



American Journal of EPIDEMIOLOGY

Volume 157

Number 2

January 15, 2003

Copyright © 2003 by The Johns Hopkins

Bloomberg School of Public Health

Sponsored by the Society for Epidemiologic Research

Published by Oxford University Press

COMMENTARIES

Invited Commentary: The Potential for Monitoring of Fecundity and the Remaining Challenges

Michael Joffe

From the Department of Epidemiology and Public Health, Imperial College Faculty of Medicine, London, United Kingdom.

Received for publication December 5, 2001; accepted for publication October 31, 2002.

In the epidemiologic study of reproductive capacity, the assessment of fecundity as a functional measure is complementary to approaches that focus on biomedical mechanisms and/or that use biomarkers such as semen quality. More research is needed on time trends, spatial patterns, and particular groups, especially those exposed to potentially toxic agents. Although specific research projects will always be important, much could be gained by general population surveillance, which could be introduced into existing multipurpose surveys and repeated periodically. The core measurement would be time to pregnancy, which can be carried out using a short, acceptable questionnaire that has good validity at the group level. This should be accompanied by questions on time periods of unprotected intercourse that do not end with conception, to avoid bias resulting from exclusion of relatively infertile couples. Information is also required on contraceptive failures, recent contraceptive use, and other covariates and possibly on behavioral variables, such as the degree of planning and persistence in trying to conceive, and couples' knowledge of fertile days of the menstrual cycle. Existing statistical methods can deal with possible biases due to "accidental" pregnancies and the effects of fertility treatment. Further methodological work is needed to avoid or measure more subtle biases, for example, to determine the best way to deal with pregnancies occurring to couples whose approach to family formation is relaxed, for whom the concept of "pregnancy planning" does not apply.

data collection; fertility; infertility; monitoring, physiologic; reproduction

Fecundity is the biologic ability to conceive given unprotected heterosexual intercourse. The unit of analysis is the couple, and the degree of fecundity depends on both partners. Until recently, little was known about the determinants of fecundity, but this situation has been altered by the development and validation of a method using the time taken to conceive (time to pregnancy). This is a functional measure, corresponding not to any specific biologic process but rather to the final common path of conception.

This paper briefly considers some methodological issues involved in monitoring the fecundity of a population. Its focus is on the feasibility of obtaining satisfactory estimates, and it does not consider some of the more technical issues, particularly those concerning how best to carry out the statistical analyses.

Biologic measures are also omitted from consideration, notably the assessment of semen quality; it is difficult to justify its use for population monitoring, as participation rates are very low (e.g., 30 percent), and bias occurs because

Reprint requests to Dr. Michael Joffe, Department of Epidemiology and Public Health, Imperial College Faculty of Medicine, St. Mary's Campus, Norfolk Place, London W2 1PG, United Kingdom (e-mail: m.joffe@ic.ac.uk).

motivation to take part tends to depend on experience or suspicion of subfecundity. Other biomarkers are also probably not appropriate for monitoring, as they are likely to indicate impairment in a specific biologic pathway rather than in fecundity as a whole.

Nevertheless it should be remembered that time to pregnancy cannot be regarded as the sole criterion of male reproductive health, as male-mediated toxicity could occur that does not alter fecundity but does adversely affect the offspring. Similar remarks apply also to female-mediated toxicity, but here no equivalent to semen quality exists.

METHODOLOGICAL ISSUES

Sampling and overall strategy

Prospective study of fecundity is difficult. Often no satisfactory sampling frame exists. Women or couples of reproductive age can be recruited, for example, from age-sex registers in general medical practice or from trade union records, but the proportion of respondents who give advance notice of starting unprotected intercourse is small (1). Another example is in the occupational context, but here the issue of intended parenthood may be sensitive enough to discourage cooperation. In any case, participation would be limited to those actively planning a pregnancy (see below). Clinical populations are too distorted by biologic and health-care selection factors to be informative about fecundity in the general population.

