High reproducibility of serum anti-Müllerian hormone measurements suggests a multi-staged follicular secretion and strengthens its role in the assessment of ovarian follicular status

Renato Fanchin1,4, Joëlle Taieb2, Daniel H. Mendèz Lozano1, Béatrice Ducot3, René Frydman1 and Jean Bouyer3

Departments of 1 Obstetrics and Gynecology and Reproductive Medicine, 2 Biochemistry and Hormonology, Clamart and 3 INSERM U569 Epidemiology, Kremlin-Bicêtre, France

4 To whom correspondence should be addressed at: Department of Obstetrics and Gynecology and Reproductive Medicine, Hôpital Antoine Béclère, 157, rue de la Porte de Trivaux, 92141, Clamart, France. E-mail: renato.fanchin@abc.ap-hop-paris.fr

BACKGROUND: Our aim was to compare the intercycle reproducibility of serum anti-Müllerian hormone (AMH) measurements with that of other markers of ovarian follicular status. METHODS: Forty-seven normo-ovulatory, infertile women underwent serum AMH, inhibin B, estradiol and FSH measurements and early antral follicle (2–12 mm in diameter) counts by transvaginal ultrasound on cycle day 3 during three consecutive menstrual cycles. Reproducibility of measurements was estimated using intra-class correlation coefficient (ICC) calculation. We also assessed the number of replicate measurements theoretically needed to reach satisfactory reliability of results. RESULTS: Serum AMH showed significantly higher reproducibility (ICC, 0.89; 95% confidence interval, 0.83–0.94) than inhibin B (0.76; 0.66–0.86; P<0.03), estradiol (0.22; 0.03–0.41; P<0.0001) and FSH levels (0.55; 0.39–0.71; P<0.01), and early antral follicle counts (0.73; 0.62–0.84; P<0.001), and reached satisfactory reliability with a single measurement. CONCLUSIONS: The improved cycle-to-cycle consistency of AMH as compared with other markers of ovarian follicular status is in keeping with its peculiar production by follicles at several developmental stages and further supports its role as a cost-effective, reliable marker of ovarian fertility potential.

Key words: anti-Müllerian hormone/intercycle variability/Müllerian-inhibiting substance/reproducibility

Introduction

Ovarian ageing is a lengthy, complex process characterized by the quantitative and qualitative attrition of ovarian follicles. From their early 30s to their early 40s, women are expected to exhaust three-quarters of their follicular reserve (Block, 1952), a phenomenon that exerts a determining effect on their fertility potential (Gougeon, 1996; Reuss et al., 1996). In an effort to quantify non-invasively the extent of follicular loss in a specific woman, a number of hormonal parameters have been proposed (Navot et al., 1987; Muasher et al., 1988; Fanchin et al., 1994, 2003a; Licciardi et al., 1995; Seifer et al., 1997), but their accuracy has been a matter of debate (Bancsi et al., 2002; Van Rooij et al., 2004). They are essentially based on the ability of early antral follicles to produce inhibin B and estradiol (E2) in response to FSH and on the modulating action of both hormones on the FSH secretion during the luteal–follicular transition in the menstrual cycle.

By definition, these hormonal tests disregard the status of other follicles that are barely or not sensitive to FSH and/or have not yet reached the antral stage, but which contribute to the functioning of subsequent menstrual cycles and to women’s fertility potential (Gougeon, 1996). In addition, they are reportedly biased by confounding variables linked to the status of follicular growth and to size discrepancies of early antral follicles during the early follicular phase (Klein et al., 1996; Fanchin et al., 2003a). It is conceivable that these limitations of traditional hormonal tests explain, at least in part, the noticeable variability of their results from one cycle to another (Scott et al., 1990; Brown et al., 1995; Scheffer et al., 1999; Hansen et al., 2003; Jain et al., 2003; Kwee et al., 2004), a phenomenon that contrasts with the overall short-term steadiness of ovarian follicular reserve and fertility potential (Gougeon, 1996).

Anti-Müllerian hormone (AMH), a glycoprotein that belongs to the transforming growth factor-β superfamily...
(Cate et al., 1986), recently has been proposed as a promising marker of the ovarian follicular status (De Vet et al., 2002; Seifer et al., 2002; Van Rooij et al., 2002), which reliability surpasses that of inhibin B, E<sub>2</sub> and FSH on cycle day 3 (Fanchin et al., 2003b). It exhibits at least three biological characteristics that are not shared by the conventional hormonal predictors of the follicular status. First, it is expressed in the granulosa cells of a wide variety of follicles that range from the large primary to the early antral stages (Baarends et al., 1995; Durlinger et al., 2002a; Weenen et al., 2004). Secondly, follicles that grow beyond the early antral follicle stage progressively lose their capacity to express AMH (Baarends et al., 1995; Fanchin et al., 2003e), which relatively preserves AMH measurements from bias linked to early follicular development (Klein et al., 1996; Fanchin et al., 2003a). Thirdly, although the mechanisms involved in the promotion or the inhibition of AMH production by granulosa cells remain undetermined, all indicate that it might be FSH independent (Durlinger et al., 2002b). Taken together, these physiological characteristics of AMH support the hypothesis that peripheral AMH measurements may provide information on the activity of a larger span of follicles with little or no influence of the hormonal–follicular dynamics at the luteal–fOLLICULAR transition.

