The Measurement of Anti-Müllerian Hormone: A Critical Appraisal

Oybek Rustamov, Alexander Smith, Stephen A. Roberts, Allen P. Yates, Cheryl Fitzgerald, Monica Krishnan, Luciano G. Nardo, and Philip W. Pemberton

Department of Reproductive Medicine (O.R., C.F.), St Mary's Hospital, Central Manchester University Hospital National Health Service (NHS) Foundation Trust, Manchester Academic Health Science Centre (MAHSC), Manchester M13 0JH, United Kingdom; Department of Clinical Biochemistry (A.S., A.P.Y., P.W.P.), Central Manchester University Hospitals NHS Foundation Trust, Manchester M13 9WL, United Kingdom; Centre for Biostatistics (S.A.R.), Institute of Population Health, MAHSC, University of Manchester, Manchester M13 9PL, United Kingdom; Manchester Royal Infirmary (M.K.), Central Manchester University Hospitals NHS Foundation Trust, Manchester M13 9WL, United Kingdom; and Reproductive Medicine and Gynecology Unit (L.G.N.), GyneHealth, Manchester M3 4DN, United Kingdom

Context: Measurement of anti-Müllerian hormone (AMH) is perceived as reliable, but the literature reveals discrepancies in reported within-subject variability and between-method conversion factors. Recent studies suggest that AMH may be prone to preanalytical instability. We therefore examined the published evidence on the performance of current and historic AMH assays in terms of the assessment of sample stability, within-patient variability, and comparability of the assay methods.

Evidence Acquisition: We reviewed studies (manuscripts or abstracts) measuring AMH, published in peer-reviewed journals between January 1, 1990, and August 1, 2013, using appropriate PubMed/Medline searches.

Evidence Synthesis: AMH levels in specimens left at room temperature for varying periods increased by 20% in one study and by almost 60% in another, depending on duration and the AMH assay used. Even at −20°C, increased AMH concentrations were observed. An increase over expected values of 20–30% or 57%, respectively, was observed after 2-fold dilution in two linearity-of-dilution studies, but not in others. Several studies investigating within-cycle variability of AMH reported conflicting results, although most studies suggest that variability of AMH within the menstrual cycle appears to be small. However, between-sample variability without regard to menstrual cycle as well as within-sample variation appears to be higher using the GenII AMH assay than with previous assays, a fact now conceded by the kit manufacturer. Studies comparing first-generation AMH assays with each other and with the GenII assay reported widely varying differences.

Conclusions: AMH may exhibit assay-specific preanalytical instability. Robust protocols for the development and validation of commercial AMH assays are required. **(*J Clin Endocrinol Metab* 99: 723–732, 2014)**

In the female, anti-Müllerian hormone (AMH), produced by granulosa cells of preantral and early antral ovarian follicles, regulates oocyte recruitment and folliculogenesis (1, 2). It can assess ovarian reserve (3–5) and guide gonadotropin stimulation in assisted reproduction technology (ART) (6). AMH is also used as a granulosa cell tumor marker, a marker of ovarian reserve after chemotherapy (7, 8), and to predict age at menopause (9, 10).

ISSN Print 0021-972X ISSN Online 1945-7197
Printed in U.S.A.

Received September 12, 2013. Accepted December 3, 2013.
First Published Online December 17, 2013

Abbreviations: AMH, anti-Müllerian hormone; ART, assisted reproduction technology; CI, confidence interval; ICC, intraclass correlation.

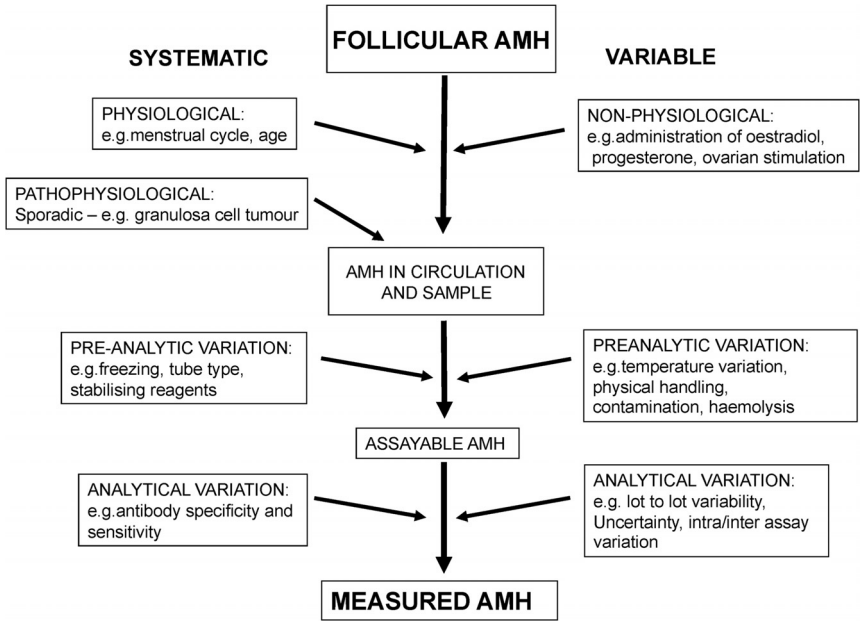

Figure 1. Biological and analytical variability of AMH.

AMH immunoassays, first developed by Hudson et al (11) in 1990, were introduced commercially by Diagnostic Systems Laboratories (DSL) and Immunotech (IOT). These assays were integrated into a second-generation AMH assay, GenII (12), by Beckman-Coulter, but recent work suggests that this new assay exhibits clinically important, within-patient, sample variability (13–15). Beckman Coulter recently confirmed this with a field safety notice (FSN 20434-3); they cite, without showing evidence for, complement interference as the problem.

"True" AMH variability comprises both biological and analytical components (Figure 1), and given the varying antibody specificity and sensitivity of different AMH assays, then logically different kits will respond to these components to varying degrees. This review considers the published literature on AMH measurement using previous and currently available assays. Potential sources of variation and their contribution to observed AMH variability were identified.

Review Structure

This review has been divided into logical subgroups. We first address the stability of AMH at different storage temperatures, then the effects of freeze/thaw cycles, and finally AMH variability in dilution studies. Secondly, the within-person variability of AMH measurement is considered, encompassing intra- and intermenstrual cycle variability and repeat sample variability in general. The final section covers AMH method comparisons, comparing older methods to each other and to the newer, now prevalent GenII method, and finishing with data on published guidance ranges concerning the use of AMH in ART. A general summary concludes the paper.

