Human Reproduction, Vol.29, No.9 pp. 1842–1845, 2014
Advanced Access publication on July 9, 2014 doi:10.1093/humrep/deu139

human reproduction

DEBATE-continued

The Bologna criteria for the definition of poor ovarian responders: is there a need for revision?

Anna Pia Ferraretti* and Luca Gianaroli

Reproductive Medicine Unit, SISMR, via Mazzini 12-Bologna, Bologna 40138, Italy

*Correspondence address. Tel: +39-51-307307; E-mail: annapia.ferraretti@sismer.it

Submitted on May 8, 2014; resubmitted on May 8, 2014; accepted on May 12, 2014

ABSTRACT: The Bologna criteria were published by the European Society of Human Reproduction and Embryology (ESHRE) in 2011 to help address the lack of a clear definition of poor ovarian responders. Since its publication, aspects of the criteria have been cited several times, whilst others have been criticized. In this debate, we re-examine the criteria (which address age, the number of oocytes retrieved and the results of ovarian reserve tests) following new evidence produced and we analyse the criticism received in an attempt to determine if revisions are needed.

Key words: poor ovarian response / ovarian stimulation / ovarian reserve

Introduction

The treatment of poor ovarian responders (PORs) has been the subject of numerous randomized trials over the last two decades. However, despite the significant amount of evidence published, the numerous definitions of PORs are striking, with 47 randomized trials published until 2011 using 41 different definitions (Polyzos and Devroey, 2011). The first systematic effort to define women with a poor ovarian response to stimulation was developed by the European Society of Human Reproduction and Embryology (ESHRE) in 2011 and published with the so-called 'Bologna criteria' (Ferraretti *et al.*, 2011). Since its publication, this article was cited 98 times (Scopus CiteAlert 2-4-2014), indicating a huge interest in the topic. The Bologna criteria can help overcome the previous difficulties with studies on PORs (Polyzos and Devroey, 2011; Reynolds *et al.*, 2013; Sunkara *et al.*, 2014), and recommendations have been made to prevent randomized trials from using 'random definitions' or the publication of meta-analyses of studies using an assortment of definitions of PORs (Polyzos and Devroey, 2011). The real applicability of the Bologna criteria should be estimated through prospective clinical trials but, at this time, only retrospective pilot studies have been published (Polyzos *et al.*, 2013a,b) because of the short interval since the ESHRE definitions were developed. Nonetheless, as expected, the Bologna criteria have already been criticized for several reasons. The present article aims to open a debate regarding whether and where the Bologna criteria need to be revised based on the criticisms received and/or new evidence produced since its publication.

The Bologna criteria

In 2011 Frydman commented that 'the Bologna criteria are not absent of criticism especially because we do not know the various mechanisms of diminishing oocyte quantity, which can have different impacts on oocyte quality and may not to be sufficient to improve the outcomes of various treatments' (Frydman, 2011). Although clearly reported in the original paper, it is important to point out here that the Bologna criteria were designed to select homogeneous groups of patients based on 'oocyte quantity' for testing in prospective randomized trials for different strategies. The Bologna criteria did not presume to distinguish between alteration in oocyte quantity versus oocyte quality from assisted reproductive technologies (ART). Recurrent poor responders cannot be homogeneous with regards to pregnancy outcome (Oudendijk *et al.*, 2012) but a standardized definition is crucial to identify the optimal management of these challenging patients. According to the ESHRE consensus on the definition of POR (Ferraretti *et al.*, 2011), at least two of the following three features must be present:

(i) Advanced maternal age ($\geq$40 years) or any other risk factor.
(ii) A previous poor ovarian response (cycles cancelled or $\leq$3 oocytes with a conventional protocol).
(iii) An abnormal ovarian reserve test (ORT) (antral follicle count (AFC) $<$5–7 follicles or anti-Mullerian hormone (AMH) $<$0.5–1.1 ng/ml).

In the absence of advanced maternal age or abnormal ORT, two previous episodes of poor ovarian response after maximal stimulation are

sufficient to define a patient as a poor responder (feature iv of the Bologna criteria).

