DEBATE—continued

Assessment of ovarian reserve

Ovarian biopsy is not a valid method for the prediction of ovarian reserve

C.B.Lambalk1,4, C.H.de Koning1, A.Flett1, Y.van Kasteren2, R.Gosden3 and R.Homburg1

1Division of Reproductive Medicine, Department of Obstetrics and Gynaecology, VU Medical Centre, Amsterdam, 2Department of Obstetrics and Gynaecology, Medisch Centrum Alkmaar, Alkmaar, The Netherlands and 3The Jones Institute for Reproductive Medicine, Eastern Virginia Medical School, Norfolk, VA, USA

4To whom correspondence should be addressed at: Division of Reproductive Medicine, Department of Obstetrics and Gynaecology, Vrije Universiteit Medical Centre, De Boelelaan 1117, Postbus 7057, 1007 MB Amsterdam, The Netherlands.
E-mail: cb.lambalk@vumc.nl

The evaluation of ovarian reserve, often critical for the elderly infertile woman, is notoriously difficult and inaccurate. The place of ovarian biopsy in this evaluation has been hotly disputed for three decades, but not resolved. To examine the feasibility of ovarian biopsy for this purpose, a project was designed to estimate the total number of oocytes in a human ovary and investigate whether any biopsy regimen is representative of the follicular reserve in an individual. Ovaries removed from patients of reproductive age during operations not involving ovarian pathology were utilized to count the number and type of follicles found in multiple biopsies of 2 and 5 mm and in the whole ovary. Representative results taking into account the total number of follicles found in the whole ovary showed that predicted values based on the biopsies were extremely varied. We concluded that due to the huge variation in the distribution of follicles across the surface of the ovary, there is no place for this procedure in clinical evaluation of reproductive ageing in the individual patient.

Key words: biopsy/ovarian reserve

Introduction

This contribution to the debate on the use of ovarian biopsies for assessing ovarian reserve (Lass, 2004; Sharara and Scott, 2004) is based on our experience while examining whether laparoscopic ovarian biopsy for such a purpose was justified.

It is obvious from the literature that it is very difficult to predict ovarian reserve and the age of menopause with accuracy. An accurate prediction would be invaluable for several reasons: planning pregnancies for those at risk of premature ovarian failure (POF) and those likely to have a later menopause, and for the determination of the status of the ovary for preservation before cancer treatment.

As life expectancy has increased, so has the desirable age for conception. This has led to the need for a way of forecasting the age when the follicular store will be extinguished. Since there is a continual fall in follicle number, with no method of replacing lost follicles, estimation of follicle number and rate of decline seems the most logical method of forecast. This is not as easy as it sounds as obtaining a measure of the total number of oocytes in the ovaries is not possible without removing them. In addition, the rate of follicle loss is not constant (Faddy et al., 1992; Gougeon, 1996).

Many techniques and observations have been used to attempt to estimate ovarian reserve and predict those with a poor chance of success in ART: age, FSH, inhibin, anti-Müllerian hormone and estradiol concentrations; dynamic tests using GnRH agonist, FSH or clomiphene citrate; measurement of ovarian volume, ovarian antral follicle count; and, the subject of our investigation, ovarian biopsy.

We designed a project to collect ovaries during surgery from women in the fertile period of life, taking systematic biopsies and counting the number of follicles in the biopsies and in the whole ovary in order to examine if one or more biopsies could be used to predict the grand total reliably.

Materials and methods

Patients

Ovaries removed from five patients of varying ages for reasons not involving ovarian pathology were utilized in this study. All were originally cycling women whose ages ranged from 30 to 46 years. The reasons for oophorectomy were unexplained abdominal pain, breast cancer and female to male transexuality.

Procedure

After surgery, the surface of the ovary was removed and divided into 12 segments and, from each segment, a biopsy of 2 mm diameter was taken with a Schumacher or other instrument. The region closest to the hilus was labelled A, the middle as B and the outermost region (closest...
to the ligamentum suspensorium) as C (see Figure 1). From each part, four biopsies were taken at regular intervals and named A1, A2, A3, A4, B1, etc. At the most distal region from the hilus, a 5 mm biopsy D was taken with Palmer forceps. All biopsies and the ovary were measured. The ovaries were fixed and embedded in paraffin wax.