Systematic Review

The terms "anti-Müllerian hormone," "AMH," "Müllerian inhibiting substance," and "MIS" were used to search the PubMed/Medline MeSH database between January 1, 1990, and August 1, 2013, for publications in English commenting on AMH sample stability, biological and sample-to-sample variability, or assay method comparison in human clinical or healthy volunteer samples. Titles and/or abstracts of 1653 articles were screened to yield the following eligible publications: 10 stability studies, 17 intra/intercycle variability studies, and 14 assay method comparability studies.

Sample Stability

Recent work has established that GenII-measured AMH is susceptible to significant preanalytical variability (13, 14), not previously acknowledged, which may have influenced results in previous studies with this assay.

Stability of unfrozen samples

Five studies examined AMH stability in samples stored either at room or fridge temperature (Table 1) (13, 16–19). Al-Qahtani et al (16), assessing the precursor of the DSL ELISA, reported that "immunoreactivity survived the storage of samples unfrozen for 4 days," but they did not record storage temperature or sample numbers. Evaluating the GenII assay, Kumar et al (18) stored 10 samples at 2 to 8°C for up to 1 week and found an average 4% variation compared to samples analyzed immediately. However, their specimens, originally reported as "fresh," appear to have been kept cool and transported overnight. Fleming and Nelson (19) reported no significant change in the GenII-assayed AMH from 51 samples stored at 4°C. Methodological information was limited, but interrogation of their data by Rustamov et al (14) suggested that AMH levels rose by an average of 27% after 7 days of storage. Zhao et al (17) reported a difference of less than 20% between DSL-assayed AMH in seven serum samples

Table 1. AMH Assay Validation: Effect of Sample Storage Conditions, Fresh/Thaw Cycles, and Linearity of Dilution

First author (Ref.)	Assay	Method	Result
Rey (21)	In-house	Effect of long-term storage at −20°C (n = 4)	AMH levels in archival samples were 230% higher than original value
Long (22)	IOT	Linearity up to 16-fold dilution (n = 3)	Observed AMH was 84–105% of expected AMH
Al-Qahtani (16)	In-house	a. Freeze/thaw stability; storage unfrozen for 4 d; b. linearity up to 32-fold dilution (n = 6)	a. Immunoreactivity survived both multiple freeze-thaw cycles and storage unfrozen for 4 d; b. dilution curves were parallel to the standard curve
Zhao (17)	DSL	Serum frozen immediately at −20°C compared to aliquots stored at 4°C or 22°C for up to 2d (n = 7)	AMH levels increased by 1% at 4°C and 9% at 22°C after 2 d compared to sample frozen immediately
Kumar (18)	GenII	a. Serum or plasma stored at 2 to 8°C or −20°C for up to 7d (n = 20); b. Serum or plasma underwent up to three freeze/thaw cycles (n = 20); c. Linearity of dilution (n = 4)	a. AMH levels were stable for up to 7 d at 2 to 8°C or −20°C; b. AMH increased by 15% in serum and by 5% in plasma after three cycles; c. Linear results obtained across the dynamic range of the assay
Preissner (23)	GenII	Linearity of dilution (n = 7)	Average agreement with expected result was 97%
Rustamov (13)	GenII	a. Stability at RT for up to 7d (n = 48); b. Storage for 5 d at −20°C or −80°C compared to fresh sample (n = 8); c. Linearity on 2-fold dilution (n = 9)	a. AMH levels increased by an average of 58% over 7 d; b. AMH levels increased by 23% at −20°C but were unchanged at −80°C; c. AMH levels were on average 157% higher than expected
Fleming (19)	GenII	a. Serum stored at 4°C for 7d (n = 51); b. Linearity of dilution (n = 10)	a. AMH levels increased by an average of 27%; b. AMH was 28 and 33% higher on 2-fold and 4-fold dilutions, respectively
Fleming (20)	GenII	a. Whole blood stored for up to 90 h at 4°C (n = 32) or 20°C (n = 21); b. Serum stored for 5 d at 20°C and 2 d at 4°C (n = 13)	a. AMH increased by 11% at 4°C and by 31% at 20°C; b. only 1% increase in AMH compared to original value
Han (15)	GenII	Serum from nonpregnant (n = 13) or early pregnant (n = 7) women stored at RT, −20°C, or −80°C for up to 7 d	In nonpregnant women, AMH increased by 26% after 7 d at RT, but was unchanged at −20°C or −80°C. In pregnant women, AMH increased by 50% at RT and 27% at −80°C after 48 h.

Abbreviation: RT, room temperature.

kept at 22°C for 48 hours when compared to aliquots from the same samples frozen immediately at −20°C. Rustamov et al (13) measured AMH (GenII) daily in 48 serum samples at room temperature for 7 days and observed an average 58% increase (from 0 to > 200%), whereas others (20) reported a 31% mean rise in GenII-assayed AMH in whole blood after 90 hours at 20°C, whereas serum AMH was virtually unchanged after prolonged storage at 20°C.

Sample stability at −20°C or −80°C and the effects of freeze/thaw

Rey et al (21) reported a significant increase in AMH (in-house assay) in samples stored at −20°C for a few weeks, attributing this to proteolysis that could be stabilized with protease inhibitor. Kumar et al (18) saw 6% variation between GenII-assayed AMH levels from 10 fresh and 10 frozen samples, whereas Rustamov et al (13) observed a 22% increase in AMH (GenII) on reanalysis of eight serum samples after 5-day storage at −20°C. These authors saw no AMH increase in serum stored at −80°C for the same period.

Linearity of dilution

Six studies examined linearity of dilution on observed AMH concentrations. Long et al (22) recovered between 84 and 105% of the expected AMH concentration (IOT, n = 3) and AMH dilution curves parallel to the standard curve were reported by Al-Qahtani (16). Kumar et al (18) (n = 4), and Preissner et al (23) (n = 7), reported GenII-assayed AMH recoveries from 95% to 104% and 97%, respectively. Sample handling information was limited in some of these studies (16, 23). Fleming and Nelson (19) (GenII, n = 10) reported variances of 8% using assay diluent and 5% using AMH-free serum after 2-fold dilution; however, interrogation of their data reveals an apparent dilutional AMH increase of 20–30% in samples stored

before dilution and analysis. In freshly collected serum, Rustamov et al (13) (GenII, n = 9) observed an average 57% increase in apparent AMH concentration after 2-fold dilution, but with considerable variation.

