A time to pregnancy questionnaire designed for long term recall: validity in Oxford, England

Michael Joffe, Laurence Villard, Zhimin Li, Rosalind Plowman, Martin Vessey

Abstract

Study objective – To establish the degree of validity of data on time to pregnancy, derived retrospectively using a short questionnaire.

Design – Information from the questionnaire was compared with data that had been collected concurrently from the same individuals.

Setting and participants – Questionnaires were mailed to 1647 women who continue to be followed up by the Oxford Family Planning Association contraceptive study, and a further 424 were approached for personal interview. Response rates were 91% and 79% respectively.

Main results – Matching was successful in 91% of pregnancies. Median recall time was 14 years (interquartile range, 11–16 years). At the group level, remarkably good agreement was found between the two sources of information, presented as cumulative percentage distributions of live births. The findings were at least as good with longer recall (>14 years) as with shorter recall. Digit performance was present to a limited degree. At the individual level, some misclassification was evident, which has implications for statistical power. For detection of clinical infertility (no conception within 12 months), the sensitivity was in the range 67%–91%, and the specificity was 92%–96%. Variations with format, duration of recall, age at delivery, year of birth, parity, social class, smoking habit, last contraceptive method, and outcome (live birth or not) were generally small, and were not statistically significant.

Conclusions – Time to pregnancy is a sensitive way of assessing reproductive function in either sex. Valid data at a group level can be derived retrospectively, with a long duration of recall, using a short questionnaire.

(J Epidemiol Community Health 1995;49:314–319)

There is increasing interest in fertility as a public health issue. It now seems clear that sperm counts have fallen by about 50% in the past half century, possibly due to a cohort effect. Clinical subfertility is a common problem: current estimates suggest that 10–20% of couples trying to conceive experience at least 12 months of inability to do so, in terms of achieving a clinically recognisable pregnancy.

However, the epidemiology of subfertility is in its infancy. There are few studies of its prevalence, and the degree of variation between different countries, regions, and subpopulations is unknown.

Time to pregnancy, in months or menstrual cycles, is a sensitive measure of the degree of fertility in either sex. For clinical purposes, a cut off point at 12 months is used conventionally, and this can easily be derived from these data. This variable can also be used to study the reproductive effects of environmental and occupational agents, by comparing populations which are exposed to such agents with non-exposed control populations. In the case of mutagenic agents, this promises to be a sensitive way of detecting germ cell mutagenesis, which could lead to heritable genetic defects, and is otherwise undetectable in humans.

Time to pregnancy can be measured either prospectively, or retrospectively by asking respondents. Prospective methods entail following subjects after recruitment until pregnancy occurs, as manifest either as a birth or other outcome of pregnancy, or by using a biological measurement such as human chorionic gonadotrophin (hCG). The use of recall is highly convenient, but raises the question of the validity of the resulting data. It has previously been shown that male factory workers in the printing and plastics industries are for the most part able to answer a questionnaire item on time to pregnancy, in the context of an interview administered study of reproductive effects of occupation, and that the resulting distribution closely resembles that expected from prospective studies. In the case of female workers, the predicted "infertile worker effect" was shown. In that study, the recall period was the worker's entire reproductive history, which ranged from a few years to more than 30 in some cases; acceptable data were obtained with 10 or even 20 years of recall.

This paper presents the findings from a validation study, comparing the reported and the recorded values corresponding to the same pregnancies, with duration of recall up to about 20 years. Previously published validation studies have focussed on much shorter recall periods, four years or less. The recorded values are derived from the database of the Oxford Family Planning Association contraceptive study (referred to throughout the paper as "the database"). The reported values have been obtained using a short questionnaire on time to pregnancy and subfertility, which was sent to subjects who are members of that study or used as the basis of an interview.
Methods

The Oxford Family Planning Association contraceptive study was started in 1968. Recruitment took place in 17 family planning clinics during the period 1968 to 1974, and a total of 17,032 women took part. Women were eligible if they were married, white British subjects aged 25 to 39 years old, and if they voluntarily agreed to participate. They were followed up annually thereafter, by means of a questionnaire that included items on change of contraceptive method and the occurrence of any birth, miscarriage, or other outcome of pregnancy during the preceding 12 months. The study has provided a number of the classic findings in the area of subfertility research, including the tendency for conception to be delayed after use of oral contraception.4

