A Validated Model of Serum Anti-Müllerian Hormone from Conception to Menopause

Thomas W. Kelsey¹, Phoebe Wright², Scott M. Nelson³, Richard A. Anderson⁴, W. Hamish B. Wallace⁵*

¹ School of Computer Science, University of St. Andrews, St. Andrews, Scotland, United Kingdom, ²University of Edinburgh, Edinburgh, Scotland, United Kingdom, ³Centre for Population and Health Sciences, University of Glasgow, Glasgow, Scotland, United Kingdom, ⁴MRC Centre for Reproductive Health, University of Edinburgh, Edinburgh, Scotland, United Kingdom, ⁵Division of Reproductive and Developmental Sciences, University of Edinburgh, Edinburgh, Scotland, United Kingdom

Abstract

Background: Anti-Müllerian hormone (AMH) is a product of growing ovarian follicles. The concentration of AMH in blood may also reflect the non-growing follicle (NGF) population, i.e. the ovarian reserve, and be of value in predicting reproductive lifespan. A full description of AMH production up to the menopause has not been previously reported.

Methodology/Principal Findings: By searching the published literature for AMH concentrations in healthy pre-menopausal females, and using our own data (combined n = 3,260) we have generated and robustly validated the first model of AMH concentration from conception to menopause. This model shows that 34% of the variation in AMH is due to age alone. We have shown that AMH peaks at age 24.5 years, followed by a decline to the menopause. We have also shown that there is a neonatal peak and a potential pre-pubertal peak. Our model allows us to generate normative data at all ages.

Conclusions/Significance: These data highlight key inflection points in ovarian follicle dynamics. This first validated model of circulating AMH in healthy females describes a transition period in early adulthood, after which AMH reflects the progressive loss of the NGF pool. The existence of a neonatal increase in gonadal activity is confirmed for females. An improved understanding of the relationship between circulating AMH and age will lead to more accurate assessment of ovarian reserve for the individual woman.

Introduction

The human ovary establishes its complete complement of primordial follicles during fetal life, with recruitment and thereby depletion of this dormant primordial follicle pool required for normal fertility but ultimately leading to reproductive senescence [1]. Primordial follicles are recruited continuously from before birth to join the early growing cohort (initial recruitment). After puberty, at every new cycle a limited number of follicles are recruited from this cohort of small growing follicles (cyclic recruitment), followed by a final selection for dominance and ovulation of a single follicle [2,3]. Thus, at any specific time, the majority of primordial follicles are held in a dormant state, and when eventually recruited most will not reach the preovulatory stage but are destined to be removed through atresia at earlier stages of follicular development. Currently, clinical assessment is unable to assess the number of primordial follicles, or their rate of loss/activation. Knowledge of these aspects of ovarian function would be of value in a range of contexts, both clinical and social/personal, as well as being of great value in promoting our understanding of how reproductive lifespan is regulated.

Anti-Müllerian hormone (AMH), is now recognized as a principal regulator of early follicular recruitment from the primordial pool [4,5], with AMH null mice demonstrating accelerated depletion of primordial follicle number and an almost three-fold increase in smaller growing follicles [6]. Furthermore this increase in number of growing follicles occurs despite lower serum follicle stimulating hormone (FSH) concentrations [7], suggesting that in the absence of AMH, follicles are more sensitive to FSH and progress through the early stages of follicular development. AMH is produced by small growing but not primordial follicles [8–10], although limited data suggest that serum AMH concentrations also correlate with primordial follicle number in humans [11] as in rodents [12]. The prepubertal endocrine environment is markedly different from the adult with low and non-cyclical gonadotropins: the relevance of this to AMH secretion is incompletely understood although follicle growth through the preantral stages and occasionally to early antral stages (i.e. across the full range of stages that secrete AMH) is observed in childhood [13]. A recent study has reported an increase in initial primordial follicular recruitment rates up to the age of puberty, and then a progressive decline to the menopause [14]. This suggests that AMH concentrations at any given age in both childhood and adulthood may mirror primordial follicular recruitment rates, rather than simply primordial follicle number. Consequently across the female lifespan, circulating AMH will potentially exhibit an initial increase followed by a non-linear decline as is well established for the primordial follicle pool [14–18].
In keeping with this, AMH concentrations in adults have been shown to decline with age [19,20]. AMH concentrations are relatively stable across the menstrual cycle [21,22] and also between cycles in the same woman [23]. Measurement of AMH is increasingly used in the prediction of ovarian response to superovulation [24]. Although large AMH cohort studies describing falling AMH concentrations in adult women (mostly from populations attending infertility clinics) have recently been published [25–28], the data for AMH concentrations in children have until recently been considerably more limited [29,30]. To date no single study has examined AMH across the lifespan in healthy females. The aim of the current study is to produce a model of serum AMH in healthy females from conception to the menopause.