Discussion: sample stability

Sample stability can be a major analytical problem, and detailed examination suggests that previous evidence, stating that commercially measured AMH is stable in storage and exhibits linearity of dilution (12, 16, 18, 19), is weak or conflicting.

No study looking at room temperature storage on IOT-assayed AMH was found, and only one was found using DSL-assayed AMH, which showed an increase of less than 20% during storage (17). Studies using the GenII assay to investigate the effect of storage on AMH variability at room temperature, in the fridge, and at −20°C reach differing conclusions, ranging from stable to an average 58% increase in measured levels. It is important to note here that sample preparation and storage before these experiments was different and could account for the observed discrepancies. The most stable storage temperature for AMH in serum appears to be −80°C (13, 16).

Linearity of dilution studies was also conflicting (13, 18, 19, 23); those reporting good linearity used samples transported or stored before baseline analysis, whereas dilution of fresh samples showed poor linearity. In late 2012, Beckman Coulter accepted that the GenII assay did not exhibit linear dilution and issued a warning on kits that samples should not be diluted. They now suggest that with the newly introduced premixing protocol, dilution should not be a problem.

This review highlights the fact that assumptions about AMH stability in serum were based on a limited number of small studies, often providing limited methodological detail (impairing detailed assessment and comparison with other studies) using samples stored or transported under unreported conditions. Furthermore, conclusions derived using one particular AMH assay have been applied to other commercial assays without independent validation.

The available data suggest that dilution of samples and/or storage or transport in suboptimal conditions can lead to an increase in apparent AMH concentration. The conditions under which this occurs in each particular AMH assay are not yet clear, and more work is required to understand the underlying mechanisms. Two alternative hypotheses have been proposed: first, that AMH may undergo proteolytic change as postulated by Rey et al (21) or conformational change as proposed by Rustamov et al (13, 14) during storage, resulting in "stabilization" of the molecule in a more immunoreactive form; second, Beckman has postulated the presence of an interferent (complement), which degrades on storage (Beckman Coulter field safety notice FSN 20434-3).

A recent case report found that a falsely high AMH level was corrected by the use of heterophilic antibody blocking tubes (24), but this does not explain elevation of AMH on storage (13).

Whatever the mechanism responsible, two solutions are available: either inhibit the process completely, or force it to completion before analysis. Rustamov et al (13) and Han et al (15) both suggest predilution of samples to force the process, a protocol now adopted by Beckman Coulter in their revised GenII assay protocol. Any solution must be robustly and independently validated both experimentally and clinically before introduction in clinical practice. Fresh optimal ranges for interpretation of AMH levels in ART will be needed, and the validity of studies carried out using unreported storage conditions may have to be re-evaluated.

Within-Person Variability

The biological components of AMH variability, such as circadian and inter/intracycle variability have been extensively studied (Table 2 and Supplemental Table 1, published on The Endocrine Society's Journals Online web site at http://jcem.endojournals.org).

Circadian variation

Bungum et al (25) evaluated circadian variability, measuring AMH (IOT) every 2 hours over 24 hours, within days 2–6 of the menstrual cycle in younger (20–30 years old) and older (35–45 years old) women. Within-individual coefficients of variation of 23% (range, 10–230%) in the younger group and 68% (range, 17–147%) in the older group were observed.

Variability within the menstrual cycle

Cook et al (26) observed significant (12%) variation in mean AMH (in-house) levels in 20 healthy women throughout different phases of the menstrual cycle. Intracycle variability of IOT-assayed AMH was reported in three publications (27–29). In two, sequential samples were stored at −20°C until analysis (27, 28). Streuli et al (29) did not report on storage. La Marca et al (27) saw no difference in mean follicular phase AMH levels (d 2, 4, and 6) in untreated, spontaneous menstrual cycles from 24 women. This group went on to report a small, insignificant change (14%) in within-group AMH variability throughout the whole menstrual cycle in 12 healthy women. However, this analysis does not appear to allow for correlations within same-patient samples. Streuli et al (29) studied intracycle variation of

