulfilled by a poor responder.

Is stratified randomization a necessity for future RCTs in poor responders satisfying the Bologna criteria?

It is true that stratified randomization, a form of restricted randomization, is a useful way of achieving balanced allocation regarding known important prognostic factors (Kernan et al., 1999). The approach ensures that the comparison will reflect the actual difference between the interventions, removing any potential confounding effects produced by these factors. Stratified randomization is valuable particularly when the sample size of the study is small or when interim analyses have been a priori specified. As sample size increases, the risk of covariate imbalances is diminished and thus the need for stratified randomization is reduced.

Stratified randomization has been shown to lower the probability of type I error (finding a significant difference in your sample when none exists in the actual population) (Feinstein and Landis, 1976) and also to increase the power of the study (the probability of detecting a significant difference in your sample that exists in the actual population) (Green and Byar, 1978). Furthermore, stratified randomization facilitates subgroup analyses (each strata is a smaller RCT with its own randomization list), so that a priori hypothesized differences in the efficacy of the intervention among the various subgroups can also be tested (Kernan et al., 1999).

However, stratified randomization is also characterized by potential disadvantages. The most significant one is the selection bias it might introduce, since it allows the recruiting investigators to ‘guess’ or ‘predict’ the allocation of the next patient who is eligible for enrollment (Berger, 2007). This is a major problem in RCTs, since it introduces a non-random (systematic) error and thus invalidates the conceptual base of an RCT, which is the avoidance of systematic errors in the allocation of subjects. On the other hand, the random presence of baseline imbalances in important covariates that might result from simple (non-stratified) randomization does not violate the fundamental statistical assumptions and thus statistical tests remain valid and inferences can still be made (Senn, 1989, 1991, 1994; Altman and Dore, 1991; Roberts and Torgerson, 1999).

After all, perfect balance between groups in a single experiment is never achieved and randomization is performed mainly to ascertain that any differences observed in known and, most importantly, unknown prognostic factors between the groups compared are by chance alone (non-systematic). Even if baseline imbalances in prognostic factors are present after simple randomization, it has been demonstrated that these can lead to increased type I error only when the impact of these prognostic factors on the target event is substantial (Feinstein and Landis, 1976). Moreover, even if we assume that imbalances in such important prognostic factors are present after simple randomization, the accrual of further data (from other well-designed RCTs) will neutralize a potential confounding effect and lead to accurate estimates. Therefore, stratified randomization should be used only when it is required, since the potential selection bias that it might introduce could eventually be more harmful than a random imbalance of baseline covariates after simple randomization.

Another important drawback of stratified randomization is the complexity it adds to what for many is already considered a very demanding task, i.e. the design and execution of an RCT. The number of strata should be kept as small as possible, not only for logistical reasons but also in order to avoid the problem of overstratification, which can possibly lead to covariate imbalances between groups due to incomplete filling of blocks (Pocock, 1979; Simon, 1982). Another reason to keep the number of strata small is the size of each subgroup. Assuming the researchers are interested in performing subgroup analyses and have prespecified these in their protocol, excessive stratification might lead to subgroups of small size and thus reduced statistical power to detect differences, while it will also increase the probability of type I error due to the multiple statistical tests performed.

For these reasons, it is recommended that when stratified randomization is employed, the number of strata should be kept as small as possible and also these strata should be constructed on the basis of known and important confounders (Pocock, 1979). Papathanasiou (2014) proposes that in future RCTs using the Bologna POR criteria patients should be randomized into eight different strata. Whether there are indeed eight different categories of PORs with sufficiently diverse prognosis to warrant stratification is currently not known and, thus, any such design for future RCTs on PORs using the Bologna criteria does not seem to be justified, given the added complexities and costs associated with such RCTs.

The Bologna criteria in clinical research: next steps

The accumulation of top-quality evidence regarding interventions that might be beneficial for poor responders is urgently needed. Clinical researchers should make an effort to produce data originating from well-designed RCTs with sufficient sample size and adequate allocation concealment and masking. In this way, precise and unbiased estimates of the efficacy of each intervention can be produced and can aid the clinician in deciding upon the optimal management of poor responders. RCTs of smaller size can be easily combined with the methods of meta-analysis especially when a uniform definition of POR (i.e. the Bologna criteria for POR) has been used.
Considering the concerns that have been expressed regarding the homogeneity of the population described by the Bologna criteria (Papathanasiou, 2014), it is imperative that this hypothesis is further explored. If high-quality evidence consistently shows that significant prognostic differences actually exist between these patients, then an attempt should be made to identify the distinct subpopulations of the Bologna PORs that should be taken into account when designing future studies. Until such evidence is available, however, it seems premature to consider adopting complex countermeasures of stratification into eight strata (Papathanasiou, 2014) for something that neither has yet been proven nor has yet been properly defined.

Conclusions

In conclusion, the management of POR has been one of the most challenging aspects of ovarian stimulation for IVF. Clinical research so far has been often hampered by the lack of a standardized definition of poor response. The use of the ESHRE Bologna criteria in future clinical studies might be a unique opportunity to produce data that can be compared and combined in a meaningful and straightforward way. Meticulously designed studies of sufficient sample size and with proper randomization, allocation concealment and masking should be employed in order to offer evidence of the highest quality regarding the optimal management of POR.

References