Biomarkers of ovarian response: current and future applications

Scott M. Nelson, Ph.D.
School of Medicine, University of Glasgow, Glasgow, United Kingdom

With our increasing appreciation that simply maximizing oocyte yield for all patients is no longer an appropriate stimulation strategy and that age alone cannot accurately predict ovarian response, there has been an explosion in the literature regarding the utility of biomarkers to predict and individualize treatment strategies. Antral follicle count (AFC) and antimüllerian hormone (AMH) have begun to dominate the clinical scene, and although frequently pitted against each other as alternatives, both may contribute and indeed be synergistic. Their underlying technologies are continuing to develop rapidly and overcome the standardization issues that have limited their development to date. In the context of in vitro fertilization (IVF), their linear relationship with oocyte yield and thereby extremes of ovarian response has led to improved pretreatment patient counseling, individualization of stimulation strategies, increased cost effectiveness, and enhanced safety. This review highlights that although biomarkers of ovarian response started in the IVF clinic, their future extends well beyond the boundaries of assisted reproduction. The automation of AMH and its introduction into the routine repertoire of clinical biochemistry has tremendous potential. A future where primary care physicians, endocrinologists, and oncologists can rapidly assess ovarian dysfunction and the ovarian reserve more accurately than with the current standard of follicle-stimulating hormone (FSH) is an exciting possibility. For women, the ability to know the duration of their own reproductive life span will be empowering and allow them to redefine the meaning of family planning. (Fertil Steril® 2013;99:963–9. © 2013 by American Society for Reproductive Medicine.)

Key Words: Anti-mullerian hormone, antral follicle count, biomarkers, OHSS, ovarian response, ovarian stimulation

Discuss: You can discuss this article with its authors and with other ASRM members at http://fertstertforum.com/nelsonsm-biomarkers-ovarian-response-amh-afc/

With our increasing appreciation that simply maximizing oocyte yield for all patients is no longer an appropriate stimulation strategy (1) and that age alone cannot accurately predict ovarian response, there has been an explosion in the literature regarding the utility of biomarkers to predict ovarian response and individualize treatment strategies. Although follicle-stimulating hormone (FSH) concentration and dynamic tests have served us well, their multiple deficiencies have been highlighted by the introduction of alternative, more informative markers such as antral follicle count (AFC) and antimüllerian hormone (AMH). The improved performance of these two biomarkers is largely due to their significantly stronger correlations with primordial follicle counts (2) (Fig. 1). Therefore, although elevated FSH remains informative and continues to be a defining characteristic of menopause (3), for assisted conception use its days may be limited. Similarly, when equivalent information can be achieved without dynamic testing, patient convenience will dictate a simpler route. At present, we are in a relative state of flux where the strengths and limitations of these newer markers continue to be fully elucidated. As with all new technologies, there is a spectrum of opinion regarding their usefulness, ranging from the early adopters to those who are potentially more reticent. The aim of this article is thus not to replicate the comprehensive systematic reviews of all biomarkers (4), but rather to focus on AFC and AMH, which are now beginning to dominate clinical practice, acknowledge their inherent limitations, and propose a vision for the future.

**THE LACK OF STANDARDIZATION**

The introduction of new technology is always accompanied by ongoing technical development and the inherent problems of changing indications for use and standards. Both AFC and AMH have suffered from these problems. For AFC, the primary issues are related to the dramatic improvements in resolution and the relative reduction in cost of the machines. These have resulted in a much larger number of...
operators using a variety of machines, with a concomitant introduction of substantial interobserver variability. Just observing several operators within a single clinic will emphasize their variable scanning techniques, knowledge regarding image optimization, inclusion criteria for antral follicles (e.g., 2–5 mm or 2–10 mm), and methodology for counting and measuring follicles. Although some attempts to standardize two-dimensional techniques have been made (5), these have been limited when compared with the formal external quality control measures and accreditation that were so successful for nuchal translucency. It is therefore not surprising that recent clinical trials sponsored by pharmaceutical companies have not depended on AFC except for use as an exclusion criterion (6, 7). For example, in this context it is largely irrelevant whether there are 24 or 39 follicles, as both are currently classed as polycystic ovaries and the patient can be confidently excluded.