Table 2. Intracycle Variability of AMH

First Author (Ref.)	Subjects	Cycles (a), Day Sampled (b)	Assay	Storage (a), Freeze/Thaw (b), Measurement (c)	Result	Authors' Conclusion
Cook (26)	Healthy, age 22–35 y, regular cycle (n = 20)	a. 1 Cycle; b. d 2/3, LH surge, LH surge + 7 d	In-house	a. −80°C; b. once; c. interassay variation eliminated	d 3: AMH = 1.4 ± 0.9 ng/mL; mid cycle: AMH = 1.7 ± 1.1 ng/mL; mid luteal: AMH = 1.4 ± 0.9 ng/mL.	Fluctuations significant ($P < .008$). AMH may have a regulatory role in folliculogenesis.
La Marca (27)	Healthy, age 21–36 y, regular cycle (n = 24)	a. Follicular phase; b. alternate days	IOT	a. −20°C; b. once	AMH did not change from d 2 to 6 in spontaneous cycles but decreased progressively in FSH-treated cycles	AMH levels did not change significantly during follicular phase of the menstrual cycle.
La Marca (28)	Healthy, age 18–24 y, regular cycle (n = 12)	a. 1 Cycle; b. alternate days, d 0 = day of LH surge	IOT	a. −20°C; b. once	Low mean AMH = 3.4 ± 1.1 ng/mL (d 14); high mean AMH = 3.9 ± 1.3 ng/mL (d 12).	AMH levels did not change significantly throughout menstrual cycle.
Lahlou (31)	Placebo-treated (n = 12)	a. 1 Cycle; b. every 3 d	DSL	NR	7 d before LH surge: AMH = 26 ± 3.2 pmol/L; peak: AMH = 19.1 ± 3.5 pmol/L; 10 d after LH surge: AMH = 25.4 ± 4.3 pmol/L.	AMH levels exhibited a diphasic pattern with levels declining significantly ($P < .05$) during LH surge.
Hehenkamp (30)	Healthy, fertile, regular cycle (n = 44)	a. 2 Cycles; b. AMH measured at each of 7 cycle phases	DSL	a. −20°C	a. Sine pattern fitted to AMH data was not significant ($P = .40$); b. 72% repeat AMH values fell within the same quintile, 28% in adjacent quintile.	AMH shows no consistent fluctuation through cycle compared to FSH, LH, E2.
van Disseldorp (10)	Data from Hehenkamp (30)				Intracycle within-subject variation of AMH was only 13% compared to 31–34% for AFC (dependent on follicle size).	AMH displays less intracycle variability than AFC.
Overbeek (37)	Data from Hehenkamp (30)				Fluctuations were larger than 0.5 μg/L in one cycle in significantly ($P = .001$) more women in the younger group than the older one.	AMH can fluctuate substantially in younger women during menstrual cycle so a single measurement could be unreliable.
Tsepelidis (32)	Healthy, age 18–35 y, regular cycles (n = 20)	a. 1 Cycle; b. d 3, 7, 10–16, 18, 21 and 25	DSL	a. −20°C; b. once	Within-cycle differences not significant ($P = .408$).	AMH levels do not vary during the menstrual cycle.
Wunder (33)	Healthy, age 20–32 y, regular cycles (n = 36)	a. 1 Cycle; b. alternate days	DSL	a. −80°C	AMH levels were statistically higher in the late follicular phase than at the time of ovulation ($P = .019$) or in the early luteal phases ($P < .0001$).	AMH levels vary significantly during the menstrual cycle.
Streuli (29)	Healthy, mean age = 24.1 y, regular cycles (n = 10)	a. 1 Cycle; b. before (LH, −10, −5, −2, −1) and after LH surge (LH, +1, +2, +10)	IOT	a. −18°C	AMH levels were statistically lower during the early luteal phase compared to early follicular phase ($P = .016$) and late luteal phase levels ($P = .02$).	In clinical practice, AMH can be measured at any time during the menstrual cycle.
Sowers (35)	Healthy, age 30–40 y, regular cycles (n = 20)	a. 1 Cycle; b. daily	DSL	a. −80°C; b. once; c. simultaneous	Higher AMH levels with significant variation between d 2–7 in the "younger ovary." Low AMH levels with little variation in the "aging ovary."	AMH varies across the menstrual cycle in the "younger ovary."
Robertson (36)	a. Age 21–35 y, regular cycles (n = 43); b. Age 45–55 y, variable cycles (n = 18)	a. 1 Cycle + initial stages of succeeding cycle; b. three times weekly	DSL	NR	No intracycle variation in AMH level was found in women in mid-reproductive life or in 33% of women with regular cycles in late reproductive age. In the remaining cycles, there was a significant ($P < .01$) 2-fold decrease in AMH in 11 cycles and a significant ($P < .01$) 4.2-fold increase between the follicular and luteal phases.	When AMH levels are substantially reduced, they become less reliable markers of ovarian reserve.
Hadlow (40)	age: 29–43 regular cycles non-PCOS (n = 12)	a. 1 Cycle; b. 5–9 samples per subject	GenII	a. −20°C within 4 h of sampling; b. once; c. simultaneous	7 of 12 women could be reclassified depending on when AMH was measured during the cycle; 2 of 12 crossed cutoffs predicting hyperstimulation.	AMH cycles varied during menstrual cycle and clinical classification of the ovarian response was altered.

Abbreviations: NR, not recorded; PCOS, polycystic ovary syndrome; AFC, antral follicle count; E2, estradiol.

AMH throughout two menstrual cycles in 10 healthy women and also reported no significant changes (<5%).

The DSL assay was used in eight studies assessing intracycle variability (30–37). Four studied sample storage at −20°C (30, 32, 34, 37), and two studied sample storage at −80°C (33, 35). No sample storage data were given in two publications (31, 36). Hehenkamp et al (30) assessed within-subject variation of AMH in 44 healthy women throughout two consecutive menstrual cycles and reported an intracycle variation of 17.4%. Lahlou et al (31)

reported a "diphasic" pattern of AMH, with a significant decrease in levels during the LH surge from 10 women at various cycle phases. Tsepelidis et al (32) reported a mean intracycle coefficient of variation of 14%, comparing group mean AMH levels in 20 women during various stages of the menstrual cycle. Wunder et al (33) reported an intracycle variability of around 30% in 36 healthy women, sampling on alternate days. They saw a marked fall around ovulation, which might have been missed with less frequent sampling intervals, as in other studies. Sowers et al (35) studied within-cycle variability in 20 healthy women but did not compute an overall estimate; instead, they selected subgroups of low and high AMH and reported significant within-cycle variability for women with high AMH but not those with low AMH—an analysis that has been questioned (38, 39). Robertson et al (36) subgrouped mean AMH levels in 61 women, observing that AMH levels were stable in women of reproductive age and ovulatory women in late reproductive age, whereas AMH in other women in late reproductive age was much more variable. Using the data from Hehenkamp et al (30), van Disseldorp et al (34) calculated intraclass correlation (ICC) and reported a within-cycle variability of 13%, although this was not clearly defined. Using the same data, Overbeek et al (37) analyzed the absolute intraindividual difference in younger (≤38 years) and older (>38 years) women. This study concluded that the AMH concentration was more variable in younger women (0.81 ± 0.59 μg/L) compared to older women (0.31 ± 0.29 μg/L) during the menstrual cycle ($P = .001$); thus, a single AMH measurement may be unreliable. A recent study using the GenII assay reported 20% intracycle variability in AMH measurements in women (n = 12) with regular ovulatory cycles (40). All the reports considered have findings consistent with a modest true systematic variability of 10–20% in the level of AMH in circulation during the menstrual cycle. Although there have been suggestions that this variability may differ between subgroups of women, these have been based on post hoc subgroup analyses, and there is no convincing evidence for such subgroups (38).

Variability between menstrual cycles

Three studies (Supplemental Table 1) evaluated AMH variability in samples taken during the early follicular phase of consecutive menstrual cycles (10, 29, 41), and three studies have reported on the variability of AMH in repeat samples from the same patient taken with no regard to the menstrual cycle (13, 42, 43). One study employed an in-house assay (41), one study used the IOT assay (29), three studies used the DSL assay (10, 42, 43), and one study (13) used the GenII assay. In four infertile women, Fanchin et al (41) assessed the early follicular phase AMH (in-house) variability across three consecutive menstrual cycles; they concluded that intersample AMH variability was characterized by an ICC of 0.89 (95% confidence interval [CI], 0.83–0.94). Streuli et al (29) calculated a between-sample coefficient of variation of 28.5% in AMH (IOT) in 10 healthy women. In 77 infertile women, van Disseldorp et al (10) found an intercycle AMH (DSL) variability of 11%. In summary, these studies suggest that the overall intercycle variability of AMH ranges from 11% (DSL) to 28% (IOT); this figure will include both biological and measurement-related variability.

