What’s in a number? The polycystic ovary revisited

Nick Raine-Fenning1,2,*

1Division of Human Development, School of Clinical Sciences, University of Nottingham, Nottingham, UK; 2Nottingham University Research & Treatment Unit in Reproduction (NURTURE), Nottingham University Hospitals NHS Trust, Queen’s Medical Centre Campus, B Floor, East Block, Derby Road, Nottingham NG7 2UH, UK

*Correspondence address. E-mail: nick.fenning@nottingham.ac.uk

So 19 is the new 12! This is certainly a long overdue change, but is the cut-off value proposed by Dewailly et al., in a well-designed study and elegant cluster analysis presented in this edition of Human Reproduction, the right one? Only time and external validation will tell but what is certain is that the old figures, proposed by the same group in 2003 (Jonard et al., 2003) and subsequently adopted by the Rotterdam consensus (2004a), need to be reconsidered urgently (2004b). We also see the introduction of anti-Müllerian hormone (AMH) as a potential new diagnostic criterion. Whether these variables can be used in conjunction or as separate entities remains to be seen, but the known close correlation between AMH and the number of antral follicles (Andersen et al., 2010; Jayaprakasan et al., 2010) suggests that they will not offer mutually exclusive information. Dewailly et al. suggest that AMH is the better of the two markers. The data are welcome as the current ultrasound criteria identify a morphological feature that has been shown to be present in up to 30% of the general population (Duijkers and Klipping, 2010; Johnstone et al., 2010; Kristensen et al., 2010). Can something that common truly be considered a ‘disease’? We will undoubtedly see a reduction in the apparent prevalence of polycystic ovaries (PCO) if the criteria are modified on these new data from Dewailly et al., but are the new thresholds enough or do we need to reconsider the diagnosis in more detail?

In this study, Dewailly et al. used cluster analysis to reconsider the threshold values suggested by the Rotterdam consensus. This was based on the hypothesis that women with PCO in isolation, or as the authors say ‘polycystic ovarian morphology’ (PCOM), and therefore no features suggestive of anovulation or hyperandrogenism, are normal and different from the control population. In essence they would simply represent the upper centile of patients as reflected by their own antral follicle and AMH reference ranges. Their decision to exclude these patients, subsequently defined after cluster analysis as ‘Group 1b’, from the control group means they have potentially excluded normal women. Are these the women we need to consider in more detail, namely those with isolated PCO? The authors validate their exclusion by saying that these women tend to have slightly higher mean serum androgens (Adams et al., 2004; Mortensen et al., 2009) or AMH levels (Johnstone et al., 2010). This will depend on the criteria one uses to define normality and any respective cut-off levels applied. We are using percentiles to define disease rather than clinically relevant features. The authors go on to suggest that PCOM in women without oligoamenorrhoea and/or hyperandrogenism may still ‘… represent a functional entity that may be considered as a silent form of PCOS—a concept originally suggested by Franks et al. (2008). They also suggest that serum AMH might be the best marker to define such a group. This brings us back to the application of thresholds, which are based on the premise that a disease is present or not: a dichotomy rather than a continuum. The cut-off, as readers will know, simply represents the best compromise between the sensitivity and specificity of the test, which in this study was 81 and 92%, respectively for a threshold of 19 follicles per ovary. It is a statistical entity as opposed to a clinically relevant discriminator.