**Results**

The validated model

Data from published studies and our own (Table 1) were used to derive the model (Figure 1). This included 3,260 data points across the age range from 0.3 (cord blood from preterm infants) to 54.3 years (Table 2). After 10-fold stratified cross-validation, the model with the highest coefficient of determination ($r^2$) was a degree 20 polynomial of the form

\[
\log_{10}(AMH + 1) = c_0 + c_1 \text{age} + c_2 \text{age}^2 + \ldots + c_{20} \text{age}^{20}
\]

with coefficients $c_i$ given in Table 3.

This model has an $r^2$ of 0.34, indicating that roughly one third of variation in AMH concentrations is due to age alone, with the remaining two thirds of the variation being due to other factors. Serum AMH peaks at 24.5 years on average, with concentrations decreasing shortly after birth, and again between eight and twelve years of age.

**Validation**

For each of the 10 folds of the validation process, the highest ranked of 215 models was a polynomial similar to the model reported above: in every case the returned coefficients, $r^2$ and peaks were similar (Table 4). Moreover, the average training error was within 1% of the average cross-validation estimate of the prediction error across 10 folds (Table 4). We therefore consider that the model derived for all 3,260 datapoints generalises well to unseen data, and hence report this as a validated model for serum AMH concentrations in the normal female population (Figure 2).