Beyond the age of 45, members of the cohort have continued to be followed up if they are still living in Britain and fall into one of two extreme categories: those who have never used the oral contraceptive pill, and those who have used it for at least eight years. Women were eligible for inclusion in the present study if they had had at least one pregnancy after recruitment into the cohort. To fulfil the aim of making comparisons within the research population, all women in the following categories were included, as their numbers were limited: (a) those who had had at least one pregnancy during the past 10 years, (b) those aged 50 or over, (c) those whose husbands had been semiskilled or unskilled manual workers at the time of recruitment, and (d) those who had ever had a miscarriage or stillbirth. A random selection was then made of the remaining cohort members. This gave a total of 2071 women from 16 of the clinics, 1647 of whom were sent a self completion questionnaire during 1990–91, and 424 of whom were approached for personal interview in early 1991. In the subset of study members who were asked to take part in the current research, 25% of their babies were born by early 1974, 50% by April 1976, and 75% by March 1979.

The basic questionnaire is attached as an appendix. In addition, a variant was used for approximately half the self completion questionnaires. This was designed to resemble that being used in a collaborative European study and aimed to assess the impact of minor changes on the accuracy of reporting. A comparison of the two versions found no important differences.12 The interview questionnaire was similar to the basic version but included additional questions that could help to improve the accuracy of response: following the key question on the number of months, “(If duration is 3 months or less) Do you mean that it was within the 1st, 2nd, 3rd, or 4th cycle?”; and, “How certain are you about this?”. Interviews were carried out by the study’s own research assistants, other clinic staff, or postgraduate research students.

Analysis

The database information exists in the form of the date (month/year) when contraception was stopped (time 0), and the dates of all births or other outcomes of pregnancy (time 1). For the database, the duration of time to pregnancy was derived by subtracting time 0 from time 1, giving the number of months up to the end of the pregnancy, and then subtracting the duration of the pregnancy, assumed to be nine months for births and three months for miscarriages (there was no specific information on gestational duration).

A complicating factor was that when this subtraction was performed, negative values for time to pregnancy were obtained in small number of instances, a logical impossibility in actuality. In some cases this was due to an immediate conception being followed by a preterm delivery, so that there were only seven or eight months between stopping contraception and the delivery of the child. In other cases, it was presumably due to a rounding error in the original data, for example cessation of contraception at the beginning of January (month 1) and delivery at the end of September (month 9), leading to a value of −1.

For the questionnaire, the duration of time to pregnancy was taken directly from the value given, in months. Pregnancies that were reported as having been conceived while contraception was being used, but “only sometimes”, were included in the analysis if a value was given for time to pregnancy. Values were excluded if contraception was being used more effectively than this, as contraceptive failures are not relevant in studying time to pregnancy.

This analysis is confined to live births, except where non-live births are specified, because the gestational lengths of pregnancies with other outcomes are subject to an unacceptable degree of variation, undermining the use of the database as a standard of comparison.

Pregnancies recorded in the database were matched with data from the questionnaires, and the values of time to pregnancy were then
non-smokers; etc), using the $\chi^2$ test to compare the extent of digit preference and the sensitivity and specificity.

**RESPONSE RATE**

In the self-completion sample, a total of 1498 subjects completed a questionnaire out of 1647 eligible, a response rate of 91.0%. In the interview sample, a total of 335 subjects completed a questionnaire out of 424 eligible, a response rate of 79.0%.

**Results**

The 1833 women reported a total of 3272 pregnancies occurring after the date of recruitment to the Oxford Family Planning Association contraceptive study, 2823 (86%) of which were conceived while not using an effective method of contraception.