The introduction of AMH makes sense as it has become a stable assay and provides a reliable measure of ovarian reserve that can be ascertained at any time in the menstrual cycle (La Marca et al., 2004). It is this combination of predictive power and simplicity, which makes AMH so attractive. Counting antral follicles is less straightforward and more time-consuming. The inherent intra- and inter-observer variation seen in measuring the number of antral follicles can be reduced but not eliminated by the use of three-dimensional (3D) ultrasound and, more recently, automated ultrasound. AMH more closely correlates with the number of small antral follicles (Pigny et al., 2003), which are harder to identify separately from larger follicles with ultrasound. The cohort of small follicles measuring 1–5 mm do seem to more accurately reflect ovarian reserve and are increased in women with polycystic ovarian syndrome as Dewailly’s group showed in their original 2003 paper (Jonard et al., 2003). These smaller follicles do generally contribute most to the overall pool of antral follicles, which is why counts that include larger follicles, such as those measuring 9 or 10 mm, are still clinically valid, as they are not significantly diluted by the cohort of antral follicles measuring 6 mm and more. Such counts are also easier to perform, and therefore more practical, as the observer needs to exclude only follicles above 9 or 10 mm. Automated 3D measures do provide reliable information on follicle number and size and can be used to individually identify each follicle (Deb et al., 2010) and appear to be more reflective of ovarian reserve and response (Franks et al., 2008; Deb et al., 2009) but their accuracy remains to be determined and requires further work. The authors suggest that AMH is more specific than AFC for the diagnosis of the polycystic ovary and go on to say that ultrasound is also ‘… less sensitive than
AMH in detecting PCOM in the non-PCOS group and in patients with mild PCOS as well. How can ultrasound be less sensitive than AMH in identifying the ultrasound-based end-point itself? This is illogical and an artefact of the reliance on antral follicle cut-offs proposed by Rotterdam and now reinforced once again by Dewailly et al. This analysis gives a misleading impression of the role of ultrasound in this context.

However one considers this new information, AMH offers an interesting and attractive option for the diagnosis of PCOS. Has ultrasound had its time in the spotlight therefore as far as the diagnosis of PCOS is concerned? What does seem important is the need to separate the ultrasound findings from the criteria used to diagnose the syndrome. Anyone who regularly undertakes pelvic ultrasound will tell you that some ovaries have a classic polycystic appearance and one that resembles the original description put forward by Adams et al. in 1985 (Adams et al., 1986). The ovary is enlarged, its stroma hyperechogenic relative to its cortex and highly vascular, and, above all, there are numerous, small follicles arranged around the periphery. These follicles are typically symmetrical and measure between 1 and 5 mm in keeping with the disordered folliculogenesis that is not only associated with the syndrome but that actually defines the underlying pathophysiological process. The polycystic ovary typically contains multiple small follicles of a similar size in comparison with a multifollicular ovary where there are numerous follicles, but of varied sizes showing a physiological progression from small through to large in keeping with normal folliculogenesis. This, I believe, cannot be shown by AMH, which only gives an overview of the antral follicle population and provides no information on the various follicle cohorts or distribution of follicle sizes within the ovary. Maybe this information is not important and maybe it is too difficult or time-consuming to obtain, but I do feel further work is required to see whether the pattern of folliculogenesis is different between normal women, women with PCOM, and those with PCOS.

We must remember that PCOS is an endocrine disorder and should be managed as such. Endocrinologists do not define the presence of a disease by the results of a blood test but by the clinical features the patient presents with. Gynaecologists making a diagnosis of PCOS should adopt the same process. The metabolic abnormalities and phenotypes associated with PCOS are not included in the definition of the syndrome, as Dewailly et al., highlight, because it is ‘…unclear whether these features are intrinsic to the disease or not’ (Moran and Teede, 2009). The ultrasound findings should, however, be considered as a solitary result and interpreted alongside the patient’s symptoms and signs evident on physical examination. Just as an endocrinologist would not make a diagnosis of thyrotoxicosis or hypothyroidism based on the results of a single thyroid function test or confirm Cushing’s syndrome on the basis of serum cortisol we should not label someone as having PCO simply according to the findings of the ultrasound scan or serum AMH levels. This, in fact, is exactly what the Rotterdam consensus proposed but has been misconstrued by some. The criteria still highlight the importance of clinical features and the diagnosis cannot be made on the basis of the ultrasound findings alone. People should not worry when defining an ovary as polycystic as this alone means nothing. It does, however, allow a diagnosis of PCOS to be made if there is coexistent oligoamenorrhoea and/or hyperandrogenism. The question remains are these two entities of equal importance and do they reflect the endocrinopathy that is PCOS? Another pertinent question is whether the diagnosis of PCOS should be made in oligoamenorrhoeic, hyperandrogenic women with normal ovaries on ultrasound? We can now add AMH to this. Surely a syndrome defined, in name at least, as one that has an ovarian origin or manifestation cannot exist if the ovary appears to be normal? Maybe it is time we renamed the PCOM: is this an opportunity to differentiate it from PCOS and emphasize that we are talking about follicles and not cysts? A ‘polycystic follicle ovary’ would be more appropriate or we can revert to the ‘multifollicular ovary’ description used in the past. Either would be more appropriate and also easier for patients to understand. It would also offer reassurance, where appropriate, and avoid potential stigmatization or misconception as to the underlying problem.