Variability between repeat samples

Variability between repeat samples without regard to menstrual cycle phase was examined in three studies (Supplemental Table 1). In a group of 20 women, using samples frozen for prolonged periods, Dorgan et al (42) demonstrated a variability of 31% (ICC, 0.78; 95% CI, 0.60–0.95) between two samples, with a median between-sample interval of 1 year. In a larger series of 186 infertile women, Rustamov et al (43) (DSL) found a coefficient of variation of 28% between repeated samples, with a median between-sample interval of 2.6 months (ICC, 0.91; 95% CI, 0.90–0.93). Rustamov et al (13) found that the coefficient of variation of repeated GenII-assayed AMH in a group of 84 infertile women was 59% (ICC, 0.84; 95% CI, 0.79–0.90), substantially higher than that reported using the DSL assay. Similarly, a recent study by Hadlow et al (40) found a within-subject GenII-assayed AMH variability of 80%. As a result, five of the 12 women studied crossed clinical cutoff levels after repeated measurements.

Discussion: within-patient variability

Evidence suggests that repeated measurement of AMH can result in clinically important variability, particularly when using the GenII assay. This questions the assumption that a single AMH measurement is acceptable in guiding individual treatment strategies in ART.

The observed concentration of any analyte measured in a blood (serum) sample is a function of its "true" concentration and the influence of a number of other factors (Figure 1). Studies examining the variability of AMH by repeated measurement of the hormone will therefore reflect both true biological variation and measurement-related variability introduced by sample handling and/or processing. Thus, within-sample interassay variability used as an indicator of assay performance may not reflect true measurement-related variability between samples because it does not take into account the contribution from preanalytical variability. Measurement-related between-sample variability can be established in part using blood samples taken simultaneously (to avoid biological vari-

Table 3. Within-Subject Comparison Between AMH Methods

First Author (Ref.)	Assays	Subjects	Simultaneous Analysis	Regression	Summary
Freour (44)	DSL vs IOT	69 Infertile women age 22–40 y	Yes	IOT = 4.01 × DSL + 0.98 (μg/L) (Deming regression)	DSL = 22% IOT ($P < .0001$)
Hehenkamp (30)	DSL vs IOT	82 Healthy women	NR	DSL = 0.495 × IOT − 0.03	DSL = 49.5% IOT
Bersinger (45)	a. DSL vs IOT; b. DSL vs IOT	a. 11 Infertile women; b. 55 infertile women	a. Yes; b. no	a. DSL = 0.180 × IOT; b. DSL = 0.325 × IOT + 0.733	a. DSL = 18% IOT; b. DSL = 33% IOT
Zhao (17)	DSL vs IOT	38 Donors	NR	IOT = 1.5 × DSL + 0.7 (ng/ml)	DSL = 66% IOT
Taieb (46)	DSL vs IOT	104 Samples	NR	DSL = 1.04 × IOT − 1.49	DSL = 96% IOT
Streuli (29)	DSL vs IOT	153 Normal and infertile	No	IOT = 1.07 × DSL − 0.29	DSL = IOT
Kumar (18)	IOT vs GenII	60 Female, 60 male volunteers	NR	IOT = 1.0 GenII	IOT = GenII
Gada (50)	DSL vs GenII	42 Women	NR	NR	DSL = 63% GenII
Preissner (23)	DSL vs GenII	206 Samples	NR	GenII = 1.53 × DSL − 0.77	DSL = 66% GenII
Lee (47)	DSL vs IOT	172 Infertile women	Yes	IOT = 1.102 × DSL − 0.042	DSL = IOT
Wallace (51)	DSL vs GenII	271 Women	NR	GenII = 1.40 × DSL − 0.62	DSL = 71% GenII
Li (48)	a. DSL vs IOT; b. DSL vs GenII; c. IOT vs GenII	56 Women with PCOS or subfertility	Yes	a. IOT = 0.97 × DSL − 2.96; b. GenII = 1.33 × DSL − 4.17; c. GenII = 1.38 × IOT − 0.68	a. DSL = IOT; b. DSL = 67% GenII; c. IOT = 62% GenII
Rustamov (13)	DSL vs GenII	Female IVF patients (n = 330), median of 2 y between samples	No	NR	DSL = 127% GenII (age-adjusted)
Pigny (49)	IOT vs GenII	59 women: 32 controls, 27 with PCOS	Yes	NR	IOT = 200% GenII

Abbreviations: NR, not recorded; IVF, in vitro fertilization; PCOS, polycystic ovary syndrome.

ability) from a group of subjects, although even this does not reflect the full variability in sample processing and storage inherent in real clinical measurement.

Because AMH is only produced by steadily growing ovarian follicles, it is plausible to predict a small true biological variability in serum, reflected in the modest 1–20% variability found within the menstrual cycle. In contrast, it appears that the magnitude of measurement-related variability of AMH is more significant: 1) within-sample interassay variation can be as high as 13%; 2) different assays display substantially different variability; and 3) AMH appears to be unstable under certain conditions of sample handling and storage (Table 1). Consequently, any modest variation in true biological AMH concentration may be overshadowed by a larger, measurement-related variability, and careful experimental designs are required to characterize such differences. In general, the reported variability in published studies should be regarded as a measure of total sample-to-sample variability, ie, the sum of biological and measurement-related variability (Figure 1).

In repeat samples, the available evidence confirms that there is a significant level of within-patient variability between measurements that are assay-dependent, greater than the estimates of within cycle variability, and therefore likely to be predominantly measurement-related. Evidence from several sources suggests that the effects of sample handling, storage, and freezing differ between commercial assays and that the newer GenII assay may be more susceptible to these changes under clinical conditions. When it has been established that the modified protocol for the GenII assay can produce reproducible results independent of storage conditions, then it will be necessary to re-examine intra- and intercycle variability of AMH.

Assay Method Comparability

AMH assay comparisons have used either same-sample aliquots or population-based data with repeat samples. Study population characteristics, sample handling, inter-method conversion formulae, and results from these comparisons are summarized in Table 3. AMH levels were almost universally compared using a laboratory-based, within-sample design. The Rustamov et al study (13) was population-based, comparing AMH results in two different samples from the same patient at different time points using two different assays.

IOT vs DSL

Table 3 summarizes eight large studies (17, 29, 30, 44–48) that compared the DSL and IOT AMH assays. They demonstrate strikingly different conversion factors, from 5-fold higher with the IOT assay to assay equivalence. Most studies carried out both analyses at the same time to avoid analytical variation (Figure 1). However, this does mean that samples were batched and frozen at −18 to −80°C before analysis, which, as already outlined, may influence preanalytical variability and contribute to the observed discrepancies in conversion factors.

