Is it time to revisit Rotterdam?

Why do we create definitions and classification systems for disease, and how do these help us? Or do they? Let us look at two complex disorders faced in reproductive medicine that have undergone repetitive attempts at definition and classification.

For polycystic ovary syndrome (PCOS) and poor ovarian response (POR), various bodies have formulated definitions to aid our approach to patients. I believe it is important to think about the goal of these processes, and then how they impact our interpretation of data and/or patient-specific management decisions. Webster defines a syndrome as a "group of signs and symptoms that occur together and characterize a particular abnormality or condition". The criteria developed should be simple, clearly defined, and reproducible.

How do these definitions/classification systems help us? If well done, they should allow for external validity and generalizability across populations. For those of us in medicine, they should allow properly designed clinical trials of well-characterized populations, so interpretation and application of data generated is possible in the broader population. Let us see how well we have succeeded.

The Rotterdam criteria for PCOS have undergone perhaps the most scrutiny. Before 2003, the so-called "National Institutes of Health criteria," established in 1990, required anovulation and hyperandrogenism and/or hyperandrogenemia with the exclusion of known entities. This definition identified what might be considered the "classic" and/or most severe phenotype affecting approximately 5%–8% of the reproductive aged population. The 2003 Rotterdam criteria (1) expanded the phenotypic definition by broadening the criteria requiring two of the following three conditions: oligoanovulation, clinical and/or biochemical hyperandrogenism, and polycystic ovaries. Controversy arose regarding the addition, with this definition, of two new phenotypes, one each that excluded the individual criterion from the National Institutes of Health definition—allowing for identification of PCOS in women who were ovulatory, or had no hyperandrogenism. A compromise definition by the Androgen Excess Society included the ovulatory PCOS phenotype but required hyperandrogenism as a key component of the PCOS definition. Additionally, recognition of how common "polycystic ovaries," as defined by the original criteria, was, this number has been reviewed and revised on at least two occasions.

These definitions indeed have broadened the definition and increased the prevalence of PCOS within the population to perhaps 15%–20% (2). Is this a good thing? More women now carry a diagnosis of PCOS, but how does this help them or us in their care? Is something a "pathological condition" if it has a prevalence of 20% of the population?

We care about women with PCOS for improving our ability to treat their concerns, be they anovulation and infertility or cosmetic concerns regarding acne and/or hirsutism, but also the medical concerns regarding long-term consequences: particularly, risk for diabetes, endometrial cancer, fatty liver, and, potentially, heart disease with known dyslipidemia. However, most studies that have evaluated metabolic risk in PCOS show the strongest association with hyperandrogenism (3). Has adding those without this component been helpful? Additionally, risk for fatty liver, also seems to correlate with androgen levels. Anovulation does continue to increase the risk for endometrial hyperplasia/cancer, but the additional "diagnosis" of PCOS does not aid in decisions regarding management. It also has been said that it is important to confer the diagnosis on ovulatory women with a high follicle number due to risk for ovarian hyperstimulation and required caution with ovulation induction. Again, given the correlation between antimüllerian hormone (AMH) levels, and antral follicle count, and ovarian hyperstimulation syndrome risk, it is not clear whether "PCOS" as a diagnosis is required to improve patient care/counseling.

What about research? If research studies isolated individual phenotypes to identify best treatment strategies, we could possible learn more about these "more mild" phenotypes and how to address specific concerns to optimize care. Unfortunately, many studies describe recruitment according to the "Rotterdam criteria," which actually lessens the reproducibility and uniformity of studies to generate data that consistently improve care. Additionally, the individual criteria themselves do not have standard definitions given variability of androgen assays, definitions of irregular cycles and anovulation, and ovarian assessment given importance of individual ultrasound equipment and examiner skills and lack of standardization of AMH assays. I would argue that the Rotterdam criteria have not helped us either care for women or design research as currently used.

Poor ovarian response (POR) during ovarian stimulation for in vitro fertilization is a core problem for practitioners. Correlation between oocyte number, especially with increasing age—in the face of increasing aneuploidy—and pregnancy outcome has made optimization of stimulation a critical aspect of care. The lack of uniformity of definition for POR led to the Bologna criteria as identification of optimal treatment strategies for individual patients was complicated by the variability in definitions used in previously reported trials.