So, which of the three entities is the most important or are all three of equal biological and clinical significance? Many, and in particular the Androgen Excess Society, would say that hyperandrogenism trumps oligoamenorrhoea. Furthermore, they would argue that PCOS is really a disorder characterized by an increased production of ovarian androgens, which in turn affect folliculogenesis and lead to the changes in the ovary seen on ultrasound and histology (Azziz et al., 2006; Dewailly et al., 2006; Franks et al., 2006). Indeed, Dewailly et al. have also recently shown, through principal component analysis, that a high antral follicle count (AFC) is itself a sign of hyperandrogenism (Dewailly et al., 2010b). An aberrant folliculogenesis may or may not result in anovulation but even if the cycle remains regular, the woman still has PCO and polycystic ovarian syndrome. Contrary to this, the current Rotterdam classification also permits the diagnosis of PCOS in anovulatory women without obvious hyperandrogenism if they have ultrasound features of the disease or, more specifically, have an AFC of more than 11 or an ovary measuring 10 ml or more. The proposal by Dewailly et al., in a previous paper, that the diagnosis of PCOS should first be made on the basis of oligo-anovulation and hyperandrogenism and that ultrasound (or AMH) be restricted to those in whom one of these criteria is absent does not address this, as it still allows the diagnosis of PCOS in the absence of hyperandrogenism (Dewailly et al., 2010a). Androgens must be important, as >90% of women with hirsutism can be shown to have PCO on ultrasound examination. Endocrine and histological features almost identical to those seen in humans can be induced in rats through the administration of androgens, which are known to induce a multifollicular pattern, increase the number of atretic follicles, promote luteinization of the theca cells and, eventually, lead to hyperthecosis of the ovary. Atretic follicles and hyperplastic thecal cells are also a significant source of androgen over-production in PCOS, which suggests that the endocrinological and histological features of the syndrome are both causative and reactive (Lam and Raine-Fenning, 2009).