Hence, we assumed that serum AMH levels should display improved cycle-to-cycle consistency as compared with the usual approaches of the ovarian follicular status. To test this hypothesis, we estimated the intercycle reproducibility of serum AMH measurements and compared it with that of conventional markers of follicular status such as serum inhibin B, E<sub>2</sub>, FSH and early antral follicle counts on cycle day 3 during three consecutive menstrual cycles.

**Materials and methods**

**Subjects**

We prospectively studied 47 infertile women, 25–40 years of age, who were undergoing infertility explorations before assisted reproductive technology. These routinely include hormone measurements performed on day 3 repeated at different menstrual cycles. All women met the following inclusion criteria: (i) both ovaries present and devoid of morphological abnormalities; (ii) regular menstrual cycle lengths ranging between 25 and 35 days; (iii) no current or past diseases affecting the ovaries or gonadotrophin or sex steroid secretion, clearance or excretion; (iv) no clinical signs of hyperandrogenism; (v) body mass indexes (BMIs) ranging from 18 to 25 kg/m<sup>2</sup>; (vi) no current hormone therapy; (vii) adequate visualization of both ovaries in transvaginal ultrasound scans; and (viii) no endometriosis. Aetiologies of infertility were sperm abnormalities (51%), tubal abnormalities (34%) and unexplained infertility (15%). An informed consent was obtained from all women and this investigation received the approval of our internal Institutional Review Board.

**Hormonal and follicular measurements**

On cycle day 3 during three consecutive menstrual cycles, women underwent serum AMH, inhibin B, E<sub>2</sub> and FSH measurements at ~9 a.m. All blood samples were obtained by venipuncture and serum was separated and frozen in aliquots at −80°C for subsequent centralized analysis. All hormone measurements for a single patient were performed within the same assay. Serum AMH levels were determined by the original ultrasensitive enzyme-linked immunosorbent assay (ELISA) (Beckman-Coulter, Villepinte, France) as described elsewhere (Long et al., 2000). The lower detection limit was 0.24 ng/ml and intra- and interassay coefficients of variation for AMH were <5 and <8%, respectively. Serum inhibin B levels were determined by double antibody ELISA (Serotec, Villabas, France) as previously described (Groome et al., 1996). The lower detection limit was 15 pg/ml, and intra- and interassay coefficients of variation for inhibin B were <6 and <9%, respectively. Serum E<sub>2</sub> and FSH levels were determined by an automated multi-analysis system using a chemiluminescence technique (Advia-Centaur, Bayer Diagnostics, Puteaux, France). For E<sub>2</sub>, the lower detection limit was 15 pg/ml, and intra- and interassay coefficients of variation were 8 and 9%, respectively. For FSH, the lower detection limit was 0.1 mIU/ml and intra- and interassay coefficients of variation were 3 and 5%, respectively.

Ovarian ultrasound scans were performed using a 3.6–8.0 MHz multi-frequency transvaginal probe (EC9-4, Sonoline Antares, Siemens S.A.S., Saint-Denis, France) by one single operator who was unaware of hormonal results, according to a methodology described previously (Fanchin et al., 2003a,b). In brief, ultrasound examinations assessed the number and sizes of early antral follicles. The early antral follicle count corresponded to the sum of all follicles that measured 2–12 mm in mean diameter (mean of two orthogonal diameters) in both ovaries. In an attempt to optimize the reliability of ovarian follicular assessment, the ultrasound scanner used was equipped with a tissue harmonic imaging system (Thomas and Rubin, 1998), which allows improved image resolution and adequate recognition of follicular borders. Intra-analysis coefficients of variation for follicular and ovarian measurements were <5% and their lower limit of detection was 0.1 mm.