IOT vs GenII

Three studies have compared the IOT and GenII assays (Table 3). Kumar (18) reported that both assays gave identical AMH concentrations. However, Li et al (48) found that the IOT assay produced AMH values 38% lower than the GenII assay, whereas Pigny et al (49) found levels that were 2-fold lower.

DSL vs GenII

Four studies analyzed same-sample aliquots using the DSL and GenII assays, either simultaneously or sequen-

tially (23, 48, 50, 51). Only Li et al (48) gave details of sample handling (Table 3). All four studies found that AMH values were 35–50% lower using the DSL compared to the GenII assay.

Rustamov et al (13) carried out a between-sample comparison of the assays, measuring AMH in fresh or briefly stored clinical samples from the same women at different times, with values adjusted for patient age (Table 3). In contrast to within-sample comparisons, this study found that the DSL assay gave results, on average, 21% higher than with the GenII assay. Although this comparison is open to other biases, it does reflect the full range of variability present in clinical samples and avoids issues associated with longer term sample storage.

Discussion: assay method comparability

It is critical for across-method comparison of clinical studies that reliable conversion factors for AMH are established. In-house assays aside, three commercially available AMH ELISAs have been widely available (IOT, DSL, and GenII), and the literature demonstrates considerable diversity in reported conversion factors between first-generation assays (DSL vs IOT) and between first- and second-generation immunoassays (DSL/IOT vs GenII).

Although most studies appear to follow the manufacturers' protocols, detailed methodological information is sometimes lacking. The assessment of within-sample difference between the two assays should involve thawing of a single sample and simultaneous analysis of two aliquots with each assay. Both aliquots experience the same preanalytical sample-handling and processing conditions; therefore, the results should be reproducible, provided the AMH samples are stable during the post-thaw analytical stage and the study populations are comparable. However, this review has identified significant discrepancies between studies, perhaps due to either significant instability of the sample or significant variation in assay performance. Studies comparing AMH levels measured using different assays in populations during routine clinical use have also come to differing conclusions (13, 51). Given the study designs that workers have used to try to ensure that samples are comparable, the finding of significant discrepancies in the observed conversion factors between assays is consistent with the proposal that AMH is subject to instability during the preanalytical stage of sample handling. This, coupled with any differential sensitivity and specificity between these commercial assays, could give rise to the observed results; ie, some assays are more sensitive than others to preanalytical effects.

AMH Guidance in ART

AMH guidance ranges to assess ovarian reserve (52) or subsequent response to treatment (53, 54) have been published. The Doctors Laboratory, using the DSL assay, advised the following ranges for ovarian reserve (<0.57 pmol/L, undetectable; 0.57–2.1 pmol/L, very low; 2.2–15.7 pmol/L, low; 15.8–28.6 pmol/L, satisfactory; 28.7–48.5 pmol/L, optimal; and >48.5 pmol/L, very high), ranges that supposedly increased by 40% on changing to the GenII assay (51). More recently, other authors have attempted to correlate AMH levels with subsequent birth rates. Brodin et al (53), using the DSL assay, observed that higher birth rates were seen in women with an AMH level >21 pmol/L, and low birth rates were seen in women who had AMH levels <1.43 pmol/L. In the United Kingdom, the National Institute for Health and Care Excellence recently issued guidance on AMH levels in the assessment of ovarian reserve in the new clinical guideline on fertility (54). They advise that an AMH level of ≤5.4 pmol/L would indicate a low response to subsequent treatment, and an AMH ≥25.0 pmol/L indicates a possible high response. Although not specifically stated, interrogation of the guideline suggests that these levels have been obtained using the DSL assay, which is no longer available in the United Kingdom.

As discussed above, the initial study of comparability between the DSL and GenII assays reported that GenII generated values 40% higher compared to the DSL assay; clinics were therefore advised to increase their treatment guidance ranges accordingly (51). However, a more recent study using fresh samples found that the original GenII assay may actually give values that are 20–30% lower; suggesting that following the above recommendation may lead to allocation of patients to inappropriate treatment groups (13). The apparent disparity in assay comparison studies implies that AMH reference ranges and guidance ranges for in vitro fertilization treatment that have been established using one assay cannot be reliably used with another assay method without full, independent validation. Similarly, caution is required when comparing the outcomes of research studies using different AMH assay methods.

General Summary

Recent publications have suggested that GenII-assayed AMH is susceptible to preanalytical change leading to significant variability in determined AMH concentration, an observation now accepted by the kit manufacturer. However, this review suggests that all AMH assays may display a differential response to preanalytical proteolysis, conformational changes of the AMH dimer, or the presence of interfering substances. The existence of appreciable sample-to-sample variability and substantial discrepancies in

between-assay conversion factors, suggests that sample instability may have been an issue with previous AMH assays but appears to be more pronounced with the currently available GenII immunoassay. The observed discrepancies may be explicable in terms of changes in AMH or assay performance that are dependent on sample handling, transport, and storage conditions, factors underreported in the literature. We strongly recommend that future studies on AMH should explicitly report on how samples are collected, processed, and stored. If it can be clearly demonstrated that the new GenII protocol drives this process to completion in all samples ensuring stability, then a re-examination of reference and guidance ranges for AMH interpretation will be necessary. There is a clear need for an international reference standard for AMH and for robust independent evaluation of commercial assays in routine clinical samples with well-defined sample handling and processing protocols. These issues of sample instability and lack of reliable interassay comparability data should be taken into account in the interpretation of available research evidence and the application of AMH measurement in clinical practice.

Acknowledgments

Address all correspondence and requests for reprints to: Dr Oybek Rustamov, Department of Reproductive Medicine, St Mary's Hospital, Central Manchester University Hospital National Health Service Foundation Trust, Manchester Academic Health Science Centre, Manchester M13 0JH, United Kingdom. E-mail: oybek.rustamov@cmft.nhs.uk.

No funding was sought for this study. Publication costs were funded from Endowment Fund 629150.

Disclosure Summary: There were no potential conflicts of interest.