This is not the first time a new threshold has been suggested for the ultrasound-based diagnosis of PCO. The number of follicles considered necessary to establish the diagnosis of PCO by ultrasonography has changed over the years from ‘more than 5’ (Yeh et al., 1987) to ‘more than 10’ (Adams et al., 1985) and then ‘at least 15’ (Fox et al., 1991; Battaglia et al., 1999). There was an obvious lack of agreement, which was why the Rotterdam group decided to simplify and objectify the ultrasound criteria. The distribution of the follicles, the brightness of the stroma and ovarian blood flow were not included, as these are subjective. This approach was somewhat pragmatic but at least allowed a relatively straightforward assessment and therefore...
various treatments for hyperinsulinaemia aimed at women with an iso-
tative change? The evidence is not convincing, and studies investigating
all? Does the diagnosis change management? and if so, is this an effec-
sereum AMH levels? Indeed, is an ultrasound description required at
likely to be affected by the ultrasound appearance of the ovaries or
women not responding to ovulation induction. Are these treatments
however, as it allows the introduction of specific treatment strategies,
reduce serum androgen levels. A diagnosis of PCOS is important,
and concerns. Treatment should aim to induce ovulation and or
reduce serum androgen levels. A diagnosis of PCOS is important,
but it still leaves us with a numerical value. Such numbers and
cut-off levels, in my opinion, mask the fact that we are dealing with
an endocrine disease that, as with all endocrinopathies, has various
manifestations that depend on various genetic and environmental
factors. It would be great if we could input several results into a
model and make a definitive diagnosis, but this is not realistic and
does not consider basic biology or physiology. We should be managing
a patient’s symptoms and not blood results or ultrasound findings.
Such management should be directed by the patient’s phenotype
and concerns. Treatment should aim to induce ovulation and or
reduce serum androgen levels. A diagnosis of PCOS is important,
however, as it allows the introduction of specific treatment strategies,
including the use of insulin-sensitizing agents and ovarian diathermy in
women not responding to ovulation induction. Are these treatments
likely to be affected by the ultrasound appearance of the ovaries or
 serum AMH levels? Indeed, is an ultrasound description required at
all? Does the diagnosis change management? and if so, is this an effec-
tive change? The evidence is not convincing, and studies investigating
various treatments for hyperinsulinaemia aimed at women with an iso-
lated ultrasound diagnosis of PCO have not shown any benefit
(Swanton et al., 2011). To some degree, this is logical, but the
current Rotterdam consensus and the use of strict ultrasound criteria
to define an ovary that is different from the norm has the potential to
confuse healthcare practitioners who incorrectly label the patient as
someone with a disease. I cannot see how bringing AMH into the
equation will address this.
Ultrasound features and serum AMH levels are relevant, however,
in determining ovarian reserve as measured by the response to con-
trolled ovarian stimulation. An exaggerated response and ovarian
hyperstimulation syndrome are more likely in women with PCOS.
However, while these risks may be more in women with PCOS,
they also affect women who have numerous antral follicles and those
with PCO who have no other features of the disease. The
values presented in this new paper may be more sensitive in predicting
such outcomes but once again, as with the diagnosis of PCOS, the
response is not an ‘all or nothing’ phenomena and the use of cut-off
levels ignores the fairly linear relationship seen between antral follicle
numbers or AMH levels and oocyte yields. Cut-off levels could be
used to offer different treatments clinically or in the context of ran-
domized controlled trials, but the development of population, age-
based nomograms is more logical, as these can be used to inform clin-
icians and patients of their likely response to ovarian stimulation at any
given AFC or AMH level. Work by our own group suggests that the
AFC is a significant predictor of ovarian hyper-response and ovarian
hyperstimulation syndrome. The antral follicle cut-off we defined of
22 is in keeping with the normal ranges defined by the Rotterdam con-
sensus, because a woman with an AFC above 22 must have at least
one polycystic ovary by definition, as this represents the lowest poss-
table total, a combination of 11 antral follicles in both ovaries, where a
diagnosis of PCOM cannot be made.
I think the concept of defining cut-offs or thresholds, call them what
you will, is illogical. The authors must be congratulated on trying to
address the commonly held belief that the current levels suggested
are too low, but I do not feel we should search for non-physiological
criteria as if they represent a Holy Grail to define what, after all, is a
syndrome and therefore a collection of symptoms, signs and test
results all rolled into one. The populations studied were nicely
defined and considered such issues, at least from a ultrasound per-
spective, but the subsequent analysis, albeit elegant and accurately
applied to the last detail, then ignored these issues. This is highlighted
in the title and first line of the discussion both of which highlight the
authors desire to identify ‘...a single marker...able to separate
accurately women with normal ovaries from those with PCOM
among a group of asymptomatic women’. I feel this ignores the con-
tinuum we see when scanning women and the fact that an ovary
cannot simply be considered normal or abnormal.
It is important to highlight, as the authors do, that these new
threshold values are also derived from a fairly small group of
women aged between 18 and 35 and are not, therefore, necessarily
applicable to women outside these age groups. Only 62 women
would have been classified as having PCOS by the current Rotterdam
criteria with a further 73 only considered to possibly be polycystic in
the final study. These inclusion criteria imply only women with hyper-
androgenism and oligoamenorrhoea have PCOS. The control popu-
lation was, as is often the case, derived from women referred to
hospital but subsequently considered to be ‘normal’ as opposed to
a true sample of the population at large. The women with PCOS
(27.6 ± 20.1–34.0 years) and those with oligoamenorrhea or hyper-
androgenism in isolation (28.7 ± 21.4–32.8) were both significantly
younger than the controls (30.0 ± 21.9–34.6 years), which will also
influence the number of follicles and serum AMH levels as both are
related with age. The definition of oligoamenorrhea was based on
an average cycle length of more than 35 days and not on serum
progesterone levels or ultrasound evidence of anovulation. The defi-
nition of hyperandrogenism was more definitive but included clinical
hyperandrogenism defined, in some cases, by the presence of acne
in more than two areas. The ultrasound data were derived as originally
described in the groups corresponding 2003 paper. Ovarian volume
was estimated from linear measures using the ellipsoid formula and
AFC were made by counting all visible follicles under 10 mm. All
assessments were performed with two-dimensional (2D) ultrasound,
which is acceptable but has been shown to be less reliable than three-
dimensional (3D) ultrasound. Interestingly, they took the mean ovarian
volume and the mean AFC rather than analyse each ovary indepen-
dently as recommended by the Rotterdam consensus, which means
that these thresholds cannot be applied to one ovary in isolation.
As the authors comment, continued advances in ultrasound equip-
ment and software mean that the resolution continues to improve and
more and more smaller follicles are now evident. The increase in
threshold from 12 to 19 may simply reflect the inclusion of smaller fol-
licles, specifically those measuring <2 mm, which again are not
included in the current Rotterdam classification. These follicles may be evident and therefore quantifiable but are they clinically relevant? The authors repeatedly suggest that their findings and revised cut-offs reflect the improvements in ultrasound. This may be true but cannot be considered in isolation and without due consideration to measurement reliability. Dewailly et al. suggest their own 3D findings equate with their 2D results, but these data are not shown and this is not in agreement with other authors’ experiences. The exact resolution of an ultrasound machine is hard to ascertain. Manufacturers rightly state that this depends on the organ of interest and a multitude of settings but ultimately relates to the pulse length, a variable not often discussed by the manufacturer, which more accurately defines the best resolution achievable under ideal circumstances. Despite the authors’ claims, this has not actually changed that much in recent years and many modern machines can resolve an image from around 0.4 mm which is roughly the size from which antral follicles become FSH-responsive. Even if we accept that the resolution of some machines is less than this, most can readily identify structures of 1 mm and, therefore, demonstrate the smaller antral follicles.