Retrospective assessment of fecundity is possible using time to pregnancy questionnaires and can be based on populations recruited in cross-sectional surveys or through a pregnancy (2). The outcome variable is ascertained by asking, "How long did you (or, your wife/partner) take to conceive this child?", measured in months (ungrouped), preceded by filter questions on contraceptive use to establish eligibility. Although a pregnancy-based sample is convenient, it excludes sterile couples and underrepresents less fecund couples. The cross-sectional approach is applicable to the general population or for those who have specific exposures of interest. If complete ascertainment of the population can be achieved, for example, former workers in an occupational study, this is equivalent to recruiting a cohort and then collecting data on outcomes at the time of the cross-sectional survey, in other words, a retrospective cohort design.

The design is economical, in that it does not require a great deal of information about people's lives. This is because of its focus on a specific period of a couple's life, when they were having unprotected intercourse; questions on lifestyle are thus restricted in time. Furthermore, because phases of contraceptive use, life after sterilization, and so on are "silent," there is no need to spend time asking about them. Potential confounding variables include recent use of hormonal contraception, socioeconomic status, and age and smoking status of both partners prior to conception.

Because the comparison is of populations, it is not essential to include information on every factor that could affect fertility. For example, in an individual case, conception could be delayed because illness or travel led to an interruption in the sexual life of a couple. At the population level,

this would be important only if it occurred frequently and if an exposure group were especially predisposed to such events. The assumption can often be made that such factors are equally balanced between groups and/or that their effect is small at the population level. The practical importance is that one does not have to acquire information at this level of detail. This is fortunate, as recall of such details is likely to be inaccurate, and intimate questioning of this nature would not be appropriate for the typical research situation, for example, in the occupational context or in a general population survey.

Questionnaire validity

It has been found that, at the group level, the validity of recall of time to pregnancy is remarkably good. An American study found that a short, self-completion questionnaire was unbiased as compared with a detailed telephone interview (3). Accuracy was unrelated to duration of recall (which was up to 4 years). A study from the Netherlands, based on the population from a prospective study with pregnancies up to 20 months previously, found stability of response in retrospective time to pregnancy questioning after 3–5 weeks and no systematic errors as compared with the prospective data (4). A British study compared a retrospective time to pregnancy questionnaire with data that had been collected annually over the previous 20 years from women recruited in family planning clinics (5). A considerable degree of misclassification was evident at the individual level, but at the group level the distributions of the concurrent and the retrospective data were virtually identical (apart from some digit preference), even with duration of recall up to 20 years (6).

There is some evidence that satisfactory data on time to pregnancy are also obtainable from men. The time to pregnancy distribution constructed from replies from English male factory workers closely resembled that expected from prospective studies, even with up to 20 years of recall (7); the men were notified in advance of the topics to be covered in the interview. In addition, in studies of time to pregnancy-related factors (8) and time trends (9), analyses based on separate samples of male and female respondents drawn from the same population have given similar results.

It is probably wise to confine data collection to pregnancies that resulted in a birth. With miscarriages and so on, it is harder to be confident about the quality of the data obtained: They are typically underreported (10, 11), so that the sample of reported miscarriages may not be representative of all those that occurred, especially for those in the more distant past. It is also more difficult to remember the date of a past miscarriage, whereas a child's birthday is readily recalled; similarly, covariates and time to pregnancy itself may be harder to remember (2).

Acceptability

The required questionnaire section is short, even with the potential confounding variables. Most important, experience shows that it is highly acceptable in a wide variety of populations and settings. This may partly be because phases of

celibacy, casual encounters, and so on are excluded by the structure of questioning and do not have to be declared explicitly.

Reluctance tends to be encountered only at the stage of asking permission to carry out a survey, for example, of factory owners, managers, and trade unionists in the occupational context, not in its actual conduct. Interviewers commonly report that they found the questionnaire to be more readily accepted than they expected beforehand. This may be because people enjoy talking about their children, and in a wide variety of cultures they find it easy to grasp the notion of how long it takes to conceive a baby and are not embarrassed to discuss it.