Successful matching with database records was achieved in 2563 of these (91%); non-matching was relatively frequent for pregnancies with outcomes other than live births. In 2320 instances (91%, but only 84% in the case of non-live births), women were able to give a value for time to pregnancy. Altogether 1746 pregnancies (75%) were conceived after stopping regular contraception, and produced a live birth. The information was based on self-completion questionnaires in 1392 instances, and on an interview in 354.

Figure 1 shows the findings at the group level for the self-completion format (both versions combined), in the form of a comparison of the two sources of data, as cumulative percentages of conceptions in each month. The curves are almost superimposed for most of the distribution. A small degree of digit preference is evident at 3, 6, 12, and 24 months. The questionnaire has a slight tendency to underestimate the number of conceptions in months 1 and 2. From month 3 upwards, the correspondence between questionnaire and database is very close. A similar comparison using the interview data gave similar results (data not shown).

These findings are derived from matched pregnancies. It is possible that reporting is less good for unmatched pregnancies. In practice, when using a questionnaire, one would have to accept all the reported data, without any criterion being available for distinguishing the two, so that this could represent an important source of inaccuracy. The crude, unmatched distribution of pregnancies was therefore plotted in the same format, in figure 2. All pregnancies from both sources are included, as long as the outcome was recorded/reported as having been a live birth, and effective contraception was recorded/reported as having been stopped before conception. There is no evidence that the matched pregnancies represent an unduly selected population with superior reporting.

Further comparisons were made, within the self-completion sample, to investigate the effects of duration of recall, age at the time of the pregnancy, year of birth, parity, social class (according to the husband's occupation), compared. At a group level, the distributions were compared as cumulative percentages, and also the extent of digit preference in the questionnaire data was examined. At an individual level, discrepancies were examined in the number of months between the recorded and the reported values for each pregnancy; also the sensitivity and specificity of the questionnaire was calculated in relation to the identification of clinical subfertility (no conception within 12 months).

Comparisons were made between the self-completion and interview formats, and between subgroups of pregnancies (recall less than/at least 14 years; first/later pregnancies; smokers/
Long term recall of time to pregnancy

Table 1  Individual level comparison of (a) self completion questionnaire and (b) interviews with the database

<table>
<thead>
<tr>
<th>Database value</th>
<th>Underestimate (2 mth)</th>
<th>Within 1 mth</th>
<th>Overestimate (2 mth)</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Self completion questionnaire</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Database value (mth)</td>
<td>0</td>
<td>68</td>
<td>32</td>
<td>226</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>77</td>
<td>23</td>
<td>274</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>19</td>
<td>15</td>
<td>213</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>20</td>
<td>16</td>
<td>173</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>31</td>
<td>20</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>&gt;4</td>
<td>44</td>
<td>27</td>
<td>364</td>
</tr>
<tr>
<td>(b) Interview</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Database value (mth)</td>
<td>0</td>
<td>57</td>
<td>43</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>73</td>
<td>27</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>4</td>
<td>10</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>30</td>
<td>20</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>16</td>
<td>24</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>&gt;4</td>
<td>37</td>
<td>36</td>
<td>104</td>
</tr>
</tbody>
</table>

For each value as derived from the database (rows), the table shows the percentage of questionnaire values that were either in complete agreement with the database or only one month discrepant from it (second column), together with the percentage where the questionnaire differed from the database value by at least two months, either as an underestimate (first column) or an overestimate (third column). Above a database value of 4, all values have been grouped, as the numbers were small.

Table 2  Digit preference, sensitivity, and specificity for selected characteristics