---

**Table 1. Serum AMH data summary.**

<table>
<thead>
<tr>
<th>Ref.</th>
<th>1st Author</th>
<th>Data</th>
<th>Assay</th>
<th>n</th>
<th>Average age</th>
<th>Age range</th>
<th>Det. lim.</th>
<th>Intra CV</th>
<th>Inter CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>[35]</td>
<td>Soto</td>
<td>Graph</td>
<td>IBC</td>
<td>58</td>
<td>30.3 (mean)</td>
<td>± 8.7 SD</td>
<td>0.10</td>
<td>5.3</td>
<td>8.7</td>
</tr>
<tr>
<td>[38]</td>
<td>Guibourdenche</td>
<td>Graph</td>
<td>IBC</td>
<td>192</td>
<td>N5</td>
<td>± 0.3–1.0</td>
<td>0.30</td>
<td>5.3</td>
<td>8.7</td>
</tr>
<tr>
<td>[39]</td>
<td>Hudecova</td>
<td>Graph</td>
<td>IBC</td>
<td>64</td>
<td>46.3 (mean)</td>
<td>± 6.4 SD</td>
<td>0.70</td>
<td>12.3</td>
<td>12.3</td>
</tr>
<tr>
<td>[40]</td>
<td>Mulders</td>
<td>Graph</td>
<td>IBC</td>
<td>82</td>
<td>29.9</td>
<td>19.6–35.6</td>
<td>N5</td>
<td>5.0</td>
<td>8.0</td>
</tr>
<tr>
<td>[41]</td>
<td>Pastor</td>
<td>Graph</td>
<td>IBC</td>
<td>42</td>
<td>N5</td>
<td>18.0–50.0</td>
<td>0.10</td>
<td>5.3</td>
<td>7.8</td>
</tr>
<tr>
<td>[42]</td>
<td>Piltonen</td>
<td>Graph</td>
<td>IBC</td>
<td>44</td>
<td>31.6 (mean)</td>
<td>21.0–44.0</td>
<td>N5</td>
<td>5.1</td>
<td>6.6</td>
</tr>
<tr>
<td>[20]</td>
<td>van Rooij</td>
<td>Graph</td>
<td>IBC</td>
<td>162</td>
<td>N5</td>
<td>25.0–46.0</td>
<td>0.05</td>
<td>5.0</td>
<td>8.0</td>
</tr>
<tr>
<td>[43]</td>
<td>Laven</td>
<td>Graph</td>
<td>IBC</td>
<td>41</td>
<td>N5</td>
<td>20.0–36.0</td>
<td>0.05</td>
<td>5.0</td>
<td>8.0</td>
</tr>
<tr>
<td>[19]</td>
<td>de Vet</td>
<td>Graph</td>
<td>IBC</td>
<td>82</td>
<td>29.0</td>
<td>± 4.0 SD</td>
<td>0.05</td>
<td>5.0</td>
<td>8.0</td>
</tr>
<tr>
<td>[44]</td>
<td>Knauf</td>
<td>Graph</td>
<td>IBC</td>
<td>83</td>
<td>34.2 (mean)</td>
<td>± 3.4 SD</td>
<td>0.03</td>
<td>11.0</td>
<td>11.0</td>
</tr>
<tr>
<td>[45]</td>
<td>Lee</td>
<td>Graph</td>
<td>IBC</td>
<td>225</td>
<td>NS</td>
<td>0.0–51.0</td>
<td>0.50</td>
<td>9.0</td>
<td>15.0</td>
</tr>
<tr>
<td>[36]</td>
<td>La Marca</td>
<td>Graph</td>
<td>IBC</td>
<td>24</td>
<td>44.0 (mean)</td>
<td>± 2.8 SD</td>
<td>0.24</td>
<td>5.0</td>
<td>8.0</td>
</tr>
<tr>
<td>[29]</td>
<td>Hagen</td>
<td>Graph</td>
<td>IBC</td>
<td>891</td>
<td>N5</td>
<td>0.0–68.0</td>
<td>0.03</td>
<td>7.8</td>
<td>11.6</td>
</tr>
<tr>
<td>[46]</td>
<td>van Beek</td>
<td>Graph</td>
<td>DSL</td>
<td>82</td>
<td>29.0</td>
<td>20.0–35.0</td>
<td>N5</td>
<td>5.0</td>
<td>15.0</td>
</tr>
<tr>
<td>[47]</td>
<td>Sanders</td>
<td>Graph</td>
<td>DSL</td>
<td>43</td>
<td>24.1 (mean)</td>
<td>0.1–51.0</td>
<td>0.01</td>
<td>NS</td>
<td>11.4</td>
</tr>
<tr>
<td>[34]</td>
<td>van Disselrop</td>
<td>Graph</td>
<td>DSL</td>
<td>144</td>
<td>37.9 (mean)</td>
<td>25.0–46.0</td>
<td>0.03</td>
<td>11.0</td>
<td>11.0</td>
</tr>
<tr>
<td>[48]</td>
<td>Tehrani</td>
<td>Graph</td>
<td>DSL</td>
<td>267</td>
<td>27.1</td>
<td>16.0–44.0</td>
<td>0.01</td>
<td>5.2</td>
<td>9.1</td>
</tr>
<tr>
<td>[49]</td>
<td>Dorgan</td>
<td>Graph</td>
<td>DSL</td>
<td>204</td>
<td>44.7 (mean)</td>
<td>33.3–54.7</td>
<td>0.06</td>
<td>8.0</td>
<td>8.0</td>
</tr>
<tr>
<td>[30]</td>
<td>Ahmed</td>
<td>Raw</td>
<td>DSL</td>
<td>128</td>
<td>8.5</td>
<td>0.5–16.5</td>
<td>0.50</td>
<td>8.0</td>
<td>8.0</td>
</tr>
<tr>
<td>[25]</td>
<td>Nelson</td>
<td>Raw</td>
<td>DSL</td>
<td>441</td>
<td>36.1</td>
<td>21.9–47.8</td>
<td>0.03</td>
<td>3.4</td>
<td>8.6</td>
</tr>
<tr>
<td>Total IBC</td>
<td></td>
<td></td>
<td></td>
<td>1,990</td>
<td>15.8</td>
<td>± 0.3–68.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total DSL</td>
<td></td>
<td></td>
<td></td>
<td>1,309</td>
<td>35.4</td>
<td>0.2–54.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total n</td>
<td></td>
<td></td>
<td></td>
<td>3,299</td>
<td>34.0</td>
<td>± 0.3–68.0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The references relate to the bibliography section of this paper. Age information is given as median and range, or as mean and standard deviation (SD), depending on which form was reported in the referenced study. Detection limits are given in ng/ml. Intra- and inter-assay coefficients of variation (CV) are percentages. NS denotes not stated. For longitudinal studies – [19,29,40,48] – we report the average age of participants at first measurement. The censored total excludes any values greater than 54.3 years (i.e. one standard deviation above the average age at menopause).