So where are we now? I know that I am not alone in thinking that the current ultrasound diagnosis of a polycystic ovary based on the presence of 12 or more antral follicles measuring 2–9 mm in diameter needs reconsideration. Many, including myself, will welcome these new data suggesting that a level closer to 19 is more appropriate and others will undoubtedly be pleased to see AMH levels added to the diagnostic armamentarium: a threshold of 35 pmol/l is suggested. I congratulate Dewailly and his team for revisiting their 2003 work and for providing us with these new reference levels. I would, however, urge caution in the application of the proposed cut-off levels in the clinical setting and ask people to treat both AFCs and AMH levels as we treat other ultrasound findings and endocrine results; namely, as one part of a much bigger picture. We should remember that PCOS is an endocrine disorder and should be managed as such, which means treatments should be driven by the clinical features of the disease rather than by blood results and ultrasound findings, which only give a guide to disease activity and progression.

References


Deb S, Batcha M, Campbell BK, Jayaprakasan K, Clewes JS, Hopkisson JF, Sjolblom C, Raine-Fenning NJ. The predictive value of the automated quantification of the number and size of small antral follicles in women undergoing ART. Hum Reprod 2009;24:2124–2132.


Dewailly D, Catteau-Jonard S, Poncet E. Which morphological investigations and how to interpret them to make the diagnosis of PCOS? Annales d’endocrinologie 2010a;71:183–188.


Duijkers IJ, Klipping C. Polycystic ovaries, as defined by the 2003 Rotterdam consensus criteria, are found to be very common in young healthy women. Gynecol Endocrinol 2010;26:152–160.


Mortensen M, Ehrmann DA, Littlejohn E, Rosenfield RL. Asymptomatic volunteers with a polycystic ovary are a functionally distinct but heterogeneous population. *J Clin Endocrinol Metab* 2009;94:1579–1586.