Accompanying this improved visualization of the ovary, the other major technical advance has been the development of three-dimensional automated follicular tracking (8, 9), which can substantially improve both intraobserver and interindividual variability (10). Although it is attractive conceptually, it is limited to one manufacturer and still requires offline analysis to ensure optimal performance, all of which have limited its widespread adoption. However, it does suggest a future that involves automated image acquisition, centralized quality control, and automated data interrogation incorporating integration of previous scan data. Collectively, this would provide health-care providers with a detailed analysis of the follicular dynamics and endometrial development.

For AMH, there have also been major technical limitations. These include the various existent forms of the assay, including the original research assays, the DSL and Immuno-tech assays, the Beckman Coulter Generation II assay, which combines the cross-species DSL antibodies with the Immuno-tech standards, the new AMH enzyme-linked immunosorbent assay (ELISA), which uses different antibodies, and the fully automated AMH assay that is due to be released by several companies (11). When the variability in the performance characteristics of these assays and the laboratories performing them, the lack of an international standards or an external quality control system, and the necessity for rapid upscaling of manufacturing capabilities are combined with the recent evidence that sample handling can dramatically alter AMH concentrations, it is not surprising that confusion and inconsistency are found in the AMH literature (12, 13). Abnormal batches of calibrators, inappropriate use of linear rather than cubic regression for standard curve interpretation, sample collection in ethylenediaminetetraacetic acid (EDTA) tubes rather than serum tubes, postage of samples before centrifuge, storage at room temperature, and poor operator reproducibility all have now been reported to be associated increases and decreases in serum AMH levels. At present, the manufacturing issues appear to be resolved, but the importance of proper sample handling (with a dramatic ∼40% increase in AMH reported at room temperature if the sample is not centrifuged immediately) remains underappreciated (13). Resolving these issues is not insurmountable, but it requires industry, researchers, and clinical pathology laboratories to provide clear guidance on their preferred assay from the point of venipuncture through to the interpretation of the results in an age- and gender-specific manner.

There are ongoing developments with respect to the measurement of AMH, which again will be subject to lack of standardization. Several groups are trying to quantify AMH within urine, which would allow it to be used as
a point-of-care or home test. At present, the pathways underlying the degradation and clearance of AMH have not been clarified, although consistent with older reports the recent evaluation of its decay kinetics in humans have shown it to follow first-order kinetics; with a mean terminal half-life of 27.6 ± 0.8 hours, complete clearance is expected after 8 days [14, 15]. The technical feasibility of measuring AMH in dried blood spots has also now been demonstrated, with a strong correlation to serum levels [16]. Although this opens up the possibility of measuring AMH in large epidemiologic cohorts with the relevant samples, it also raises the possibility of a variety of tests both home and postal that may use the small volume obtained from a finger-prick stick. Just as for whole blood, where it appears that AMH is not stable and has values that can increase dramatically over the course of a week (in contrast to serum where it is stable at room temperature and over multiple freeze-thaw cycles), care will need to be taken regarding sample handling [17]. Clearly, with all of these ongoing assay developments, urgent agreement on an international standard for AMH is critical.

AMH AND AFC: THE TWO FACES OF THE MOON

Given the multiple technical limitations, it is incredible that AFC or AMH exhibit such strong and similar associations with the size of the primordial follicle pool and follicular recruitment rates [2, 18–20]. The similarity in the strength of relationship reflects that it is the same 2–6 mm follicles that are seen on ultrasound that produce AMH [21]. Consequently, analyses of AFC against AMH have consistently shown strong correlations [22, 23]. There are inevitably outliers and discordance between the AFC and AMH; however, even presuming this is not simply due to the technical aspects outlined previously, this is not to be unexpected. The circulating AMH level will reflect the output of all granulosa cells within the follicles; this production is not consistent and will reflect the size of the follicles, the potential granulosa cell mass, the state of maturation of the granulosa cells, the intrafollicular environment, and the genetics of the individual, in whom with a variety of single-nucleotide polymorphisms (SNPs) may potentially influence AMH production [24]. However, this discordance also points to why we should potentially consider combining both AFC and AMH when we being to think about predicting ovarian response and potentially the reproductive life span.

