Feasibility of studying subfertility using retrospective self reports

MICHAEL JOFFE
From St Mary’s Hospital Medical School, London.

ABSTRACT During an investigation of possible reproductive effects of environmental agents, 261 male and 155 female workers were interviewed concerning subfertility at some time in the past: the time taken to conceive, for all births; and the occurrence of one or more infertile phases lasting for 6 months or more. When these two variables were compared, the quality of reporting was acceptable in 89-7% of instances, and data editing enabled accuracy to be improved. Reporting was more reliable with shorter duration of recall, and female workers’ reports were somewhat more reliable than those of male workers. The distribution of time taken to conceive was similar for male workers to that observed in previously published prospective series, though with a higher estimate of subfertility when infertile phases were also considered. Comparison with published estimates of reduced fertility appeared to be reassuring. As predicted, the equivalent comparisons for female workers showed the presence of a strong selection effect.

Impaired fertility is a good indicator of the reproductive toxicity of environmental agents, involving a series of biological processes from sperm and egg development to early (unrecognised) fetal loss. As subfertility may be a result of mutagenic activity, and as fetotoxicity may also be manifest as an increased frequency of congenital malformations, the study of subfertility may have implications which go beyond its own importance as a disturbance of reproductive function.

The majority of fertilised ova are lost before any manifestations of pregnancy can be recognised: in prospective studies, the rate of early fetal loss has been estimated as greater than 40%, at least three quarters of which are unrecognised.1 2

Methods of study have included biological approaches such as sperm examination3 and measurement of human chorionic gonadotropin,4 and documentation of the occurrence and spacing of births compared with US national statistics.5 6

Comparatively little is known about the factors influencing the ability to conceive. Epidemiological investigation is at a relatively early stage: studies based on volunteer samples are still being published in reputable journals.7-9

If the outcome of interest is taken to be the ability of a couple to conceive when trying to do so, many different approaches are possible in carrying out population based research. Clinical infertility, defined as a delay in conception of at least 1 year, can be studied epidemiologically, for example using a case-control design.10 Other measures have included the proportion who are unable to conceive after 2 years11 or 3 years12 13 of unprotected intercourse, or alternatively, the absence of a live birth after a specified period of time.14 Sterility, defined in terms of total inability either to conceive12 13 or to have a live birth,14 is considerably rarer.

The use of time taken to conceive has been advocated by Baird et al as a suitable outcome variable for epidemiological research concerning environmental agents which might affect reproduction.8 This is a measure of “fecundability,” the probability of pregnancy in each menstrual cycle. Calculations of statistical power lead to the estimate that a 50% reduction in mean fecundability could be detected with 80% probability using data from 55 exposed and 55 non-exposed couples, a sample size which is easily achievable in retrospective studies. The authors tested the feasibility of using retrospective reports of time to pregnancy for epidemiological studies, by comparing the distribution so obtained with published data from a prospective study. Their research sample was 681 pregnant women who volunteered to take part, and the recall period did not exceed 2-5 years.

The research reported here investigated the possibility of using a similar approach in the context of a retrospective survey of an occupationally defined population. However the measurement of conception
Feasibility of studying subfertility using retrospective self reports

time proposed is considerably simpler than that proposed by Baird et al., requiring only one additional questionnaire item for each birth. This less sophisticated technique is justifiable if at least one of the following two assumptions be true: that relatively subtle influences (cycle length, sporadic use of contraception, periods of sexual abstinence due to illness or travel) are quantitatively unimportant when groups rather than individuals are considered; or that they balance out in the comparison of different groups.

A measure of infertility is also used, the occurrence of phases of infertility lasting for at least 6 months, the durations of which are elicited. This has an advantage for epidemiological purposes over the dichotomous 12 months criterion of clinical infertility, as it includes milder degrees of dysfunction while also allowing the full range of impairment to be studied; analysis can be performed using a longer cut off point if desired for comparability. Further, the continuum of infertile phases overlaps substantially with that of conception times, so that two measures are available for all infertile phases lasting at least 6 months which lead to a birth.

The present paper assesses the quality of the information thus obtained in two ways: first by means of an internal comparison of the two variables, reports of conception times with reports of infertile phases; and second by means of a comparison with published data, both of the distribution of conception times and of other estimates of reduced fertility. As part of the internal comparison, the reporting of male and female workers is assessed separately, and the effect of duration of recall is examined. The recall periods are long, in some instances longer than 30 years.