The cut-off points chosen in the 'Bologna Criteria'

Age

The ovarian response to follicle-stimulating hormone (FSH) decreases with advancing age because of the physiological decline of the ovarian follicle pool. However, in ART, age alone is not an accurate marker to predict ovarian response, unless an extreme cut-off is adopted. The decline of the follicle pool shows a sudden acceleration at ~37 years of age (Rosen *et al.*, 2012) but during ovarian stimulation for IVF, this decline reaches a maximum manifestation only after the age of 43, when the risk of a poor ovarian response is >70% (Ferraretti *et al.*, 2011). However, the age threshold in the Bologna criteria was 40 years old, based on the principle that no extreme cut-offs are necessary when using a combination of features to identify high risk populations. Several women, aged 40–42 years, undergoing their first cycle, are still able to produce more than three follicles or three oocytes, while women aged ≥40 years and also having an abnormal ORT or a previous poor ovarian response may be identified as PORs based on the evidence available. Thus, consistent studies on a large series of women are needed for clarification. However, when designing clinical trials in ART, female age is a crucial factor to consider as age alone remains the best marker of oocyte quality and the best single predictor of ongoing pregnancy in both normal and poor responders; none of the ORTs have added any value (Broer *et al.*, 2013).

Number of collected oocytes

Since the aim of IVF treatment is a live birth, an ovarian response is 'poor' when a patient has a significantly lower live birth rate compared with women from whom 'more' oocytes were retrieved. The oocyte number *per se* is not an index of oocyte quality, but a sub-optimal number of oocytes for insemination is a limiting factor for *in vitro* embryo's selection and transfer of viable embryos. The ESHRE consensus adopted the threshold of three oocytes; fewer than three oocytes was defined as a 'poor' response to conventional ovarian stimulation because that was the definition most frequently used in literature at the time, although it was based on limited scientific evidence. Two recent studies based on large clinical databases (Bhattacharya *et al.*, 2013; Polyzos *et al.*, 2014) confirmed that the threshold of three oocytes adopted by the ESHRE consensus is adequate to identify patients with a poorer prognosis for live birth rates.

ORTs

At the time of the ESHRE consensus, a large amount of research already considered AMH concentration and AFC more reliable and accurate than basal FSH (or other tests) in predicting ovarian response. Only these two markers were therefore included in the criteria. At that time, however, the cut-off values reported in the literature were extremely varied and the consensus adopted a range between 0.5 and 1.1 ng/ml for AMH and between <5 and <7 for AFC. A more recent update (La Marca and Sunkara, 2014) was not able to identify precise cut-off values that were universally adopted; ranges of values are still reported as acceptable for the prediction of poor response. According to this update, the Bologna criteria should only be adjusted for AMH (from 0.7 to 1.3 ng/ml) and the range for AFC should remain the same (<5 to <7). In the future, the variables related to the measurement of AMH (assay methods) and AFC (dimension of the follicle populations counted) will likely be standardized, and once a consensus on the thresholds of these markers is reached the Bologna criteria can be updated.

Risk factors other than age are associated with a poor ovarian response

A criticism was raised in a Letter to the Editor of Human Reproduction (Younis, 2012) regarding the brief mention and ill-defined risk factors for a poor ovarian response at a young age. During the ESHRE consensus process, a detailed definition of the risk factors was not addressed because, as underlined by Dr Younis, some are well established, some are still controversial and some novel candidates may be identified in the near future. Yet, we agree that a complete list of risk factors will lead to a more simple and reproducible definition of a POR. A revised and updated evaluation of the risk factors may be needed, in accordance with the current knowledge and evidence. However, it is clear that not one factor, so far, is able to identify or predict young PORs with high accuracy. The approach adopted by the ESHRE consensus, using risk factors as criteria to define PORs (or expected PORs) only when present in combination with other criteria, remains valid.

Homogeneity of the population fulfilling the Bologna criteria

The article by Papathanasiou (2014) published in this issue raised criticism regarding the crucial point of the ESHRE consensus: the ability to select a homogeneous population of PORs for testing in future trials. The largest part of this article ('how to achieve balanced patient allocation into study groups') is based on the remarks that the Bologna criteria may include various subpopulations with 'diverse baseline characteristics and unknown clinical prognosis'. This would produce potential bias and methodological pitfalls in the design of clinical studies. In PORs, the 'clinical prognosis' is obviously related to both the low number of recruitable follicles (reduced ovarian reserves) and to some baseline characteristics, such as age (or other factors responsible for PORs), affecting oocyte quality and the ongoing pregnancy rate.