The goal of the Bologna criteria was to select homogeneous groups based on oocyte quantity allowing for prospective studies to identify best treatment modalities. Again, a 2 of 3 schema was used including: advanced maternal age ($\geq$40) or other risk factors for POR, previous poor response ($\leq$3 oocytes with conventional stimulation), and abnormal ovarian reserve testing (antral follicle count of $<$5 or AMH level of $<$0.5–1.1 ng/mL) (4). Although this was perhaps a useful first step, and met the goal of being simple and reproducible, there are several concerns regarding the application of the Bologna criteria. The first is that this definition does not separate those who have a low but appropriate response (the patient who has diminished ovarian reserve) from the individual with intrinsic capacity who underperforms. This is critical, and also reflects the lack of focus on the ovarian reserve testing, particularly antral follicle count in which expected response can be predicted and is relevant for counseling purposes before the first

stimulation start (pointing out the second flaw—requirement for a prior poor response). The use of ovarian reserve testing to predict expected response and counsel patients regarding expectations with "stimulation" is key to defining and optimizing care for women with POR. From the standpoint of optimizing study design, knowing the expected response—and a focus on those with diminished ovarian reserve, using ovarian reserve criteria for study inclusion, will optimize external validity and generalizability. For treatment outcome of oocyte "number," ovarian reserve is the most important predictor. However, the other concern regarding the Bologna criteria and this becomes more relevant as we drive improvement in clinical trial design to focus on the outcome of importance—live birth—is the impact of patient age.

The patient-oriented strategies encompassing individualized oocyte number (POSEIDON) criteria had as a goal recognizing this distinction and separating, based on age (quality), and ovarian reserve (quantity), probability for successful pregnancy (5). The patient-oriented strategies encompassing individualized oocyte number criteria, thus, addressed the primary concerns with the Bologna criteria: separation of appropriate but low response from poor response, and focus on the impact of age regarding "importance" of oocyte number. Additionally, POSEIDON uses ovarian reserve testing prominently in the criteria allowing stratification before the first stimulation. These criteria added stratification to reduce heterogeneity and provide individualization for research design and treatment recommendations. The clinical focus, equating to likelihood for pregnancy, added success as defined by ability to produce the number of eggs required to produce one euploid embryo. Studies already have started using individual phenotypes to optimize outcomes.

Is it possible to do something similar for PCOS? Although we can argue regarding the significance of androgen and/or anovulation for the diagnosis, it is critical that we find a common "language" to describe study participants to allow interpretation of the study and application of the findings. The interpretation should have clinical relevance, such as in POSEIDON predicting chances for successful pregnancy.

What end points/outcomes are most relevant for individuals with PCOS? If the outcome in PCOS is likewise successful pregnancy, then inclusion criteria would consist of phenotypes with anovulation, then potentially stratified on the basis of the presence/absence of hyperandrogenism and body mass index. Interestingly, if our primary goal is protection from long-term risk, phenotypes that include hyperandrogenism seem to be most important (3). So it can be seen that isolation of individual phenotypes becomes important if we clarify our desired outcome. This improves homogeneity of the population for study, strengthens the outcomes, and allows more appropriate application of study results to each individual phenotype and to individual patients. If studies, specifically called out enrollment included only those with ovulatory PCOS or only those with anovulation and hyperandrogenism, we could better apply the knowledge gained and be more likely to see reproducibility in other studies using the same population.

I know that the international PCOS guidelines will be revisited soon. I put out a plea to include an updated, and more clinically meaningful, classification system that encourages focused research improving health for women who carry this diagnosis based on their individual clinical characteristics.

Marcelle I. Cedars, M.D.
Division of Reproductive Endocrinology, Department of Obstetrics, Gynecology and Reproductive Sciences, University of California at San Francisco, San Francisco, California

https://doi.org/10.1016/j.fertnstert.2022.01.031

DIALOG: You can discuss this article with its authors and other readers at https://www.fertstertdialog.com/posts/34665

REFERENCES

1. Rotterdam ESHRE/ASRM-Sponsored PCOS Consensus Workshop Group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome. Fertil Steril 2004;81:19–25.
2. March WA, Moore VM, Willson KJ, Phillips DI, Norman RJ, Davies MJ. The prevalence of polycystic ovary syndrome in a community sample assessed under contrasting diagnostic criteria. Hum Reprod 2010;25:544–51.
3. Barber TM, Wass JA, McCarthy MI, Franks S. Metabolic characteristics of women with polycystic ovaries and oligo-amenorrhoea but normal androgen levels: implications for the management of polycystic ovary syndrome. Clin Endocrinol (Oxf) 2007;66:513–7.
4. 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;26:1616–24.
5. Alviggi C, Andersen CY, Buehler K, Conforti A, De Placido G, Esteves SC, et al. A new more detailed stratification of low responders to ovarian stimulation: from a poor ovarian response to a low prognosis concept. Fertil Steril 2015;105:1452–3.