It has previously been suggested that any occupationally recruited sample of women is likely to be subject to the "infertile worker effect"12: because a large proportion of women leave the paid workforce at least temporarily after the birth of a child, an employed female population will tend to contain a relative excess of infertile women, some of whom are involuntarily infertile. The reports of female workers are therefore predicted to contain higher estimates of subfertility than those of male workers.

Methods

POPULATION
The sample consisted of the current employees of five factories in the printing and plastic industries. Plant A was a small printing ink manufacturing company which employed 37 people. Plants B and C were printing companies which employed 114 and 238 people respectively. Plants D and E belonged to the same company and were engaged in the manufacture of plastic cartons and in printing thereon; they employed 109 and 36 people respectively.

METHOD
The study was carried out using the standard method previously developed by the author, designed especially for use in research on the reproductive effects of occupational exposures. This includes a standard questionnaire to collect information retrospectively from occupationally defined populations. It consists of a reproductive history, a brief occupational history, and information on other factors which could confound the analysis of the relationship of outcome and exposure. It is available in separate versions for male and female workers. Although designed to be used also in a self completion format, the research reported here was based on interviews. This is the preferred method of administration, in order to achieve an optimal response rate. Data on exposure were obtained from the participating companies, as well as from the interviews.

The questionnaire contains two items relevant to subfertility. For each birth the respondent is asked the question, "For how long had you both been trying to conceive before you/your wife or partner became pregnant with this child—or were you not planning to have a child at that time?" This is coded as ≤3 months, >3≤6 months, >6≤12 months, >1≤2 years, and >2 years, with an additional code for "wasn't planning to have a child then".

Secondly, subjects are asked about phases of infertility which they may have experienced, using a sentence to set the context for the question: "Some couples find that they do not conceive as quickly as they would like. Have you ever gone 6 months or longer when you were having intercourse regularly without contraception, but you/your wife or partner did not get pregnant?" For up to three such infertile phases, respondents are asked the dates when each began and ended, to the nearest month. For those phases which are not still continuing, the reason is asked why they came to an end: "became pregnant", "relationship ended" and "other".

The two questions are asked in separate subsections of the questionnaire. There are four intervening questions. Interviewers have not been instructed to cross check replies from the later question with information given previously.

STATISTICAL ANALYSIS
For the comparison of male and female workers' reports, the χ2 test was used. In examining duration of recall, the χ2 test for trend in proportions was used.
Results

RESPONSE RATES

The response rate was 78%. Of the 22% non-respondents, 17% refused to take part and 5% were on leave.

A total of 416 people were interviewed, 261 men and 155 women. The number of births was 814, of which 456 were reported as “unplanned”, 127 subjects (30.5%) reported at least one infertile phase which had lasted 6 months or longer. Many births and infertile phases had occurred before the respondent started work in the current job, or in any similar job. These are included in the present report, which is not concerned with occupational exposures.

INTERNAL COMPARISON

If reporting were perfect, all phases of infertility lasting at least 6 months would have been reported in reply to the question on infertile phases, and in addition all those which ended in births would have been reported in terms of the time taken to conceive. There were 214 such reports from at least one of these two sources. These were tabulated, the completeness of the information was examined, and the duration and the time of occurrence were compared. A further comparison was carried out, with other births and with miscarriages, in order to discover whether a pregnancy had been mistakenly reported as occurring at the same time as an infertile phase.

The results are shown in table 1. The information obtained was perfectly consistent and did not lack any data in 37.9% of instances, and in 27.1% the pairs of reports contained minor discrepancies between the replies to the two questions; a typical error was to report the end of an infertile phase not as the date of conception but as the date of birth, which could be corrected by cross checking against information on births. In 17.3% of cases the reports were incomplete. In 7.5% of instances a conception reported as “unplanned”, and therefore without a conception time, coincided with the report of an infertile phase, and there was no discrepancy in any other respect. In 10.3% of instances the information from different sources was in sufficiently serious conflict that the report needed to be excluded from analysis.