As described in the introduction, the Bologna criteria aimed to identify homogeneous populations of women with reduced ovarian reserves, not women with similar prognoses for pregnancy. In the validation process of the ESHRE consensus, oocyte number and oocyte quality (or ovarian reserve and the prognosis for pregnancy) need to be distinguished. It is feasible (but not yet proven because of the spurious definitions of PORs used for many years) that the prognosis for pregnancy may differ among the population of PORs. As in normal responders, the most important factor affecting oocyte quality in PORs is age. The retrospective study cited by Papathanasiou to suggest a 'potential different prognosis of pregnancy' in subpopulations of PORs (Klinkert *et al.*, 2004) is a clear example of an 'inaccurate' study design. The live births were compared for different aged groups of women: expected PORs aged >40 years and

unexpected PORs aged ≤40 years. When the end-point is a live birth, the first condition required for any clinical trial in IVF is to select similarly aged women. Why should this not be observed in PORs? The authors of the ESHRE consensus did not consider it necessary to recommend this approach for future trials because it should be implied.

To properly investigate if the 'clinical prognosis' can be improved by different interventions in these challenging patients, realistic criteria are needed to identify women with reduced ovarian reserves in a simple and reproducible manner to avoid bias related to patient selection. Efforts should be mainly focused on improving the prognosis in young PORs, since age cannot be 'reduced' by any intervention. Prospective controlled trials using standardized definition of PORs will be able to finally produce evidence contributing to the most interesting, but still controversial, areas of research regarding the use of androgen or androgen-modulating agents (Bosdou *et al.*, 2012) and the value of growth hormone (De Ziegler *et al.*, 2011) in PORs.

To clarify once again, the purpose of the Bologna criteria was not to identify women with similar prognoses of pregnancy. The 'crucial question' remains if the Bologna criteria are really able to identify patients with reduced ovarian reserves or subpopulations of patients having different degrees of ovarian reserves. Papathanasiou describes different subpopulations, all fulfilling the Bologna criteria, that can be selected based on possible combinations of risk factors, ORT results and previous IVF attempts (see Table I of Papathanasiou, 2014). Some of these subpopulations are described as patients having normal ORTs (2a, 3a, and 3b). Actually, in the ESHRE consensus paper, it is never written that women had 'normal ORTs'; rather it states 'in the absence' of abnormal ORTs (see feature iv of the Bologna criteria). The rationale was to identify even women not tested with ORTs (if fulfilling the other criteria) as PORs because at the time of the ESHRE consensus the use of AFC and AMH was still limited. Today, it is realistic to believe that after one (or more) poor ovarian responses, the women will be tested for AMH or AFC before including them in clinical trials on PORs. In cases with results inside the normal range, there are two possibilities: the first episode of POR was an occasional finding and could be absent in subsequent cycles or the women have normal ORTs, but ovaries resistant to FSH/LH stimulation. In the first case, women will not meet any more of the present Bologna criteria. In the second case, if such a population exists it has to be excluded from the trial on PORs because the underlying pathology may be very different and may require different therapeutic approaches. Some doubts could remain in cases with borderlines ORT results.

Following recent evidence that documented the accuracy of AMH and AFC in the prediction of poor ovarian response, some could believe that such tests can be sufficient to 'identify' the POR population and, therefore, that the Bologna criteria are not necessary anymore. Today, the AFC or AMH measurement is recommended to individualize ovarian stimulation in clinical practice (The Practice Committee of ASRM, 2012; La Marca and Sunkara, 2014); yet, for scientific purposes, these new markers alone still remain insufficient for accurately identifying women with the highest probability of being a real POR. According to the ESHRE consensus, more than one criterion should be contemporaneously present in each subject.

Conclusions

The minimal criteria of the ESHRE consensus are not 'perfect' and may need to be revised and implemented. Thus, comments and criticisms are welcome. However, based on updated literature reports, minimal revisions should be necessary at this point. Until there is evidence on the contrary, the Bologna criteria still remain the best realistic attempt to identify, in a simple and easy reproducible manner, women with the highest probability having 'similar' reduced ovarian reserves. This is the first step to compare results and to produce evidence in the controversial issue of what kind of strategy should be proposed in this situation. Nevertheless, we completely agree that, besides age, oocyte quality of these women may be differently influenced by the various mechanisms (Frydman, 2011) and risk factors (Younis, 2012) underlying the reduction of their ovarian reserve; future researches in these fields are urgently needed to investigate the real impact of any intervention on the clinical prognosis of PORs.