The initial studies examining the relationship of AFC and AMH with oocyte yield used stimulation strategies such as the long-course agonist, which attained maximal follicular recruitment and were heralded for their beautiful linear relationships [25, 26]. This work has now been replicated for antagonist strategies; although the correlation coefficients are still strong, they are weaker [26]. This, of course, is not surprising given that a major strength of these strategies is the altered follicular recruitment pattern that underlies the inherent reduction in the risk of ovarian hyperstimulation syndrome (OHSS) [27–29]. This means that, when using these strategies, formulas derived for agonist strategies to calculate the likely number of oocytes retrieved for any given AMH concentration will not apply, and new mathematical models will be required. This underlines the concept that for any given AFC or AMH there is a potential oocyte yield, but by altering the stimulation strategy it can be extensively modified.

Given the linear relationships of AFC and AMH with oocyte yield and their strong correlation with each other, one would anticipate that they could predict the extremes of ovarian response and the clinical scenarios of poor response and OHSS. Indeed, several systematic reviews, meta analyses, and individual-patient data meta-analyses have all shown that both AFC and AMH have essentially equivalent performance characteristics for the prediction of both poor and excessive ovarian response [30, 31]. Accounting for their similarities and the technical limitations already discussed, many have suggested that one or the other should be used in preference for response prediction. However, our recent individual-patient data meta-analysis potentially shows the way forward by harnessing their combined strengths [32]. Combining ultrasound and biochemical markers for prediction of outcomes is not new for our speciality; Down syndrome screening is the classic example of where baseline phenotypic data such as maternal age and smoking status combine with an ultrasound marker [nuchal translucency] and biochemical markers [PAPP-A and free β-human chorionic gonadotropin] to provide a prediction model with the best sensitivity and specificity. By combining age, AMH, and AFC, optimal response prediction can be achieved, with a statistically significant increase in the area under the receiver operator curve from 0.81 up to 0.85 if all three are used as compared to just two variables [32].

WHY PREDICT OVARIAN RESPONSE?

Perhaps a unique strength of IVF is our ability to repeat the procedure—with the inherent increase in cumulative live-birth rates [33]. Although the majority of our patients will not get pregnant in the first treatment cycle, with repeated attempts combining multiple fresh and frozen cycles the overall prospect of a successful outcome for many is good. We therefore need to avoid iatrogenic complications, ensure that patients are counselled appropriately, have realistic expectations of the outcome of their ovarian stimulation, and ensure that clinicians choose the optimal stimulation strategy even in that very first treatment cycle. That AFC and AMH can predict ovarian response accurately enables clinicians and thereby patients to be informed about all these critical steps.

For example, at one extreme of the response spectrum we can identify women who are at risk of OHSS and can adjust our stimulation strategy to incorporate gonadotropin-releasing hormone (GnRH) antagonists [28]. We thereby minimize the risk of this potentially fatal complication, but potentially even more importantly we have the ability to completely eliminate it by adopting a GnRH-agonist trigger before oocyte retrieval [34]. This unique approach has particular benefits for women undergoing altruistic oocyte donation, removing much of the integral risk of IVF [35, 36]. Although
Antimüllerian hormone (AMH) stratified individualization of treatment as used by the author. Ovarian response categories dictate risk, and treatment strategies are designed to minimize risk while maximizing oocyte yield within each response category. Negligible response means that the conventional criteria for triggering (three follicles ≥ 17 mm) is unlikely to be achieved. For all antagonist cycles with an excessive response, an agonist trigger is adopted. The AMH measurements are for the AMH Gen II assay, and the values are in pmol/L. The suggested antral follicle count (AFC) thresholds are based on the correlation of AMH and AFC and associated response category literature (30, 65).


Clearly more oocytes could be retrieved if a long-course agonist approach is used, this would be at the cost of patient safety if an excessive response is predicted. Conversely, maximizing follicular recruitment would seem appropriate if a poor response is anticipated.