References

1. **Durlinger AL, Kramer P, Karels B, et al.** Control of primordial follicle recruitment by anti-Müllerian hormone in the mouse ovary. *Endocrinology*. 1999;140:5789–5796.
2. **Durlinger AL, Gruijters MJ, Kramer P, et al.** Anti-Müllerian hormone attenuates the effects of FSH on follicle development in the mouse ovary. *Endocrinology*. 2001;142:4891–4899.
3. **van Rooij IA, Broekmans FJ, te Velde ER, et al.** Serum anti-Müllerian hormone levels: a novel measure of ovarian reserve. *Hum Reprod*. 2002;17:3065–3071.
4. **Nelson SM, Yates RW, Fleming R.** Serum anti-Müllerian hormone and FSH: prediction of live birth and extremes of response in stimulated cycles–implications for individualization of therapy. *Hum Reprod*. 2007;22:2414–2421.
5. **Nardo LG, Gelbaya TA, Wilkinson H, et al.** Circulating basal anti-Müllerian hormone levels as predictor of ovarian response in women undergoing ovarian stimulation for in vitro fertilization. *Fertil Steril*. 2009;92:1586–1593.
6. **Yates AP, Rustamov O, Roberts SA, et al.** Anti-Mullerian hormone-tailored stimulation protocols improve outcomes whilst reducing adverse effects and costs of IVF. *Hum Reprod*. 2011;26:2353–2362.
7. **Lane AH, Lee MM, Fuller AF Jr, Kehas DJ, Donahoe PK, MacLaughlin DT.** Diagnostic utility of Müllerian inhibiting substance determination in patients with primary and recurrent granulosa cell tumors. *Gynecol Oncol*. 1999;73:51–55.
8. **Anderson RA.** Pretreatment serum anti-müllerian hormone predicts long-term ovarian function and bone mass after chemotherapy for early breast cancer. *J Clin Endocrinol Metab*. 2011;96:1336–1343.
9. **Broer SL, Eijkemans MJ, Scheffer GJ, et al.** Anti-mullerian hormone predicts menopause: a long-term follow-up study in normoovulatory women. *J Clin Endocrinol Metab*. 2011;96:2532–2539.
10. **van Disseldorp J, Lambalk CB, Kwee J, et al.** Comparison of inter- and intra-cycle variability of anti-Mullerian hormone and antral follicle counts. *Hum Reprod*. 2010;25:221–227.
11. **Hudson PL, Dougas I, Donahoe PK, et al.** An immunoassay to detect human müllerian inhibiting substance in males and females during normal development. *J Clin Endocrinol Metab*. 1990;70:16–22.
12. **Nelson SM, La Marca A.** The journey from the old to the new AMH assay: how to avoid getting lost in the values. *Reprod Biomed Online*. 2011;23:411–420.
13. **Rustamov O, Smith A, Roberts SA, et al.** Anti-Mullerian hormone: poor assay reproducibility in a large cohort of subjects suggests sample instability. *Hum Reprod*. 2012;27:3085–3091.
14. **Rustamov O, Smith A, Roberts SA, et al.** Author reply: reproducibility of AMH. *Hum Reprod*. 2012;27:3641–3642.
15. **Han X, McShane M, Sahertian R, White C, Ledger W.** Pre-mixing samples with assay buffer is an essential pre-requisite for reproducible Antimullerian Hormone (AMH) measurement using the Beckman Coulter Gen II assay (Gen II). *Hum Reprod* 2013;28(suppl 1):i76–i78 (abstract).
16. **Al-Qahtani A, Muttukrishna S, Appasamy M, et al.** Development of a sensitive enzyme immunoassay for anti-Müllerian hormone and the evaluation of potential clinical applications in males and females. *Clin Endocrinol (Oxf)*. 2005;63:267–273.
17. **Zhao J, Ng SY, Rivnay B, Leader BS.** Serum anti-Müllerian hormone (AMH) measurement: inter-assay agreement and temperature stability. *Fertil Steril*. 2007;88(suppl 1):S174.
18. **Kumar A, Kalra B, Patel A, McDavid L, Roudebush WE.** Development of a second generation anti-Müllerian hormone (AMH) ELISA. *J Immunol Methods*. 2010;362:51–59.
19. **Fleming R, Nelson SM.** Reproducibility of AMH. *Hum Reprod*. 2012;27:3639–3641; author reply 3641–3642.
20. **Fleming R, Fairbairn C, Blaney C, Lucas D, Gaudoin M.** Stability of AMH measurement in blood and avoidance of proteolytic changes. *Reprod Biomed Online*. 2013;26:130–132.
21. **Rey R, Lordereau-Richard I, Carel JC, et al.** Anti-müllerian hormone and testosterone serum levels are inversely during normal and precocious pubertal development. *J Clin Endocrinol Metab*. 1993;77:1220–1226.
22. **Long WQ, Ranchin V, Pautier P, et al.** Detection of minimal levels of serum anti-Müllerian hormone during follow-up of patients with ovarian granulosa cell tumor by means of a highly sensitive enzyme-linked immunosorbent assay. *J Clin Endocrinol Metab*. 2000;85:540–544.
23. **Preissner CM, Morbeck DE, Gada RP, Grebe SK.** Validation of a second generation assay for anti-Mullerian hormone. *Clin Chem*. 2010;56(suppl):A54.
24. **Cappy H, Pigny P, Leroy-Billiard M, Dewailly D, Catteau-Jonard S.** Falsely elevated serum antimüllerian hormone level in a context of heterophilic interference. *Fertil Steril*. 2013;99:1729–1732.
25. **Bungum L, Jacobsson AK, Rosén F, et al.** Circadian variation in concentration of anti-Müllerian hormone in regularly menstruating females: relation to age, gonadotrophin and sex steroid levels. *Hum Reprod*. 2011;26:678–684.
26. **Cook CL, Siow Y, Taylor S, Fallat ME.** Serum müllerian-inhibiting