References

Bhattacharya S, Maheshwari A, Mollison J. Factors associated with failed treatment: are women embarking on their first IVF cycles? *PLoS One* 2013. doi: 10.1371.

Bosdou JK, Venetis CA, Kolibianakis EM, Toulis KA, Goulis DG, Zepiridis L, Tarlatzis BC. The use of androgens or androgen-modulating agents in poor responders undergoing in vitro fertilization: a systematic review and meta-analysis. *Hum Reprod Update* 2012;**18**:127–145.

Broer SL, van Disseldorp J, Broeze KA, Dolleman M, Opmeer C, Bossuyt P, Eijkemans MJ, Mol BW, Broekmans FJ. on behalf of the IMPORT study group. Added value of ovarian reserve testing on patient characteristics in the prediction of ovarian response and ongoing pregnancy: an individualized patient data approach. *Hum Reprod Update* 2013;**19**:26–36.

De Ziegler D, Streuli I, Meldrum DR, Chapron C. The value of growth hormone supplements in ART for poor ovarian responders. *Fertil Steril* 2011;**96**:1069–1076.

Ferraretti AP, La Marca A, Fauser BC, Tarlatzis B, Nargund G, Gianaroli L. on behalf of the 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;**26**:1616–1624.

Frydman R. Poor responders: still a problem. *Fertil Steril* 2011;**96**:1057.

Klinkert ER, Broekmans FJ, Looman CW, Te Velde ER. A poor ovarian response in the first cycle is not necessary related to a poor prognosis in subsequent cycles. *Fertil Steril* 2004;**81**:1247–1253.

La Marca A, Sunkara SK. Individualization of controlled ovarian stimulation in IVF using ovarian reserve markers: from theory to practice. *Hum Reprod Update* 2014;**20**:124–140.

Oudendijk JF, Yarde F, Eijkemans MJC, Broekmans FJM, Broer SL. The poor responders in IVF: is the prognosis always poor? A systematic review. *Hum Reprod Update* 2012;**18**:1–11.

Papathanasiou A. Implementing the ESHRE "poor responder" criteria in research studies: methodological implications. *Hum Reprod* 2014; **29**:1835–1838.

Polyzos NP, Devroey P. A systematic review of randomized trials for the treatment of poor ovarian responders: is there any light at the end of the tunnel? *Fertil Steril* 2011;**96**:1058–1061.

Polyzos NP, DeVos M, Corona R, Vloeberghs V, Ortega-Hrepich C, Stoop D, Touraye H. Addition of HP HMG after corifollitropin alfa in antagonist-treated poor ovarian responders: a pilot study. *Hum Reprod* 2013a; **28**:1254–1260.

Polyzos NP, DeVos M, Humaidan P, Stoop D, Ortega-Hrepich C, Devroy P, Touraye H. Corifollitropin alfa followed by rFSH in a GnRH antagonist protocol for poor ovarian responders: an observational study. *Fertil Steril* 2013b;**99**:422–426.

Polyzos NP, Nwoye M, Corona R, Blockeel C, Stoop D, Haentjens P, Camus M, Tournaye H. Live birth rates in Bologna poor responders treated with ovarian stimulation for IVF/ICSI. *RBM Online* 2014;**28**:469–474.

Reynolds KA, Omurtag KR, Jimenez PT, Rhee JS, Tuuli MG, Jungheim ES. Cycle cancelation and pregnancy after luteal estradiol priming in women defined as poor responders: a systematic review and meta-analysis. *Hum Reprod* 2013;**28**:2981–2989.

Rosen MP, Johnstone E, McCulloch CE, Sternfeld B, Reijo-Pera RE, Cedars MI. A characterization of the relationship of ovarian reserve markers with age. *Fertil Steril* 2012;**97**:238–243.

Sunkara SK, Coomarasamy A, Fris 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; **101**:147–153.

The Practice Committee of the American Society for Reproductive Medicine. Testing and interpreting measures of ovarian reserve: a committee opinion. *Fertil Steril* 2012;**98**:1407–1415.

Younis JS. The Bologna criteria for poor ovarian response; has the job been accomplished? *Hum Reprod* 2012;**27**:1874–1879.