**Histology**

All biopsies were sliced at 5 μm intervals, mounted on slides and stained with haematoxylin and eosin.

All sections were examined under the light microscope at 100× magnification, and the number of primordial, early primary, 1–2 layers of granulosa, 3–4 layers, multilayered and Graafian (antral) follicles counted (Faddy et al., 1987). Only follicles where the nucleus could be seen were counted in order to prevent counting any one follicle twice. An attempt was made at following follicles from section to section to reduce the likelihood of double counting or of missing follicles with indistinct nuclei.

The remaining ovary was sliced at 5 μm intervals. Sections were sampled at regular intervals of one in 25 and stained. Every 50th section was examined, counted and grouped using the same method as for the biopsies, but all visible follicles were counted irrespective of the presence of nuclei. The distance between sections (250 μm) ensured that no follicle was counted twice. This was the reference for the total number of follicles in the ovary.

**Analysis**

**Whole ovary.** The maximum size of a primordial follicle is 50 μm. As sections were taken at 5 μm but only counted every 50th, potentially there were small follicles in the 250 μm space between observed sections. Each primordial follicle appears in up to 10 subsequent sections. This means only one primordial follicle was observed for every five that may have been present. The following conversion was therefore used to interpret the data (Zuckerman and Weir, 1977).

In the case of primordial follicles, with a maximum diameter of 50 μm, the correction was as follows:

\[ \text{5 μm/50 μm} \times \frac{1}{50} \times \text{no. of follicles/section} \]

In addition, the number of follicles within the biopsies must be added to the total.

**Biopsies.** The quantity of follicles recorded in the biopsies is an absolute number and no correction factor was applied. In order to make predictions regarding the number of follicles within the ovary, the surface area of the biopsy was taken as a proportion of the surface area of the ovary. This is because the cortex of the ovary lies at the surface and the medulla is virtually devoid of follicles.

The area of the biopsy was calculated as a circle (πr²) and the surface area of the whole ovary was calculated as an ellipsoid as illustrated below. The area of the hilus was subtracted from the total area of the ovary, as it is known to be devoid of oocytes. The area of the hilus was assumed to equal to 5 mm² in all cases.

The surface area of an ellipse = \[4π\left(\frac{r_1^2r_2^2 + r_1^2r_3^2 + r_2^2r_3^2}{3}\right)\]

Where \(r_{1-3}\) = the radii in the three axes of the ellipse.

The surface areas of the 2 and 5 mm biopsies were 3.14 and 19.64 mm², respectively.

**Results**

Although five ovaries were collected, following a thorough examination of three of them it became very apparent that even numerous biopsies could not predict the total number of follicles in an ovary. We therefore abandoned the detailed examination of further ovaries as this was thought to be a waste of considerable time and effort without a worthwhile return.

We present here the results from three ovaries. The first patient who donated an ovary (ovary A) to the study was 33 years old with two children and had a regular menstrual cycle at the time of oophorectomy. The reason for surgery was intractable, very severe lower abdominal pain, unresponsive to any treatment given. After extensive, multidisciplinary discussions, it was decided to meet the patient's request to perform a hysterectomy and bilateral oophorectomy. All organs removed were found to be perfectly normal on pathological examination.

The other ovaries (ovaries B and C) were from women aged 33 and 34 years, respectively, who underwent oophorectomy because of female to male transexuality. Both had received testosterone (250 mg i.m. once every 2 weeks) treatment for 3 months prior to the surgery. Their ovaries were enlarged and had a polycystic appearance.

Table I summarizes the numbers of follicles found in each of the biopsies and those found in the whole ovaries. The total counts in the whole of ovaries A and C entirely matched with their age, fitting exactly on the curve of normal values constructed by Faddy et al. (1992). Ovary B, however, was deviant, and the number of follicles counted was equivalent to that found in women around 40 years of age.

The numbers of follicles varied greatly across the various biopsies. A summary of predictions of follicle numbers based on biopsies is also given in Table I.

There was again great variation in the predicted total number, based on transformation of the biopsy count to the whole ovary. The average predicted number found in the 12 biopsies of 2 mm was also very deviant from the whole counts. Taking into account the ‘true’ total number of follicles found in the whole ovary, the predicted values based on the biopsies varied tremendously from +223% to −97% in ovary A, +3619% to −100% in ovary B and +492% to −42% in ovary C.