QUALITY OF REPORTING: GENDER AND DURATION OF RECALL

Table 2 shows a comparison of the quality of reporting by male and female workers. Quality of reporting was judged by checking reports of infertile phases against information on pregnancies, as described above. The proportion of serious discrepancies was similar for men and women. Women were more likely to give perfectly concordant results than men, or to give accounts containing minor discrepancies. Men’s reports were correspondingly more liable to contain missing information, and to report as “unplanned” pregnancies with conception times in excess of 6 months.

Table 2 Accuracy of reporting: comparison of men and women

<table>
<thead>
<tr>
<th></th>
<th>Male workers</th>
<th>Female workers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No</td>
<td>%</td>
</tr>
<tr>
<td>Reports agree perfectly</td>
<td>41</td>
<td>32.8</td>
</tr>
<tr>
<td>Minor discrepancy</td>
<td>26</td>
<td>20.8</td>
</tr>
<tr>
<td>Missing information</td>
<td>32</td>
<td>25.6</td>
</tr>
<tr>
<td>Reported as “unplanned”</td>
<td>13</td>
<td>10.4</td>
</tr>
<tr>
<td>Serious discrepancy: excluded</td>
<td>13</td>
<td>10.4</td>
</tr>
<tr>
<td>Total</td>
<td>125</td>
<td>100.0</td>
</tr>
</tbody>
</table>

* p < 0.01; † p < 0.025; ‡ p < 0.001

A similar comparison is shown in table 3, for the length of time which had elapsed between the event reported and the date of interview, calculated from the beginning of the attempt at conception. Perfect concordance was strongly time related, varying from 60.5% for events within the past 10 years to 14.8% for those which occurred more than 30 years ago. Discrepancies were also clearly time related: minor discrepancies varied from 13.2-48.1%, and serious discrepancies from 2.6-15.0%. No trend was observed in the tendency to omit information or to report pregnancies as unplanned.

BIRTHS: DISTRIBUTION OF TIME TO CONCEPTION

Figures 1 and 2 show comparisons of data from the present study, separately for male and female workers, with two previously published prospective studies of women who had stopped using contraception.16 17 Aggregate figures have been used, combining all parities and all methods of prior contraception.

In fig 1, the curves for male and female workers are based only on replies to the question concerning time
Feasibility of studying subfertility using retrospective self reports

Table 3 Accuracy of reporting: duration of recall

<table>
<thead>
<tr>
<th></th>
<th>Over 30 years (n = 27)</th>
<th>21–30 years (n = 80)</th>
<th>11–20 years (n = 69)</th>
<th>Within 10 years (n = 38)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td>Reports agree perfectly</td>
<td>14·8</td>
<td>32·5</td>
<td>40·6</td>
<td>60·5 †</td>
</tr>
<tr>
<td>Minor discrepancy</td>
<td>48·1</td>
<td>30·0</td>
<td>23·2</td>
<td>13·2 †</td>
</tr>
<tr>
<td>Missing information</td>
<td>18·5</td>
<td>16·3</td>
<td>17·4</td>
<td>18·4</td>
</tr>
<tr>
<td>Reported as “unplanned”</td>
<td>7·4</td>
<td>6·3</td>
<td>10·1</td>
<td>5·3</td>
</tr>
<tr>
<td>Serious discrepancy: excluded</td>
<td>11·1</td>
<td>15·0</td>
<td>8·7</td>
<td>2·6 *</td>
</tr>
<tr>
<td>Total</td>
<td>100·0</td>
<td>100·0</td>
<td>100·0</td>
<td>100·0</td>
</tr>
</tbody>
</table>

* p < 0·10; † p < 0·01; ‡ p < 0·001

available on the time taken to conceive using the variable concerning infertile phases.

Inaccurate reporting might be recognisable in either or both of two ways: the shape of the curve might differ from those derived from the prospective studies; and there might be higher or lower estimates of the proportion of the population who have conceived at various stages.

Figure 1 shows that for male workers, the curves are remarkably similar and the estimates corresponded to the prospective studies up to a conception time of 12 months, but at 24 months there was a relative deficit in reported conceptions. In the case of female workers, though the shape of the curve is comparable, the estimates of the proportion conceiving were consistently lower throughout the range.

In figure 2, while the shapes of the distributions remain comparable, the estimates of impaired fertility are increased. In the case of male workers, this was apparent by 12 months, and the proportion who had not conceived at the end of 24 months was approximately double that obtained in the long term prospective study. Female workers reported substantially more subfertility.