**Statistics**

Measures of central tendency and variability used were the median and the range, respectively. To assess within-subject reproducibility of hormonal and ultrasonographic results, we calculated the intra-class correlation coefficient (ICC) and its 95% confidence intervals (Shrout and Fleiss, 1979) for each parameter. The ICC is the ratio of the between-subject variability over the total variability, the latter including between- and within-subject variability. In a series of measurements performed on different individuals, the ICC can be interpreted as the correlation coefficient between repeated measurements made in one of them. ICC ranges from 0 and 1.00, and values exceeding 0.80 usually indicate adequate reproducibility. If ICC values remain below 0.80, a single measurement is likely to be insufficient to categorize an individual with respect to the measured parameter. In this case, using the formula of Shrout and Fleiss (1979), we calculated the minimum number of replicate measurements that would be necessary to achieve satisfactory reproducibility of this mean (i.e. ICC ≥ 0.80). Values of ICC of different parameters were compared using a z-score test. Differences between paired variables were assessed by the Wilcoxon signed-rank test. The relationship between two different continuous variables was assessed by correlation coefficient. The Fisher r to z test was used to determine if the correlation coefficient (r) was significantly different from zero. A P-value <0.05 was considered statistically significant.
Results
Median age of participants was 33 years (range, 25–40), BMI was 20.7 kg/m\(^2\) (range, 18.2–24.8) and first and second menstrual cycle lengths were similar at 28 days (range, 25–34). The data distribution of serum hormone levels and of the number of early antral follicles on cycle day 3 during the three consecutive menstrual cycles is depicted in Figure 1. As shown, for cycles I, II and III, median serum AMH levels were 0.93 ng/ml (range, 0.24–6.40), 0.95 ng/ml (range, 0.24–7.08) and 0.81 ng/ml (range, 0.24–5.11), respectively. Median serum inhibin B levels were 88 pg/ml (range, 30–182), 86 pg/ml (range, 15–196) and 85 pg/ml (range, 15–138), respectively. Median serum E\(_2\) levels were 28 pg/ml (range, 15–111), 31 pg/ml (range, 15–206) and 32 pg/ml (range, 15–109), respectively. Median serum FSH levels were 7.7 mIU/ml (range, 3.6–19.3), 6.8 pg/ml (range, 2.9–18.3) and 7.8 pg/ml (range, 4.3–16.1), respectively. Median counts of early antral follicles (2–12 mm) were 11 (range, 1–28), 12 (range, 2–32) and 11 (range, 1–27), respectively. The longitudinal differences among serum AMH, inhibin B, E\(_2\), FSH and follicle measurements on cycles I, II and III did not reach statistical significance. During the three consecutive menstrual cycles, the number of antral follicles exceeded, in a small cohort of participants (\(n = 5\)), the threshold value reported in the definition of multifollicular ovaries as seen in polycystic ovary syndrome, i.e. 24 follicles (Rotterdam ESHRE/ASRM-sponsored PCOS Consensus Workshop Group, 2004). Conversely, this phenomenon was not observed in more than one cycle per patient.

During the three cycles, serum AMH levels were positively correlated with early antral follicle counts (\(r = 0.80, P < 0.0001\); \(r = 0.59, P < 0.0001\); and \(r = 0.71, P < 0.0001\), respectively) and with inhibin B (\(r = 0.32, P < 0.03\); \(r = 0.28, P < 0.05\); and \(r = 0.45, P < 0.001\), respectively) and tended to be negatively correlated with FSH (\(r = -0.28, P < 0.05\); \(r = -0.28, P < 0.06\); and \(r = -0.46, P < 0.001\), respectively). Serum E\(_2\) levels did not correlate significantly with any other variable during the three observation cycles. Similarly, menstrual cycle lengths, BMIs and aetiologies of infertility were not significantly related to hormonal and ultrasonographic data. As expected, patients’ ages were negatively correlated with serum AMH levels during the three consecutive menstrual cycles (\(r = -0.32, P < 0.03\); \(r = -0.35, P < 0.02\); and \(r = -0.30, P < 0.04\), respectively), but did not show any significant relationship with the remaining hormonal and ultrasonographic parameters.

Data on ICC values and the number of repeated measurements to achieve satisfactory reliability of the mean are summarized in Table I. As shown, ICC values for AMH were significantly higher than for inhibin B (\(P < 0.03\)), E\(_2\) (\(P < 0.0001\), FSH (\(P < 0.01\)) and early antral follicle count (\(P < 0.001\)), and required only one measurement to achieve satisfactory reliability of the mean.

Figure 1. Box-and-whiskers plot depicting serum AMH, inhibin B, E\(_2\) and FSH levels, and early antral follicle (2–12 mm in diameter) counts on day 3 during three consecutive menstrual cycles. Horizontal lines inside the boxes represent median levels. Upper and lower limits of the boxes and whiskers represent the 75th and 25th centiles and 90th and 10th centiles, respectively. Values outside the 90th and 10th centile limits are represented as dots. Note that longitudinal differences from one cycle to another did not reach statistical significance for any variable.