doi:10.1371/journal.pone.0022024.t001
We have generated the first validated model of serum AMH in the healthy human population from conception to menopausal ages. Our model shows that serum AMH concentrations peak at age 24.5 years for the average case, and suggests that two thirds of the variation in AMH concentrations for healthy females is due to factors other than age.

We have shown that serum AMH falls shortly after birth, with concentrations only increasing again after about two years of age. This feature is in line with evidence of a mini-puberty seen in neonatal girls [31,32], although more clearly characterized in boys [33], and with a recent longitudinal study of AMH in female neonatal blood [29] (the data from which were included in our study). Our model also shows that serum AMH concentration falls between the ages of eight and twelve, before rising to a peak in the mid-twenties. This fall may be an artefact of our model derivation process rather than a true reproductive biological event. As the fall coincides with the initial increases in gonadotropin concentrations of early puberty, it is possible that it reflects changes in the proportions of follicles at different stages of growth with increasing numbers progressing to antral stages rather than becoming atretic early on [3]. AMH is produced by early growing follicles at all stages up to the early antral stage [8] but it is unknown which follicle class contributes most to circulating concentrations. The rising granulosa cell mass (and thus AMH production per follicle) will be balanced by progressively declining numbers of follicles at each stage of growth [2,3].

Our incomplete understanding of how early follicle recruitment and growth are regulated means that any such interpretation is speculative but our results both suggest this avenue of future research, and give useful indications of effect sizes and ages for the design of such investigations.

The increase in AMH during the postnatal period, supported by a recent longitudinal analysis in the first 3 months of life [29], is likely to be analogous to the well established transient rise in testosterone and inhibin B in boys at that time [33]. This is likely to reflect the relatively high gonadotrophin concentrations that are present which will support more advanced follicle development than occurs in the remainder of the prepubertal period. Consistent with this, the ovary shows follicle growth to the early antral stage from birth [13]. The continuing rise in AMH through childhood is striking, and parallels rising follicle growth initiation from the very large pool at these ages [14]. The rising AMH production would therefore act to limit follicle growth activation [5,6], thus a point of inflection when follicle recruitment starts to slow, and which is followed by a decline in AMH concentrations is predictable: our data demonstrate the age at which this occurs.

Our observed increase in AMH concentrations beyond the age of 12 years, is in contrast to the recent analysis by Hagen et al. which suggested in a cross-sectional study that AMH did not change from age 8 to age 25 or indeed relative to pubertal stage [29]. Censoring our own data between these two ages would also suggest that AMH does not vary markedly, however, this would lose the power of the broader picture afforded by the entire age range analysis performed here. The lack of an increase beyond age 12 as suggested by Hagen et al., is in marked contrast to current and previous studies, which have all suggested a peak in early adulthood. A recent analysis of 9,601 infertility patients [25] and a smaller study of 82 healthy subjects both report that AMH concentrations decrease after age 20 [19]. Although

![Figure 1. Serum AMH data.](doi:10.1371/journal.pone.0022024.g001)
two studies have reported peak AMH at age 31–33 [34,35], these limitations – including the accuracy of the health status of subjects

The columns are the 21 coefficients $c_k$ for the degree 20 polynomial returned by TableCurve2D that gave the highest $r^2$ for the dataset under consideration. For fold $k_i$, the dataset consisted of all points except the $k_i$ test data. After validation of the degree 20 polynomial (Table 3), its coefficients were calculated for the entire dataset

The 3,260 AMH datapoints described in Table 1, split into ages.

Sample sizes at each age for the AMH model.

The 3,260 AMH datapoints described in Table 1, split into ages. $n$ denotes the size of the subset of the data associated with an age in years.

