Cost-effective analysis of oocyte cryopreservation: stunning similarities but differences remain

Dear Sir,

Recently, our two groups independently used decision analysis to estimate the cost effectiveness of oocyte cryopreservation for social reasons (van Loendersloot et al., 2011; Hirshfeld-Cytron et al., 2012). Although we both employed similar decision analysis techniques, striking differences were obtained in the results, with the US-based study of Hirshfeld-Cytron et al. noting that oocyte cryopreservation cost an additional $135 520 per live birth, while the European study of Van Loendersloot et al. noted that if 61% of women utilized their frozen gametes, oocyte cryopreservation was cost-effective at an additional $24 600 per live birth.

A detailed review of our methods reveals several differences that contributed to the remarkable differences in our findings. For example, Hirshfeld-Cytron et al.’s model was based on a 25-year-old woman who undergoes one cycle of oocyte preservation and who, attempting to conceive for 6 months at age 40, then uses the cryopreserved oocytes. If still unsuccessful, she then turns to IVF for four cycles. In contrast, Van Loendersloot et al. had a 35-year-old woman undergo three oocyte preservation cycles and then use these directly at age 40. If still unsuccessful, then the next strategy was natural conception for 3 years. This group also incorporated miscarriage costs into the analysis. Cost estimates were different between the two studies, the costs of IVF and frozen embryo transfer being much lower in the European study. Probabilities of success with cryopreserved oocytes (36% live birth per cycle compared with 25% live birth rate per cycle) and IVF (38% after three IVF cycles compared with 30% after four IVF cycles) at age 40 also were more optimistic in the European analysis.

These differences in clinical scenarios, cost and probability estimates likely reflect practice differences and interpretation of data that are available in the literature. Despite differences in the results of our studies, we both agree on the need to begin to study these clinical scenarios before drawing conclusions with regard to the cost efficacy of the technique. The differences in our results also reflect the need for larger collaborative studies reporting on oocyte cryopreservation success over a wide range of ages with IVF costs. Ultimately, new technology needs to be investigated before implemented. Likewise, accurate estimates of oocyte cryopreservation success and cost are imperative for societal allocation of resources and patient consultation. We believe that continued decision analysis on the incorporation of assisted reproduction techniques, unlikely and difficult to study in randomized control fashion, is imperative to continue to move the field forward.

References


J. Hirshfeld-Cytron1,2,*, L.L. van Loendersloot3, B.W. Mol3, M. Goddijn3, W.A. Grobman2, L.M. Moolenaar3 and M.P. Milad2

1Division of Reproductive Endocrinology and Infertility, Department of Obstetrics and Gynecology, University of Illinois Medical Center, 820 S Wood Street MC 808, Chicago, IL 60612, USA
2Department of Obstetrics and Gynecology, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA
3Center for Reproductive Medicine, Department of Obstetrics and Gynaecology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands

*Correspondence address. Tel: +1-312-355-2740; E-mail: hirshfel@uic.edu

doi:10.1093/humrep/des339

Advanced Access publication on September 27, 2012

Reproductibility of AMH

Sir,

We read with interest the paper by Rustamov et al. (2012), which assesses the performance of the AMH DSL and Gen II assays and questions the validity of previous papers (Kumar et al., 2010; Wallace et al., 2011). In accordance with National Committee for Clinical Laboratory Standards Guidelines, the correct way to compare and validate the AMH assays is by simultaneous testing of the two methods in the same serum samples as previously reported (Kumar et al., 2010; Wallace et al., 2011). The study by Wallace et al. was carried out in three separate centres and clearly demonstrated how robust the test is when conducted in well-controlled environments. This result was then confirmed by a further independent laboratory (The Doctors Laboratory) with these additional results reported in the Beckman Coulter Technical Bulletin UK2010-MIB-004. Comparing two sequential populations whose constitution may have changed over time and draw conclusions regarding the relative concentration values measured by tests performed at different times by different personnel possibly under different conditions is not an
The AMH assays have progressed through a series of evolutionary steps, with the recent Gen II assay designed to use the calibrators (or standard curve) of the previously available assay marketed by Immunotech (IOT, France), and the antibodies used were those previously used in the DSL (Texas, USA) kit (Nelson and La Marca, 2011). Previous comparative analysis of the performance of the DSL and IOT assays demonstrated that DSL values were ~40% lower than that observed for the Immunotech (Taieb et al., 2008). Correspondingly, it was not a surprise when the paper by Wallace et al. confirmed that the values produced by the Gen II kit were similar to the IOT kit.

To confirm that the Gen II and IOT assays have not changed since these tests were carried out, we took a random sample of 15 recent reports from the UK National External Quality Assessment Service analyses of serum. In this series, each value is constituted from >50 laboratories for Gen II, and six labs for the IOT. The mean value for these tests was 21.4 pmol/l in the Gen II series and 21.0 pmol/l in the IOT series. This clearly demonstrates that both tests maintain their strong agreement, and that the Gen II assay sits exactly where it would be anticipated. It should be added that the variance of the Gen II was lower than that for IOT, indicating its simple reliability.