ESTIMATES OF REDUCED FECUNDITY

Four comparisons are shown in table 4, three of them concerned with first births (primary) and one with subsequent births (secondary). The concepts “sterility”, “infertility” and “reduced fecundity” are respectively intended to convey total life long inability to bear a child, long term infertility which was not necessarily permanent, and inability to conceive which lasted for 2 years or more. The data from the present study are compared with the US National Survey of Family Growth; with an English historical study of births, which is assumed to represent “natural” fertility for that population; and with a national survey carried out in Denmark. Definitions are given in the table, corresponding to each estimate. Square brackets enclose those estimates for which the definition used in that study differed substantially.

taken to conceive. Births reported as unplanned were not included. Figure 2 shows the distribution when all phases of infertility have been included, whether or not they ended with conception, after seriously conflicting reports have been excluded. Conceptions reported as “unplanned” have been included if information was
from the concept forming the basis for the comparison.

The estimates for male workers were found to be similar to those derived from the other studies, with a slight tendency for Primary Reduced Fecundity to be more frequent and Secondary Reduced Fecundity less frequent. In the case of female workers, all the estimates of reduced fertility were substantially increased in comparison with those from all the other studies.

Discussion

The results of the study are considered from the viewpoint of the ability of the method to generate data of acceptable quality.

Consistency of reporting was examined by comparing conception times, reports of infertile phases and other information on the timing of pregnancies. In 89-7% of instances the information was of acceptable quality, though omissions and minor discrepancies were not uncommon, and in some cases conceptions which had occurred after more than 6 months were reported as unplanned. These would be missed if the questionnaire contained only the single item on time taken to conceive.

A second justification for the inclusion of infertile phases as well as conception times is that comparison of reports makes it possible to exclude those which contain serious discrepancies. This internal quality check also enables editing of data to be carried out, for example if a respondent has mistakenly added the 9 months of pregnancy to the infertile phase. It would be possible for interviewers to carry out a quick consistency check during the interview, and thereby to assist respondents in their efforts to recall past events.

The third and most important reason for studying infertile phases directly is that conception times alone provide an underestimate of subfertility.

There are three possible reasons for the apparent relative excess of subfertility found in male worker's reports in the present study: a true excess of subfertility in this population, bias in the comparison series, and bias in the present method.

A true excess of subfertility might possibly be present owing to a social class effect, since the occupations represented here were all manual. It could not have been due to occupational exposures, as similar distributions were obtained when the data were restricted to the period prior to starting work in that factory or any previous similar job.

Bias in the comparison studies could theoretically have occurred as a result of loss to follow up, especially in the prospective study of relatively long duration, but in practice this was not of sufficient magnitude to have produced the observed difference in rates of subfertility. A more plausible possible explanation is that attenders of family planning clinics may be more fertile compared with the general population.

The possibility remains that the present method overestimates subfertility to some extent, for example owing to the exclusion of pregnancies reported as unplanned (see below). This would be important if it differentially affected different research populations, as comparisons would then be erroneously exaggerated or diminished. A bias of this type has been

### Table 4 Comparison of estimates of reduced fertility

<table>
<thead>
<tr>
<th></th>
<th>Primary &quot;sterility&quot; %</th>
<th>Primary &quot;infertility&quot; %</th>
<th>Primary reduced fecundity %</th>
<th>Secondary reduced fecundity %</th>
</tr>
</thead>
<tbody>
<tr>
<td>US National Survey of Family Growth(^{13})</td>
<td>[2.6(^{a})]</td>
<td>[3.3(^{b})]</td>
<td>10.9(^{a})</td>
<td>5.3(^{b})</td>
</tr>
<tr>
<td>English historical data(^{14})</td>
<td>4.6(^{b})</td>
<td>[23.4(^{b})]</td>
<td>8.9(^{b})</td>
<td>9.0(^{b})</td>
</tr>
<tr>
<td>Danish national survey(^{11})</td>
<td>Male workers</td>
<td>2.6(^{c})</td>
<td>3.3(^{e})</td>
<td>12.1(^{h})</td>
</tr>
<tr>
<td>Female workers</td>
<td>9.6(^{c})</td>
<td>15.4(^{e})</td>
<td>29.8(^{b})</td>
<td>13.5(^{i})</td>
</tr>
</tbody>
</table>

See text for explanation of square brackets.