At present, the value of a mixed-strategy program has yet to be fully elucidated, but certainly for centers where agonist strategies still dominate we have previously shown the advantage of this approach over conventional dose adjustment and long-course agonists for all (37). Our current approach to AMH stratified treatment is depicted in Figure 2. Alternative strategies for all categories of ovarian response, including in particular normal and poor responders have been reported, with the optimal strategy still to be fully elucidated. For anticipated normal responders, we have continued to use agonist-based strategies, due to the similarity of the point estimates but tighter confidence intervals for ongoing pregnancy (28 randomized, controlled trials: odds ratio [OR] 0.87; 95% confidence interval [CI], 0.77–1.00) and live births (9 randomized, controlled trials: OR 0.86; 95% CI, 0.69–1.08) favoring agonist-based rather than antagonist-based strategies (28).

For potential poor responders, we currently use a flare strategy because of its reduced treatment burden and ability to capitalize on endogenous luteinizing hormone (LH) activity, in accordance with recent studies supporting a beneficial role of LH in older women (38). Although with very low AMH levels (negligible category) the risk of not achieving the usual trigger criteria (3 follicles ≥ 17 mm) is high (39), pregnancies are achievable, and we are therefore prepared to stimulate all women rather than excluding them from our program or only offering oocyte donation (37).

The role of adjuvants such as androgen-modulating agents at this end of the ovarian response spectrum remains to be fully clarified (40). Inevitably, clinicians will have their preferred stimulation strategies for each category, and these will continue to evolve pending the development of new therapeutic options and evidence from randomized controlled trials. Pretreatment categorization of anticipated ovarian response before the first treatment cycle has been a critical step forward, and this approach is now being adopted in ongoing randomized controlled trials (41).

The ability to predict a very poor response has resulted in some investigators withholding the first treatment cycle if a very low AMH concentration is detected, with an overall improvement in results of the program and substantial cost savings (42). We have previously argued that this was not appropriate because even women with AMH concentrations at the limit of functional sensitivity have a significant chance of conception through IVF (43). Inevitably, this will be lower than a woman of the same age with a higher ovarian reserve, as both age and ovarian reserve are independent predictors of live birth after IVF (44). To completely withhold treatment and not actually confirm a predicted poor response purely based on an AFC or AMH would seem inappropriate. This is particularly the case as neither have been incorporated into cost-effectiveness models, with other more accurate population level models, such as IVF, predict available and now being used to guide national access criteria (45, 46).

Whether knowing the anticipated oocyte response has a beneficial psychological effect for the couple and thereby reduces cycle dropout has not been formally evaluated. In our own practice, the ovarian assessment report is performed before the initial consultation, and it provides the predicted oocyte response category, including poor or negligible response. This does appear to set patient expectations appropriately, particularly at the bottom end of the spectrum where only a few oocytes may be retrieved. Given that many women do not fully appreciate the detrimental effect of age on oocyte number, the ability to guide them on their overall success by use of a combination of their age as a surrogate for oocyte quality as well as AMH and AFC to indicate oocyte yield has been a powerful tool (44). It is likely in the future that with standardization of AMH and AFC measurement and stimulation strategies, multivariate prediction models with tight confidence intervals will be able to be created for generating customized, individualized reports. Steps on this path have already been made for dose adjustment, using a combination of AMH, FSH, age, smoking status, and body mass index (47, 48).

A FUTURE FOR OVARIAN RESPONSE BIOMARKERS BEYOND ART

So although the future for blood biomarkers is bright, potentially the greatest impact will be not in the IVF clinic but among other health-care providers. For years, nonreproductive medicine specialists have been measuring FSH to assess...
ovarian function, and it is a standard first-line investigation by primary care physicians and gynecologists for women with menstrual irregularities. The dominance of FSH is in part due to our familiarity with its biological pathway, which we learn as undergraduates, as well as its ease of measurement on an automated platform and consequently its price. However, it clearly has a much lower correlation with primordial follicle counts and follicular recruitment rates and has limited ability to diagnose ovarian dysfunction, including polycystic ovarian syndrome (PCOS). As an alternative blood test, AMH would appear to overcome these shortcomings, particularly once automation and competition make it as available, routine, and cheap as FSH measurement currently is. This critical step will open up AMH to all health-care providers, and will drive a wide range of new and novel uses in a future where AMH measurement dominates (Fig. 3).