Discussion
The present study aimed at estimating the reproducibility of repeated serum AMH measurements in consecutive menstrual cycles and comparing it with the intercycle reproducibility of other putative markers of the ovarian follicular status. Some

925
physiological singularities of AMH, which are not shared by the remaining clinical approaches, such as multi-staged follicular production (Baarends et al., 1995; Durlinger et al., 2002a; Weenen et al., 2004), declining expression during final follicular maturation (Baarends et al., 1995; Fanchin et al., 2003c) and presumable FSH independence (Durlinger et al., 2002b), led us to hypothesize that AMH measurements might show optimum intercycle consistency as compared with inhibin B, E₂ and FSH levels and early antral follicle counts on cycle day 3.

To investigate this issue, the present study monitored a cohort of selected infertile patients during three consecutive menstrual cycles, and employed a statistical methodology that allowed us to integrate between-subject variability into the overall intercycle variability of measurements. For this, we preferred to calculate the ICC rather than the coefficient of variation because the coefficient of variation takes into account only the within-subject variability. Hence, when two different parameters show similar within-subject variability, their coefficients of variation will be the same, even if the first one has a greater between-subject variability than the second one. In this hypothetical case, the reproducibility of the first parameter would be higher. The determination of ICC prevents such a methodological flaw.

Our findings indicated that the intercycle variability of serum AMH levels is lower compared with that of the remaining hormonal parameters analysed. These results may be, at least in part, explained by the complex regulation of inhibin B, E₂ and FSH secretions during the luteal–follicular transition in the menstrual cycle. During this period, inhibin B and E₂ production by early antral follicles modulates their own stimulation by FSH. Early and/or intense FSH release may accelerate antral follicle growth (Klein et al., 1996) and prematurely increase inhibin B and/or E₂ levels on cycle day 3 (Fanchin et al., 2003a). In addition, in contrast to inhibin B, the ability of early antral follicles to produce E₂ in response to FSH is scarce and variable, probably because of their incipient aromatase activity (Erickson et al., 1979; Gougeon, 1996). This offers a possible explanation for the relatively improved reproducibility of serum inhibin B (ICC = 0.76) as compared with serum E₂ (ICC = 0.22) and FSH (ICC = 0.55) measurements on day 3. In line with this, to reach adequate reliability, our results indicated that serum E₂ and FSH measurements should be repeated at least during 14 and three menstrual cycles, respectively, a requirement that is clinically cumbersome.

As previously reported (De Vet et al., 2002; Van Rooij et al., 2002; Pigny et al., 2003), serum AMH levels showed, during the three consecutive menstrual cycles studied, a positive correlation with the number of early antral follicles, which tended to surpass that of the remaining hormones. These results are in keeping with our previous observations (Fanchin et al., 2003b) and indicate a conspicuous contribution of early antral follicles to peripheral AMH levels. Incidentally, the observation that in a small fraction of patients the number of follicles exceeded the threshold value reported in the definition of multifollicular ovaries, as seen in polycystic ovary syndrome (Rotterdam ESHRE/ASRM-Sponsored PCOS consensus workshop group, 2004), explains the fact that some of them tended to present abnormally high AMH and inhibin B levels (Pigny et al., 2003). Further, the number of early antral follicles showed subtle, yet noticeable, variations from one cycle to another, thereby corroborating previous reports (Scheffer et al., 1999; Hansen et al., 2003). In contrast, serum AMH levels appeared less prone to short-term variations than early antral follicle counts. Incidentally, coefficients of correlation values between AMH levels and follicle number were not constant during the three cycles analysed (r = 0.80, r = 0.59 and r = 0.71, respectively). Taken together, these results provide further support for the hypothesis that peripheral AMH levels are also influenced by the activity of follicles at other developmental stages.

In conclusion, the present study demonstrated that cycle-to-cycle reproducibility of serum AMH levels is higher than other markers of ovarian follicular status, such as serum inhibin B, E₂ and FSH and early antral follicle counts on day 3 of the menstrual cycle. Expanding the results of previous studies that recognized AMH levels as a reliable reflector of the long-term process of ovarian ageing (De Vet et al., 2002; Van Rooij et al., 2002), our observation of a short-term steadiness of AMH levels fits with the stability of women’s fertility potential from one cycle to another. In addition, the reliability of a single measurement represents a cost-effective and practical advantage of AMH over other hormonal markers, in particular, E₂ and FSH. Furthermore, due to its peculiar ontogenesis and regulation, the possibility of measuring AMH levels at alternative phases of the menstrual cycle rather than the first days of the follicular phase should be addressed in further studies. To clarify this issue, our group is currently conducting a prospective trial to detail the dynamics of peripheral AMH levels throughout the menstrual cycle. Additional studies are, however, needed to challenge the present results and to confirm the clinical usefulness of single AMH measurements, considered either alone or in combination with other markers such as early antral follicle counts, to assess reliably the ovarian follicular status and fertility potential.

926
References


Submitted on September 20, 2004; accepted on October 25, 2004