To assess repeat assay performance with the Gen II assay we recently undertook a series of sample re-evaluations in women <40 years of age, whose AMH value was recorded as below the 25th centile. This group was specifically chosen due to the emotive nature of a low AMH value for an individual woman in her 30s. The samples were stored at 4°C for 7 days prior to repeat assay. Table I shows the results of two evaluations of 51 samples, with repeat tests effected following 1 week of refrigerated storage. There was strong concordance between the two values despite operating at the lower limit of sensitivity. We have previously suggested that AMH can be used to stratify anticipated response and of the 51 samples, 50 were initially categorized as 'low' response prediction in an IVF setting. The repeat tests, operating close to the lower limit of sensitivity, showed 48 (96%) were categorized at the same level. Collectively, this suggests satisfactory assay performance towards the lower concentration limits.

In discussing the different results determined by the authors and the previously published comparative tests, the authors refer to a concept of ‘degradation of the specimen in one or both of the assays’. There may be some confusion here between assay performance and sample preparation. It has been known for many years that the AMH molecule undergoes proteolytic modification, which is why the more modern assays were developed, with antibodies targeted at the ‘mature’ region of the molecule. These assays have been well validated for use in whole blood, when the serum is maintained in contact with blood cells (Fleming et al., submitted for publication). The effect can be completely avoided by appropriate processing of the sample.

Lastly the authors describe a dilution test which suggests that the results produced by the Gen II kit may be unreliable. We have performed such a test with samples previously showing ‘high’ concentrations of AMH in them (n = 10), and using (i) the supplied assay diluent and also (ii) blank serum as diluent. Table II shows the estimated concentrations determined in the different circumstances. The duplicate measurements during this evaluation showed a variance of 8 and 5% in the two circumstances. These results demonstrate a good linearity and reliability of the test, as initially reported (Kumar et al., 2010).

Collectively, the previous published work and the evidence presented here question the validity of the conclusions presented by Rustamov et al., and, although care must be taken with sample preparation (Fleming et al., submitted for publication), strongly supports the use and clinical application of the new Gen II test for AMH.

### References


Reply: Reproducibility of AMH

Sir,

We are grateful for the opportunity to respond to the comments in the letter from Drs Fleming and Nelson concerning our recent paper (Rustamov et al., 2012). It was our intention in publishing to stimulate discussion and further research in this area; however, we consider that they appear to have misunderstood several points about our study.

The authors adopt a very narrow definition of assay reliability and we would not dispute that this assay performs adequately on the sample in which these tests have been performed. Fleming and Nelson point out that the correct way to validate a new assay is to simultaneously compare results in the same sample—this is exactly what we did when the new Gen II assay was introduced and we obtained results which were in agreement with other published comparisons. The issue is not about how two assays may agree in measuring AMH in a single sample, but about what has happened to the AMH in that sample prior to analysis. We presented evidence that, in routine clinical samples processed strictly within the protocols specified by the manufacturer, more variability is found. We speculated that this may be due to some aspects of sample preparation, about which Fleming and Nelson are in agreement. It does seem that the various AMH assays differ in their sensitivity to these factors in ways which are not understood. Moreover, we have demonstrated in a series of real clinical samples (which again we stress were prepared strictly in accordance with the assay manufacturers protocols) that a basic property of any assay, linearity to dilution, is violated.

These differences may well be due to pre-analytical aspects of the sampling process. The work described by Fleming and Nelson appears to have been carried out in stored samples while our own work was performed either on fresh samples or serum which had usually been frozen for only a few days, the assays being performed in the same lab by the same staff as previous assays. An earlier publication (Rey et al., 1993) showed that long-term storage of serum at −20°C caused an approximate doubling of the AMH concentration compared with fresh samples, an effect that could be prevented by the addition of proteolytic inhibitors. Our work is consistent with an interpretation that proteolytic or conformational change in the AMH dimer occurs prior to analysis. Differences in measured AMH levels seen in samples stored under different conditions may reflect the extent of this change prior to analysis; if the process has gone to completion, the samples will appear to be stable; if ‘activation’ has not occurred at all, then results may appear to double under assay conditions or dilution. Most samples will lie somewhere between these extremes as we found in the 7-day room temperature storage experiment.

We have been unable to find comprehensive reports in the literature concerning AMH assay validation and details of the exact methodology employed when handling blood samples from patients are lacking. For example Wallace et al., 2011, quoted by Fleming and Scott, made no mention whatsoever of how serum samples were processed or stored. Linearity studies were performed using serum pools containing known concentrations of AMH (by definition already assayed for AMH, i.e. not fresh serum) which had been diluted using serum from post-menopausal women which had no detectable AMH (presumably also confirmed by assay). None of this material is ‘fresh’ serum, so presumably the ‘phenomena’ which lead to a higher AMH reading will already have occurred and the measurable AMH levels will be stable in these pools if the process has gone to completion. If the change has already occurred in their pooled samples, they will not be able to reproduce our results.

Fleming and Nelson provide additional data, which reassuringly confirm our observations and which we look forward to seeing fully published. They show a significant 27% increase in measured AMH over 7 days at 4°C, compared with our 58% in separated serum samples stored at room temperature for the same period. Their dilution data also shows an average increase in recovery of 20–30% on dilution with diluent or serum which although not as dramatic at the 57% seen in our samples is nonetheless consistent with the effects we observed. It will be interesting to see more detailed data here, as we noted that the non-linearity on dilution does seem to be sample dependent. Larger numbers of better-characterized samples are needed to understand this phenomenon.

We hope that this evidence might help to identify further experiments, which will lead to an understanding of the cause of this variability and ultimately lead to an assay which fulfils the promise that AMH measurement offers.

References