**Discussion**

Steele et al. (1970), the first to advocate the use of ovarian biopsy for the investigation of amenorrhoea, concluded that the tissue obtained was representative of the gonad as a whole, particularly with regard to its complement of germ cells, an
opinion supported by some subsequent authors (Sykes and Ginsburg, 1971; Black and Govan, 1972; Zographos et al., 1973; Egger, 1975; Fayez and Jonas, 1976; Motashaw et al., 1977). Some (Zographos et al., 1973; Egger, 1975) assumed one sample of at least 5 mm × 5 mm was necessary, while others argued that larger biopsies were required (Steele et al., 1977). Some (Zographos et al., 1973; Egger, 1975) assumed one sample of at least 5 mm × 5 mm was necessary, while others argued that larger biopsies were required (Steele et al., 1977; Sykes and Ginsburg, 1971).

Sutton (1974) was the first to seriously question the value of the ovarian biopsy and, based on his studies, he concluded at that time that it would be wrong to give the patient too grave a prognosis based on ovarian biopsy. He found that spontaneous pregnancies had occurred even in the complete absence of follicles within biopsy samples.

With advances in instrumentation and techniques, taking biopsies has become easier and more consistent. It remains, however, an invasive and, as such, hazardous, expensive and stressful operation.

It seems that the 1970s saw ovarian biopsy as an important advance in gynaecological diagnosis and prognosis, and since then it has become an issue of debate as to the reliability of these data (Khastgir et al., 1994; Wallach, 1995). Lass et al. (1997) found that ovarian volume was not correlated with follicular density in women under 35 years old but was highly correlated in women over 35 years old. They also found that follicular density decreases with advancing age. Women older than 35 had only a third of the concentration of follicles of younger women. FSH concentrations were not correlated with follicular density in women under 35 years old but was highly correlated in women over 35 years old. They also found that follicular density decreases with advancing age. Women older than 35 had only a third of the concentration of follicles of younger women. FSH concentrations were not correlated with follicular density. Recently, Schmidt et al. (2003) showed that follicle density varied greatly in small pieces of cortex, rendering information from biopsies unreliable.

| Opinion supported by some subsequent authors (Sykes and Ginsburg, 1971; Black and Govan, 1972; Zographos et al., 1973; Egger, 1975; Fayez and Jonas, 1976; Motashaw et al., 1977). Some (Zographos et al., 1973; Egger, 1975) assumed one sample of at least 5 mm × 5 mm was necessary, while others argued that larger biopsies were required (Steele et al., 1977; Sykes and Ginsburg, 1971). Sutton (1974) was the first to seriously question the value of the ovarian biopsy and, based on his studies, he concluded at that time that it would be wrong to give the patient too grave a prognosis based on ovarian biopsy. He found that spontaneous pregnancies had occurred even in the complete absence of follicles within biopsy samples. With advances in instrumentation and techniques, taking biopsies has become easier and more consistent. It remains, however, an invasive and, as such, hazardous, expensive and stressful operation. It seems that the 1970s saw ovarian biopsy as an important advance in gynaecological diagnosis and prognosis, and since then it has become an issue of debate as to the reliability of these data (Khastgir et al., 1994; Wallach, 1995). Lass et al. (1997) found that ovarian volume was not correlated with follicular density in women under 35 years old but was highly correlated in women over 35 years old. They also found that follicular density decreases with advancing age. Women older than 35 had only a third of the concentration of follicles of younger women. FSH concentrations were not correlated with follicular density. Recently, Schmidt et al. (2003) showed that follicle density varied greatly in small pieces of cortex, rendering information from biopsies unreliable. Is the lack of follicles in a single biopsy, or indeed any number of biopsies, enough to make accurate diagnoses or are they more misleading than prognosticating? If the data were reliable and could be used to counsel patients, then the risks associated with laparoscopic biopsy may be tolerable. We found that in individual cases, one or many more ‘standard’ biopsies could not reliably indicate the total numbers of follicles actually present in an ovary. In an attempt to estimate the number of follicles in the whole ovary, we converted biopsy information into total ovary information by transformation based on surface area. However, this approach did not yield acceptable results. One could argue that the surface area in the androgen-induced polycystic ovary (PCO)-like ovaries is enlarged and thus could affect the final predicted total using our method. The same would apply, however, to any method based on volume estimation. PCO is very common and, as such, the issue of estimating ovarian reserve also frequently applies. Every method in which biopsies are employed to predict a total need a translation via some morphometric characteristic. In ovaries, this is hazardous because their shape can rapidly change and is extremely dependent on their biological activity. From ovary A, which was most probably an entirely normal ovary, a number of biopsies had only a few follicles. This was a regular cycling 33-year-old woman who may well have been diagnosed with ovarian failure if these sparse biopsies had