<table>
<thead>
<tr>
<th>No</th>
<th>Digit preference up to 12 mth only</th>
<th>Validity in ascertainment of clinical infertility</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Preferred values (%)</td>
<td>Sensitivity (%)</td>
</tr>
<tr>
<td>----</td>
<td>----------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>Format</td>
<td>Self completion</td>
<td>1392</td>
</tr>
<tr>
<td></td>
<td>Interview</td>
<td>354</td>
</tr>
<tr>
<td>Duration of recall</td>
<td>Less than 14 years</td>
<td>510</td>
</tr>
<tr>
<td></td>
<td>14 years or more</td>
<td>882</td>
</tr>
<tr>
<td>Age at delivery</td>
<td>Less than 32 years</td>
<td>808</td>
</tr>
<tr>
<td></td>
<td>32 years or more</td>
<td>584</td>
</tr>
<tr>
<td>Year of birth</td>
<td>Before 1946</td>
<td>729</td>
</tr>
<tr>
<td></td>
<td>1946 or later</td>
<td>663</td>
</tr>
<tr>
<td>Parity</td>
<td>First pregnancy</td>
<td>513</td>
</tr>
<tr>
<td></td>
<td>Later pregnancy</td>
<td>879</td>
</tr>
<tr>
<td>Social class a</td>
<td>I &amp; II</td>
<td>736</td>
</tr>
<tr>
<td></td>
<td>III, IV &amp; V</td>
<td>656</td>
</tr>
<tr>
<td></td>
<td>IV &amp; V only</td>
<td>78</td>
</tr>
<tr>
<td>Smoking habit</td>
<td>Yes</td>
<td>326</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>1046</td>
</tr>
<tr>
<td>Last contraceptive method</td>
<td>Oral</td>
<td>277</td>
</tr>
<tr>
<td></td>
<td>Not oral</td>
<td>576</td>
</tr>
<tr>
<td>Outcome</td>
<td>Live birth</td>
<td>1392</td>
</tr>
<tr>
<td></td>
<td>Non-live birth</td>
<td>187</td>
</tr>
</tbody>
</table>

* Social class I & II denote professional and managerial workers; III denotes clerical and skilled manual workers; IV & V denote semi- and un-skilled manual workers, respectively.

Discussion

In both the interview and self completion formats, a high degree of agreement was found between the questionnaire and the database in the comparisons at a group level. The finding for a recall period greater than 14 years was remarkable, and unexpected.

The extent of digit preference was limited, and only marginally higher in the self completion format. It is therefore clearly useful to ask respondents to give the actual number of months, rather than grouped categories (for example, up to 3 months; 4 to 6 months; 7 to 12 months, etc) as has sometimes been done on the assumption that only crude estimates are obtainable.

At the individual level, a degree of misclassification occurred which is important even for group comparisons, because of the effect on statistical power. Theoretically, studies can be carried out with a rather small sample in each exposure defined study group: a figure as low as 100, half of whom are exposed, has been suggested as sufficient to detect a 50% reduction in fecundability. However, that calculation depended on a lower level of misclassification than was observed here. A previous study found less misclassification with a duration of recall up to only four years, but used a detailed retrospective interview as the standard of comparison.

Recall of time to pregnancy does not depend on the occurrence of a perceived problem, in fact, the reverse seems to be true: the degree of misclassification was higher for larger values of the variable. This finding is fortunate, in that the majority of conceptions occur within the first few months, so that this is where the statistical power lies, and secondly, the influence of medical treatment is not a factor in this early part of the distribution. This reinforces the view that research on time to pregnancy is best regarded not as a comparison of proportions of a population who have a clinical problem, but as a comparison of whole distributions of this variable between population groups defined in terms of exposure characteristics.

In terms of the reporting of clinical infertility, both formats of the questionnaire performed reasonably well: sensitivities ranged from 67 to 91%, and all specificities were over 90%.

smoking habit, and last method of contraception (oral or not).

In general, the cumulative percentage graphs for each subgroup were little different from figure 1. As an example, figure 3 shows the findings for duration of recall longer than 14 years (the median). The correspondence between the curves for database and questionnaire is at least as good as it was for the whole population.

Results at the individual level are shown in table 1, separately for self completion questionnaires and interviews. For each database value from 0 to 4 months, then in aggregated values for more than 4 months, the percentage of questionnaire values is shown which was either close (within one month), or an underestimate, or an overestimate. Some misclassification is evident in both formats, neither being clearly superior to the other: the degree was greater with relatively high values of time to pregnancy.

The additional questions in the interview format (see above) did not lead to any perceptible gain in accuracy, and resulted in smaller numbers being available for analysis (data available on request).