The model is based entirely on cross-sectional data (the small amount of longitudinal data was treated as if cross-sectional). Further longitudinal validation in similarly large populations will provide additional confirmation.

It should be noted that our methodological choices have no effect on the qualitative nature of our results. If we convert to DSL values instead of IBC, do not add one to the log-adjusted values for case of exposition, do not censor at age 54.3, allow models with more parameters (or restrict to models with slightly fewer parameters), use $k \times 2$ or bootstrap cross-validation, we still obtain a validated model with similar peaks, $r^2$ and level of generalisation to unseen data. The current introduction of the GenII AMH assay (Beckman Coulter) uses the same standards as the IBC assay, thus we anticipate that the normative model presented (Figure 2) will be valid for values obtained using that system.

This comprehensive statistical analysis of 3,260 healthy infants, children and women has facilitated the first validated normative

The columns are the 21 coefficients $c_k$ for the degree 20 polynomial returned by TableCurve2D that gave the highest $r^2$ for the dataset under consideration. For fold $k_i$, the dataset consisted of all points except the $k_i$ test data. After validation of the degree 20 polynomial (Table 3), its coefficients were calculated for the entire dataset (final column). We report this model as our validated model of serum AMH from conception to menopause.

Table 3. Sample sizes at each age for the AMH model.

Table 3. 10-fold cross-validation models.

Table 2. Sample sizes at each age for the AMH model.

Table 3. 10-fold cross-validation models.

Table 2. Sample sizes at each age for the AMH model.

Table 3. 10-fold cross-validation models.
model of age related circulating AMH from conception to the menopause. The model provides a means for interpretation of how an individual’s serum AMH concentration compares with population norms.

Materials and Methods

Data acquisition

Studies involving serum AMH measurements of human females were identified by performing PubMed and Medline searches and searching individual journals (including Menopause, Fertility and Sterility, Human Reproduction and the Journal of Clinical Endocrinology and Metabolism) using the search terms AMH, Mullerian inhibiting substance, ovarian reserve and polycystic ovarian syndrome. The references of included studies (Table One; [19,20,29,34–36,38–49]) were checked to identify further relevant studies to be processed. Data was selected for this analysis only for subjects who were not known to be infertile, or have an identified chronic illness. Hence all subjects were either in control groups from controlled studies or from prospective studies of the healthy population. Any data from subjects with a chronic disease or undergoing infertility assessment or investigation were excluded from the study. In the main, the data was from pre-menopausal women. In three studies the menopausal status of the women was not stated [45,47,49]. Data from fetal blood (n = 25) and cord blood (n = 53) of infants [38] were included. Longitudinal data – from [19,29,40,48] – were recorded as cross-sectional values. The data were extracted from graphs using Plot Digitizer software [50] to convert datapoints on the graphs into numerical data. Repeated datapoints were isolated by requiring that the acquired dataset

<table>
<thead>
<tr>
<th>Table 4. 10-fold cross-validation results.</th>
</tr>
</thead>
<tbody>
<tr>
<td>k0</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>MSE train</td>
</tr>
<tr>
<td>MSE test</td>
</tr>
<tr>
<td>r²</td>
</tr>
<tr>
<td>Age at peak AMH</td>
</tr>
</tbody>
</table>

The mean squared error (MSE) for the model with highest r² is given for both the training set (90% of the dataset) and the test set (the remaining 10%) for each of the ten folds. Since – both for individual folds and on average – the errors are similar, we consider the model to be validated. The r² and peak ages are for the highest ranked model returned by TableCurve2D for each fold.

doi:10.1371/journal.pone.0022024.t004

Figure 2. The normal range for serum AMH in girls and women. The red line is the log-unadjusted validated AMH model using IBC assay values. The blue and green lines are the 68% and 95% prediction limits for the model (plus and minus one and two standard deviations respectively).

doi:10.1371/journal.pone.0022024.g002
matched the descriptive statistics provided in the supporting paper (Table One).