substance levels during normal menstrual cycles. *Fertil Steril*. 2000; 73:859–861.
27. **La Marca A, Malmusi S, Giulini S, et al.** Anti-Müllerian hormone plasma levels in spontaneous menstrual cycle and during treatment with FSH to induce ovulation. *Hum Reprod*. 2004;19:2738–2741.
28. **La Marca A, Stabile G, Artenisio AC, Volpe A.** Serum anti-Mullerian hormone throughout the human menstrual cycle. *Hum Reprod*. 2006;21:3103–3107.
29. **Streuli I, Fraisse T, Chapron C, Bijaoui G, Bischof P, de Ziegler D.** Clinical uses of anti-Müllerian hormone assays: pitfalls and promises. *Fertil Steril*. 2009;91:226–230.
30. **Hehenkamp WJ, Looman CW, Themmen AP, de Jong FH, Te Velde ER, Broekmans FJ.** Anti-Müllerian hormone levels in the spontaneous menstrual cycle do not show substantial fluctuation. *J Clin Endocrinol Metab*. 2006;91:4057–4063.
31. **Lahlou N, Chabbert-Buffet N, Gainer E, Roger M, Bouchard P.** Diphasic pattern of anti-Mullerian hormone (AMH) in ovulatory cycles as evidenced by means of ultra-sensitive assay: new insights into ovarian function. *Fertil Steril*. 2006;86(suppl):S11.
32. **Tsepelidis S, Devreker F, Demeestere I, Flahaut A, Gervy Ch, Englert Y.** Stable serum levels of anti-Müllerian hormone during the menstrual cycle: a prospective study in normo-ovulatory women. *Hum Reprod*. 2007;22:1837–1840.
33. **Wunder DM, Bersinger NA, Yared M, Kretschmer R, Birkhäuser MH.** Statistically significant changes of antimüllerian hormone and inhibin levels during the physiologic menstrual cycle in reproductive age women. *Fertil Steril*. 2008;89:927–933.
34. **van Disseldorp J, Faddy MJ, Themmen AP, et al.** Relationship of serum antimüllerian hormone concentration to age at menopause. *J Clin Endocrinol Metab*. 2008;93:2129–2134.
35. **Sowers M, McConnell D, Gast K, et al.** Anti-Müllerian hormone and inhibin B variability during normal menstrual cycles. *Fertil Steril*. 2010;94:1482–1486.
36. **Robertson DM, Hale GE, Fraser IS, Hughes CL, Burger HG.** Changes in serum antimüllerian hormone levels across the ovulatory menstrual cycle in late reproductive age. *Menopause*. 2011;18:521–524.
37. **Overbeek A, Broekmans FJ, Hehenkamp WJ, et al.** Intra-cycle fluctuations of anti-Müllerian hormone in normal women with a regular cycle: a re-analysis. *Reprod Biomed Online*. 2012;24:664–669.
38. **Roberts SA.** Variability in anti-Mullerian hormone levels: a comment on Sowers et al, "Anti-Müllerian hormone and inhibin B variability during normal menstrual cycles." *Fertil Steril*. 2010;94:e59.
39. **Sowers M, McConnell D, Gast K, et al.** Author reply: Variability in anti-Mullerian hormone levels: a comment on Sowers et al, "Anti-Müllerian hormone and inhibin B variability during normal menstrual cycles." *Fertil Steril*. 2010;94:e60.
40. **Hadlow N, Longhurst K, McClements A, Natalwala J, Brown SJ, Matson PL.** Variation in antimüllerian hormone concentration during the menstrual cycle may change the clinical classification of the ovarian response. *Fertil Steril*. 2013;99:1791–1797.
41. **Fanchin R, Taieb J, Lozano DH, Ducot B, Frydman R, Bouyer J.** High reproducibility of serum anti-Mullerian hormone measurements suggests a multi-staged follicular secretion and strengthens its role in the assessment of ovarian follicular status. *Hum Reprod*. 2005;20:923–927.
42. **Dorgan JF, Spittle CS, Egleston BL, Shaw CM, Kahle LL, Brinton LA.** Assay reproducibility and within-person variation of Müllerian inhibiting substance. *Fertil Steril*. 2010;94:301–304.
43. **Rustamov O, Pemberton PW, Roberts SA, et al.** The reproducibility of serum anti-Müllerian hormone in subfertile women: within and between patient variability. *Fertil Steril*. 2011;95:1185–1187.
44. **Fréour T, Mirallié S, Bach-Ngohou K, Denis M, Barrière P, Masson D.** Measurement of serum anti-Müllerian hormone by Beckman Coulter ELISA and DSL ELISA: comparison and relevance in assisted reproduction technology (ART). *Clin Chim Acta*. 2007;375: 162–164.
45. **Bersinger NA, Wunder D, Birkhäuser MH, Guibourdenche J.** Measurement of anti-mullerian hormone by Beckman Coulter ELISA and DSL ELISA in assisted reproduction: differences between serum and follicular fluid. *Clin Chim Acta*. 2007;384:174–175.
46. **Taieb J, Belville C, Coussieu C, Guibourdenche J, Picard JY, Di clemente N.** Two immunoassays for antimullerian hormone measurement: analytical and clinical performances [in French]. *Ann Biol Clin (Paris)*. 2008;66:537–547.
47. **Lee JR, Kim SH, Jee BC, Suh CS, Kim KC, Moon SY.** Antimüllerian hormone as a predictor of controlled ovarian hyperstimulation outcome: comparison of two commercial immunoassay kits. *Fertil Steril*. 2011;95:2602–2604.
48. **Li HW, Ng EH, Wong BP, Anderson RA, Ho PC, Yeung WS.** Correlation between three assay systems for anti-Müllerian hormone (AMH) determination. *J Assist Reprod Genet*. 2012;29:1443–1446.
49. **Pigny P, Dassonneville A, Catteau-Jonard S, Decanter C, Dewailly D.** Comparative analysis of two-widely used immunoassays for the measurement of serum AMH in women. *Hum Reprod*. 2013; 28(suppl 1):i311–i316.
50. **Gada R, Hughes P, Amols M, et al.** Validation and comparison of AMH serum levels using the original active MIS/AMH ELISA to the new active AMH Gen II ELISA. *Fertil Steril*. 2011;95(suppl):S23.
51. **Wallace AM, Faye SA, Fleming R, Nelson SM.** A multicentre evaluation of the new Beckman Coulter anti-Mullerian hormone immunoassay (AMH Gen II). *Ann Clin Biochem*. 2011;48:370–373.
52. **The Doctors Laboratory.** Lab Report newsletter: Winter 2007/2008. TDL Test Update. Anti-Müllerian Hormone (AMH). http://www.tdlpathology.com/about-tdl/publications/lab-report-newsletter/winter-20072008. Accessed October 2, 2013.
53. **Brodin T, Hadziosmanovic N, Berglund L, Olovsson M, Holte J.** Antimullerian hormone levels are strongly associated with live birth rates after assisted reproduction. *J Clin Endocrinol Metab*. 2013; 98(3):1107–1104.
54. **National Institute for Health and Care Excellence.** NICE clinical guideline CG156. Fertility: assessment and treatment for people with fertility problems. http://www.nice.org.uk/Guidance/cg156. Accessed October 2, 2013.