Table I. Number of follicles

<table>
<thead>
<tr>
<th>Biopsy coordinates</th>
<th>Counted in biopsy</th>
<th>Total contents predicted from biopsy</th>
<th>Percentage deviation from count of whole ovary</th>
<th>Counted in biopsy</th>
<th>Total contents predicted from biopsy</th>
<th>Percentage deviation from count of whole ovary</th>
<th>Counted in biopsy</th>
<th>Total contents predicted from biopsy</th>
<th>Percentage deviation from count of whole ovary</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>95</td>
<td>36 860</td>
<td>+19</td>
<td>8</td>
<td>16 938</td>
<td>+572</td>
<td>92</td>
<td>174 536</td>
<td>+492</td>
</tr>
<tr>
<td>A2</td>
<td>20</td>
<td>7760</td>
<td>−75</td>
<td>13</td>
<td>27 524</td>
<td>+992</td>
<td>15</td>
<td>28 456</td>
<td>−3</td>
</tr>
<tr>
<td>A3</td>
<td>55</td>
<td>21 340</td>
<td>−30</td>
<td>5</td>
<td>10 586</td>
<td>+320</td>
<td>9</td>
<td>17 073</td>
<td>−42</td>
</tr>
<tr>
<td>A4</td>
<td>21</td>
<td>8148</td>
<td>−74</td>
<td>0</td>
<td>0</td>
<td>−100</td>
<td>32</td>
<td>57 709</td>
<td>+96</td>
</tr>
<tr>
<td>B1</td>
<td>46</td>
<td>17 848</td>
<td>−42</td>
<td>17</td>
<td>35 992</td>
<td>+1427</td>
<td>32</td>
<td>60 709</td>
<td>+106</td>
</tr>
<tr>
<td>B2</td>
<td>32</td>
<td>12 416</td>
<td>−60</td>
<td>0</td>
<td>0</td>
<td>−100</td>
<td>88</td>
<td>166 948</td>
<td>+466</td>
</tr>
<tr>
<td>B3</td>
<td>23</td>
<td>8924</td>
<td>−71</td>
<td>0</td>
<td>0</td>
<td>−100</td>
<td>36</td>
<td>68 296</td>
<td>+132</td>
</tr>
<tr>
<td>B4</td>
<td>8</td>
<td>3104</td>
<td>−90</td>
<td>6</td>
<td>12 703</td>
<td>+403</td>
<td>10</td>
<td>17 074</td>
<td>+42</td>
</tr>
<tr>
<td>C1</td>
<td>9</td>
<td>4392</td>
<td>−86</td>
<td>6</td>
<td>12 703</td>
<td>+403</td>
<td>10</td>
<td>18 972</td>
<td>−36</td>
</tr>
<tr>
<td>C2</td>
<td>257</td>
<td>99 716</td>
<td>+223</td>
<td>2</td>
<td>4234</td>
<td>+68</td>
<td>91</td>
<td>172 640</td>
<td>+486</td>
</tr>
<tr>
<td>C3</td>
<td>16</td>
<td>6208</td>
<td>−80</td>
<td>0</td>
<td>0</td>
<td>−100</td>
<td>9</td>
<td>17 074</td>
<td>+42</td>
</tr>
<tr>
<td>C4</td>
<td>2</td>
<td>776</td>
<td>−67</td>
<td>7</td>
<td>14 820</td>
<td>+488</td>
<td>0</td>
<td>136 433</td>
<td>+363</td>
</tr>
<tr>
<td>D</td>
<td>161</td>
<td>9982</td>
<td>−68</td>
<td>277</td>
<td>93 763</td>
<td>+3619</td>
<td>82</td>
<td>23 961</td>
<td>−19</td>
</tr>
</tbody>
</table>

Mean ± SD of all 2 mm biopsies

18 958 ± 27 289

11 292 ± 11 608

77 993 ± 65 700

a Primordial follicles, primary follicles, 1–2 layer follicles, 2–3 layer follicles, multilayer follicles and antral follicles.
b Estimate based on quantitative morphological analysis of fully sectioned ovary plus those counted in the biopsies.
c See Materials and methods for explanation of coordinates; biopsies A–C were 2 mm in diameter; biopsy D was 5 mm in diameter.
been taken, but who obviously had a fairly healthy store of follicles.