Table 2 examines digit preference, and the validity (sensitivity and specificity) of the questionnaires for detecting clinical subfertility. For digit preference, the table gives the percentage of all values up to 12 months which were given as "preferred" values (3, 6 and 12). Some differences are apparent, but none was statistically significant.
However, the positive predictive value was below 50%.

The almost equally good performance of the self-completion questionnaire differs from the findings of Zielhuis et al. In that study, the duration of recall was only two years or less. Our interviewers commented that women often appeared to be searching their memories for the information, and unable to access it in the short time available. In this respect, self-completion questionnaires could have an advantage. One approach, unfortunately not tried in the present study, is to provide a list of the items (not the actual questions) to be covered in the interview, before it occurs, enabling the respondent to prepare mentally. This method has been used successfully before: male factory workers were able to give replies enabling an apparently valid time to pregnancy distribution to be constructed.7

As frequently happens in validation studies, close examination of data from the database has shown that it is not perfect as a gold standard. In general, as discrepancies between database and questionnaire could sometimes reflect imprecision in the database rather than the questionnaire, the results reported here may underestimate the true degree of validity of the questionnaire data.

On the other hand, it is possible that the present population is somewhat more accurate than others: these women have been accustomed to completing a short form each year for many years, covering pregnancies and changes in contraceptive method, with dates. However, they have never been asked questions on time to pregnancy, or on subfertility more generally.

The findings reported here apply only to live births. There were too few miscarriages and stillbirths to give stable estimates. As the ascertainment and the timing are less good retrospectively for miscarriages than for births, and also the time to pregnancy distribution is shifted towards the right (less fertile) for pregnancies ending in a miscarriage compared with those ending in a live birth,9 it seems advisable to standardise research, so that studies using time to pregnancy as the outcome variable are restricted to pregnancies that ended in a birth.

Biological mechanisms do not correspond directly with manifest outcomes of pregnancy.15 One advantage of a research design that is formulated in terms of exposure characteristics (a retrospective cohort study in the present case), is that multiple outcomes can be studied. As an exposure that affects time to pregnancy may also affect the risk of miscarriage, studies could also include as an additional outcome variable the proportion of recognised pregnancies that ended in miscarriage, with appropriate precautions being taken to ensure valid data.16

Previous research on the validity of recall of birth weight and gestational age has shown that recall periods of up to 20 years are feasible for epidemiological research.17 The present study indicates that this time to pregnancy questionnaire can also be used retrospectively with a recall time of comparable magnitude. This is an important strength, because for a defined cohort such as an occupational category, a large proportion of subjects will have had (or fathered) at least one pregnancy, making generation of an adequate sample size comparatively easy.

Appendix

For Each Pregnancy:

1 Did this pregnancy end in a live birth, or a miscarriage, or what? (live birth, stillbirth, miscarriage, ectopic or tubal pregnancy, other)

2 When did it end? (ie date of the child’s birth, or of miscarriage etc) (day, month, year)

For Live Births and Stillbirths Only (omit twins):

3 How much did this baby weigh? (Ibs, oz)

4 During the month this pregnancy was conceived, were you or your husband using any form of birth control? (Yes, No)

If No: please go to question 7

If Yes:

5 What method of birth control were you using? (oral (the pill), intrauterine device (coil, loop), condoms, diaphragm (cap), rhythm or withdrawal, other)

6 Were you using this all the time, nearly all the time, or only sometimes? (only sometimes, nearly all the time, all the time)

If all the time or nearly all the time: please go on to the next pregnancy.

If No or only sometimes:

7 How many months did it take you to get pregnant? (please write in)

Please Go on to the Next Pregnancy.

This study was supported by a grant from Birthright. We would like to thank Dr. Rosemary Painter, Jane Wadsworth, Luke Whitaker, and Dr. David Yeates for their help with the computing and statistics. The research would not have been possible without the cooperation of the research assistants who carry out the annual data collection for the Oxford Family Planning Association contraceptive study.

References

10. Zielhuis GA, Hulscher MEJL, Florack EML. Validity and
Long term recall of time to pregnancy