Response bias

It is now a common experience in survey work that response rates tend to be disappointing. Exceptions to this may occur, for example, with pregnant or recently delivered women and in the context of a regular medical examination, for example, in the Italian occupational health system (12). The problem tends to be exacerbated if the study design combines time to pregnancy assessment with a request for the man to donate a semen sample (13). The size of any bias resulting from a poor response rate depends also on the likelihood that nonrespondents differ systematically from respondents. In a survey with the express purpose of studying fecundity, this problem may be intractable. However, if questionnaire items on fecundity can be embedded in a survey that already exists for more general purposes, response is unlikely to be strongly related to the degree of fecundity.

Experience in the United Kingdom with the Omnibus Survey (9), which is run monthly by the Office for National Statistics, and with the 1958 birth cohort (National Child Development Survey) (8) shows that refusal to answer fecundity-related questions is rare. The other form of item nonresponse, inability to answer the question, is somewhat more common with men than women (e.g., 14 percent and 7 percent, respectively, in the National Child Development Survey). The overall response rates of these surveys are approximately 70 percent and, although this is far from ideal, nonresponse is unlikely to be biased in relation to fecundity.

Influence of fertility treatment

The statistical analysis of time to pregnancy data involves survival analysis, in which the outcome is measured as the number of months taken to conceive, rather than as a dichotomy (yes/no or present/absent) as in much of epidemiology. This allows right censoring to be used, in which the months are removed from both the numerator and the denominator after the date of starting medical treatment, when this information is available, so the censoring date indicates that conception occurred at some later time. There is some loss of information involved, but this is outweighed by the advantage of having an estimate that is not biased by the possible effect of treatment.

More generally, if specific information is not available on when each couple sought treatment, time to pregnancy anal-

yses generally use right censoring, for example, at 14, 10, or 7 months. In general, assistance with conception is not sought in the early months of trying to conceive, and when treatment is effective it takes some time before conception occurs.

In either case, the loss of information is unimportant as most conceptions occur in the early months, and therefore little statistical power is lost. An additional advantage of the use of censoring is that recall of the duration of relatively long periods of infertility is less accurate (5).

However, even with the use of censoring, care must be taken when comparing populations with different levels of infertility treatment, for example, time trends or international differences. This is because couples who have had successful treatment, without which they would not have conceived, are included in the population as having a (censored) time to pregnancy value. As such couples are relatively infertile, this can paradoxically lead to apparently lower fertility in a population with more successful treatment.

Truncation bias

If a representative cross-sectional sample of the population were interviewed in September 2003, a couple who had commenced unprotected intercourse 12 months earlier would have a pregnancy (and therefore a time to pregnancy value) only if they conceived within that time; the less fecund couples would thus be excluded from the population at risk. It is therefore crucial that any analysis of time to pregnancy relating to time trends uses categories based on the "starting time," when unprotected intercourse started, rather than the date of conception or birth. If this is not done, truncation effects can occur (14, 15). This bias has the effect of artificially overestimating fecundity in the most recent category. A similar effect can occur at the beginning of the study period. The relevance of truncation bias is not only to time trend analyses but also to the study of exposures that have altered over time, for example, in occupational studies.

Aspects of fecundity not covered by time to pregnancy

A time to pregnancy value is eligible for acceptance only if conception occurred in the absence of an effective method of contraception. Although it is simple to ask whether or not a pregnancy resulted from contraceptive failure and to exclude those that did, it is possible that, in a comparison of different populations (e.g., in a time trend analysis), there could be a systematic difference in the probability of this being reported, which could distort the time to pregnancy distribution. In practice, time to pregnancy analysis is routinely checked by seeing whether the exposure variable is related to the proportion of "accidental" conceptions (16). In addition, the standard time to pregnancy regression model is rerun after excluding rapid conceptions (0 or 1 month) (16). These checks allow bias of this type to be detected.