For example, for a woman attending her primary care physician or gynecologist with menstrual cycle irregularity, assessment of AMH will enable stratification into possible diagnostic categories (49). PCOS would be likely if the AMH level is elevated, with values >35 pmol/L consistent with polycystic ovary morphology (50). Although AFC will still currently be required to meet the diagnostic criteria, the suggestion of a likely diagnosis before transvaginal scanning is a huge step for the physicians for whom the scan is not available. At the other extreme, low levels were initially reported as being due to premature ovarian insufficiency, but it is now appreciated that in conditions in which there is impaired follicular recruitment (such as prolonged suppression with GnRH agonists, pregnancy, or hypogonadotrophic hypogonadism) AMH can also occasionally be low (51, 52). Clarification of these relative extremes of ovarian reserve can then be achieved by measuring FSH.

For the woman about to embark on ovarian surgery assessment, preoperative and postoperative AMH measurements will provide accurate estimates of the impact of the surgery on her ovarian reserve. A recent meta-analysis quantified the impact of endometrioma removal as a reduction in AMH of 1.5 ng/mL, which is equivalent to a more than 10-year age-related decline (53). For industrial partners, it will allow assessment of new technologies and their impact on the ovarian reserve, as surgical techniques are increasingly recognized as an important determinant in the detrimental impact of surgery on primordial follicle count.

For the young woman with cancer, considering her future fertility is part of “survivorship” and life beyond cancer. Although discussing fertility preservation is now routine for oncologists, recognition that AMH can better predict than age which women will become infertile and have chemotherapy-related amenorrhoea is a step forward (54). Quantification of pretreatment ovarian reserve will facilitate counseling those women whose commencement of chemotherapy must be delayed to allow oocyte/embryo storage as they already have a critically low ovarian reserve. Even in prepubertal girls with cancer, AMH has been shown to be able to predict posttreatment chemotherapy-related amenorrhoea (55). For oncology trials, AMH will become an increasingly important secondary outcome measure that allows accurate quantification of the detrimental impact of specific strategies on the ovarian reserve.

Finally, the greatest step forward may be in the accurate prediction of the reproductive life span. Although AMH becomes undetectable approximately 5 years in advance of the menopause (56), the more striking revelation is the ability to predict the onset of the menopause (57, 58). In several prospective and retrospective cohorts, it has been shown that young women with a low age-specific AMH have a reduced reproductive life span and an earlier menopause. It is yet unclear as to whether this will also translate to reduced fertility as suggested by te Velde and Pearson (59), where a decade of subfertility and sterility precedes menopause; only a single study has suggested that women with low AMH have reduced fecundity (60). This study has yet to be replicated, with another with 6 months of follow-up observation suggesting no effect on fecundity (61). A positive association between AMH and fecundity would potentially imply that young women with a low AMH levels irrespective of regular ovulatory cycles have a lower oocyte quality; which would potentially underlie their reduced fecundity. At present, the evidence for AMH being associated with oocyte quality is not consistent, and large data sets with prolonged follow-up will be required to confirm an independent effect beyond that of age, particularly in regards to the time to natural pregnancy. Although the rate of decline at a population level has been modeled (62–64), any given women may cross centiles, and thus repeated measurements would be required. This could be integrated with other health screening programs. Collectively the ability to identify a woman who will have a shorter reproductive life span will allow the redefining of
family planning. It will no longer be associated with solely contraception but rather with allowing women to plan effectively when to have a family in the context of their professional and reproductive life span.

CONCLUSION

Our current ability to predict ovarian response has been transformed in recent years with the recognition of the strong linear relationships of AMH and AFC with ovarian reserve. The automation of the AMH assay will be a major step forward and allow rapid access to estimation of the ovarian reserve to all health-care providers. Inevitably, ovarian biomarkers will continue to develop dramatically, with the integration of novel discovery platforms and computational medicine. However, while we await these developments, we can ensure that we harness the collective power of the currently available biomarkers to ensure optimal patient care and true personalization of ovarian stimulation.

Acknowledgments: The author thanks Antonio La Marca for helpful discussions regarding the content of the manuscript.

REFERENCES