We combined the resulting dataset with two sets of raw Scottish data. The first consisted of individual serum AMH measurements (n = 441, median age 36.1 years, max. age 47.8, min age 21.9) from a cohort of women whose partners were known to have severe male factor infertility requiring ICSI, and where no other female cause of infertility had been identified [25]. Individual patient serum AMH measurements were undertaken between July 2006 and October 2009 in the biochemical laboratories of the University of Glasgow, the Glasgow Centre for Reproductive Medicine and the Glasgow Royal Infirmary. All three facilities were providing centralised AMH testing for infertility clinics within the United Kingdom. Ethical approval for studies involving these data was obtained from NHS Scotland. Subjects were informed that data may be analysed anonymously. Individual ethical approval has not been taken for studies involving this data, as it is routine anonymous clinical data which is covered by the general UK National Health Service ethics for analysis of routine biochemical data, provided it is anonymous. The second dataset was supplied by Ahmed et al. [30] and consists of 128 measurements taken from subjects aged 0.5–16.5 years.

Serum AMH values were standardised to give AMH measurements in ng/ml using the conversion formula 1 pmol/1 = 7.145 ng/ml.

The resulting data were considered separately depending on the assay used to obtain serum AMH values. The first dataset (n = 1,990, median age 15.8 years, max. age 68.0, min age −0.3) came from those studies in which the serum concentrations of AMH were determined using enzyme-linked immunosassay kits IBC (Immunotech Beckman Coulter Company, France). The second dataset (n = 1,309, median age 33.4 years, max. age 54.7, min age 0.2) came from studies in which the enzyme-immunometric assay Active MIS/AMH ELISA kits DSL (Diagnostic Systems Laboratories Inc., TX, USA) were used. We converted the DSL data into IBC values using the conversion formula

2.02 × DSL = IBC

which has a reported r² of 0.85 [51], and censored 39 datapoints over 54.3 years (mean age at menopause plus one standard deviation [52].) The resulting dataset consists of 3,260 serum AMH concentrations at known ages, and approximates circulating AMH concentrations in the healthy population from conception to menopause.

Data analysis

Stratified 10-fold cross-validation was performed using standard techniques [53]. The dataset was split into 10 distinct subsets, k₀, . . ., k₉, of nearly equal size, each with similar mean, median, minimum and maximum. For fold i, kᵢ was retained as test data, with the remaining 90% of the data used for training purposes. For each training set we added zero AMH values at conception, in order to force models through the only known AMH concentration at any age. Since variability increases with AMH concentration, we log-adjusted the data (after adding one to each value so that zero AMH on a chart represents zero serum AMH). All data were analysed as cross-sectional values (i.e. longitudinal patterns were not considered). We then fitted 215 mathematical models to the i-th test data using TableCurve-2D (Systat Software Inc., San Jose, California, USA), and ranked the results by coefficient of determination, r². Each model defines a generic type of curve and has parameters which, when instantiated gives a specific curve of that type. For each model we calculated values for the parameters that maximise the r² coefficient. The Levenberg-Marquardt non-linear curve-fitting algorithm was used throughout, with convergence to 5 significant figures after a maximum of 1,000 iterations. The highest ranked model was chosen as the best model for the test data, and the mean square error and r² were calculated after removing the artificial zero values at conception.

For validation purposes, the mean square error of the kᵢ data for the i-th model was calculated and compared to the mean square error of training data for the same model. In other words, the cross-validation estimate of the prediction error of the model was compared to the training error of the model. Training error is expected to be underestimated (due to overfitting of the chosen data) and the cross-validation estimate of the prediction error is expected to be overestimated (due to underfitting of unseen data). We consider a model to be validated if the differences between these errors is small.

For each fold the highest ranked model was a degree 20 polynomial of the form

\[ \log_{10}(AMH + 1) = c₀ + c₁age + c₂age² + \ldots + c_{20}age^{20} \]

Hence, after validation (Tables 3 and 4), the 21 parameters for this model were derived for the entire dataset (n = 3,260), again using TableCurve2D. The resulting model is reported as our validated model (Figure 2).

Author Contributions

Conceived and designed the experiments: TWK PW SMN RAA WHBW. Performed the experiments: TWK PW SMN RAA WHBW. Analyzed the data: TWK PW SMN RAA WHBW. Contributed reagents/materials/analysis tools: TWK PW SMN RAA WHBW. Wrote the paper: TWK PW SMN RAA WHBW.

References