The quantity of oocytes in the biopsies was extremely variable even between closely related areas. This implies an inherent unreliability and unpredictability of individual biopsies.

In terms of absolute prediction, no region appeared to be more informative than any other. The mean of the 12 smaller biopsies proved no more accurate than individual biopsy data. The averages in ovary A and C are a reasonable assessment of reproductive potential. Whether women have ±20 000–70 000 follicles, as the estimate suggested, or ±30 000 as both the serial sections showed, they still seem able to reproduce. The average predicted contents in ovary B obviously overestimated the reproductive potential equivalent by some 10 years.

The original idea of the study was to take a maximum of 12 systematic small biopsies and investigate by computer randomization if a number of biopsies much less than 12 could be found which would approximate the true count and thus provide a clinically feasible procedure. However, since not even the average of 12 biopsies came close and because of the immense inter-biopsy variation that was apparent, it was decided not to go ahead with this analysis.

Cysts, any large follicles or recent corpora lutea as well as abnormal ovarian shape could lead to sampling difficulties. These are inherent to the organ. We took the biopsies following removal of the ovaries, but even under these ideal circumstances, sampling and processing appeared highly susceptible to error.

The whole ovary sections were of sufficient distance apart to ensure that no follicle was counted twice. In the biopsies, however, occasionally, oocytes would appear in a series of sections, disappear and no nucleus became apparent. An attempt was made at following follicles from section to section to reduce the likelihood of double counting or of missing follicles with indistinct nuclei. The methods described and used for predictions and estimations throughout the current investigation rely on correction factors which are based upon assumptions regarding cell size, orientation and shape. However, none of the quantitative morphology techniques will overcome the inherent inaccuracy resulting from the unequal distribution.

Unequal distribution of follicles across the cortex of the ovary is the most obvious explanation for the great interbiopsy variation of numbers of follicles per mm² surface area. Our observations are in agreement with those recently published by Schmidt et al. (2003). They measured cortical follicle density per fragment volume in several entire ovaries and this varied from 0.007 to 166 follicles/mm³.

There are several points in the above from which unequal distribution of oocytes may arise. The arrangement and growth pattern of the secondary sex cords into the ovary causes unequal dispersal of oocytes. The distribution of the follicles throughout the ovarian cortex from the stem cells is unlikely to be complete and will result in patches/nests of cells readily observable in the cortex. Blood supply has an influence and is an example of the fact that the ovary is an ever-changing organ. The presence of a corpus luteum will also affect the density. All these effects will contribute to the ‘patchiness’ of the follicular pattern and are likely to ensure that any biopsy taken will yield an inordinately large estimation if a nest is hit or a reserved estimation if they are missed.

An additional, important explanation for the lack of reliability of the follicle counts in the biopsy for predicting total counts is the fact that the very small numbers present in some specimens will give poor statistical precision in predicting the larger numbers because the impact of stochastic variation will be greater.

In conclusion, taking one or more ovarian biopsy does not seem to be the right procedure to estimate reliably the number of follicles in an ovary in the individual case.

This is largely the consequence of the unequal distribution of follicles across the surface of the ovary, in addition to technical shortcomings. Therefore, there is no place for this procedure in clinical evaluation of reproductive ageing. For research purposes, it should only be used with caution and only for estimation of statistics for patients in groups whose sizes are such that they would compensate for the inherent extreme intra-individual and inter-individual spread of the parameter (Webber et al., 2003).

Our negative experiences with this procedure are presented here in the context of the revival of interest in the subject and the subsequent debate. We hope that this contribution will add to the opinion that there is no place for the use of ovarian biopsy for the estimation of ovarian reserve.

Acknowledgements
We thank the personnel of the department of Cell Biology at the Vrije Universiteit, Medical Centre for their help in facilitating the histological processing.

References