- \(^{a}\) Non-surgically sterile, para 0, as defined by respondents; surgically sterile subjects excluded from the population at risk
- \(^{b}\) Never had a child
- \(^{c}\) At least 8 years' exposure; no births reported
- \(^{d}\) Same as \(^{a}\), plus those who did not conceive within 3 years of marriage (unprotected intercourse), para 0
- \(^{e}\) At least 2 years' exposure with no conception, para 0; no conception even at the end of this phase
- \(^{f}\) Same as \(^{d}\), but currently married women only
- \(^{g}\) No birth within 2 years of marriage
- \(^{h}\) At least 2 years' exposure with no conception, para 0; includes infertile phases which ended with conception as well as those classified in \(^{e}\)
- \(^{i}\) Same as \(^{f}\), but para 1 or more
- \(^{j}\) Same as \(^{h}\), but para 1 or more
observed in occupational studies of miscarriages\textsuperscript{18,19}; workers designated as non-exposed tended systematically to underreport miscarriages. One solution is a research strategy which does not rely on the use of non-exposed workers as a control group.\textsuperscript{20}

The comparison of estimates of reduced fertility suggested that those for male workers appeared to be comparable with estimates derived from other studies, but the comparison is beset with problems of definition. The Danish national study\textsuperscript{11} used similar definitions as the present investigation but had the same limitation, that it was based on retrospectively collected data, albeit with a shorter recall period. The other two comparisons have the strength that the studies concerned utilised different research methods, but the differences in definition have to be taken into account.

The predicted infertile worker effect was observed in the present study. The magnitude of this bias is likely to vary with differing patterns of women's employment, depending on such factors as the availability of child care.

There are at least two ways of overcoming the infertile worker effect. The preferable method is to include former workers in the sample, which depends on the availability of accurate lists, efficiency of tracing, and success in recruitment to the study. An alternative method is to design the research so that comparison is made with current female workers in non-exposed populations, but this depends heavily on the assumption that the two groups have had similar patterns of age at child bearing and of subsequent work related behaviour.

Out of 814 births, 456 (56\%) were reported as unplanned. Conception times were thus not available, except for births which could be matched with reports of an infertile phase. It is probable that a significant proportion of "unplanned" conceptions were not unwanted nor entirely unexpected, so that their designation as unplanned may reflect different cultural approaches to family planning. This issue could be addressed by a modification of the questionnaire, separating the issues of plannedness and time taken to conceive by dividing the present questionnaire item into two parts, with the intention of obtaining a reply on conception time from those reporting a pregnancy as unplanned, whenever this is possible.

The unit of analysis for all the results presented here has been the pregnancy in the case of conception times, and all exposures to possible pregnancy in the case of infertile phases. Many respondents reported two or more events, and these have been included in the analysis. The assumption of independence implicit in this method ignores the likely intercorrelation of values at different times for each couple. This might be expected to reduce the dispersion of values, and although no major impact would be expected on these results, such an assumption would not be justified when undertaking a statistical comparison of subgroups, for example of workers who have been exposed to different environmental agents. The appropriate method of statistical analysis of fecundability utilises the beta distribution.\textsuperscript{21}

In conclusion, studies of male workers can include retrospectively collected information on subfertility, in the context of research into the possible effects of occupational exposures on reproductive processes. The use of just two questionnaire items generates data which appear to be of sufficient reliability to allow comparisons of workers exposed to different environmental agents, though a problem remains concerning pregnancies reported as unplanned. Caution is required if the recall period is long, especially more than 20 years. The possibility of selective recall, for example relative underreporting of problems by workers designated as non-exposed, needs also to be borne in mind. Data from female workers are of relatively better quality, but present problems for study design and/or analysis owing to a strong selection effect.

The work was supported by a European Commission project grant. I would like to thank all those at the five participating factories who helped with the research; as well as Claire-Marie Fortin, Sheila Williams and Eva Alberman.

Address for correspondence and reprints: Dr Michael Joffe, Lecturer in Community Medicine, St Mary's Hospital Medical School, London W2 1PG.

References


Accepted for publication February 1989