A more difficult issue is deliberate terminations of pregnancy. It is not feasible to obtain reliable data on terminated pregnancies in a survey, and in most cultural contexts it is probably best not to try. The presence of a consequent bias

can only be detected by using the available statistics on abortion rates in a comparable population, if these exist (9).

At the other end of the spectrum of fecundity, some couples experience one or more infertile phases that do not end in a pregnancy. As these are not ascertained by the time to pregnancy question, study of time to pregnancy on its own understates the population experience of subfertility. For monitoring purposes this may not matter, as it is likely that any agent that causes an increase in serious subfertility would also shift the time to pregnancy distribution among those couples who did conceive.

In principle, it is possible to obtain information on periods of unprotected intercourse that did not end in pregnancy, using a minimum period of, for example, 3 or 6 months as a threshold for reporting, and to include this variable on infertile phases along with the time to pregnancy variable (7, 17). Survival analysis methods can handle this situation without difficulty by censoring when conception did not occur. In practice, however, ascertainment of such intervals is less reliable than for time to pregnancy. It may be that, with relatively short recall periods, good data are achievable. Methodological improvements in this area are desirable.

Behavioral intermediaries between biologic capacity and biologic outcome

In principle, the eligibility for acceptance of a time to pregnancy value or the duration of an infertile phase is based on a biologic criterion, exposure to intercourse without effective contraception. Many studies have used planned pregnancies (and unsuccessful attempts at conceiving) as a proxy for this, but they then exclude those couples whose approach to reproduction is less tightly controlled. This could introduce bias if such couples tend to have different characteristics, for example, if they are more likely to be smokers. A different approach is to accept a time to pregnancy value as eligible if it was not truly "accidental," for example, a contraceptive failure, which thus includes a larger and more representative section of the population. This issue requires further methodological work.

Although it is likely that time to pregnancy is affected by the frequency of intercourse, this is not an appropriate topic for population surveillance, partly because the quality of reporting is likely to be inadequate. It can often be assumed that the frequency distribution is similar in the different comparison groups, but this is not necessarily so. In the case of exposure to a chemical agent that alters libido, it would be inappropriate to adjust for frequency of intercourse, even if it could be measured, as it is on the causal path between the agent and the outcome (time to pregnancy).

A more subtle aspect is that the probability of conception may depend on the motivation of the couple. It has been suggested that the degree of persistence is an important determinant (18). It may therefore be advisable to incorporate the degree of planning and persistence into the data collection and analysis, together with information on the type and rigor of contraceptive method in use in the time period shortly before the starting time.

Similarly, couples' knowledge of the timing of the woman's maximal probability of conception may influence

the probability of conceiving among planners. This has been suggested as a possible reason for the apparent increase in fecundity over recent decades in Great Britain (9).

CONCLUSION

The scientific and public concern about "declining sperm counts" (19, 20) and the difficulty in deciding whether the reported observations are due to methodological problems (21, 22) highlight the need for a method of monitoring the fecundity of both sexes. Low participation rates make semen quality unsuitable for population surveillance, and no equivalent biomarker is available for women. It is possible, however, to monitor time to pregnancy as a functional measure of fecundity. This provides a measure of minor impairment, which is probably more appropriate as it is a sensitive measure and therefore more useful as an early warning; although major disturbances in fecundity may be of more obvious concern, especially to individual couples, no suitable monitoring method exists. It is not known to what extent minor and major impairments of fecundity have common risk factors. However, even minor impairment could be important as an indicator of biologic damage, especially if it involved a genetic mechanism, as heritable defects could also affect the offspring and future generations (2).

A short questionnaire module is available that is readily acceptable to respondents and able to provide information on time to pregnancy that is valid at the group level. It is also necessary to include couples who are at either end of the fecundity spectrum. In the case of those with reduced fertility, methodological improvements are needed.

By embedding data collection in population surveys that are conducted for more general purposes, the problem of response bias can be overcome. Use of censoring in the statistical analysis makes it possible to allow for the effects of fertility treatment. In addition, the analysis needs to be designed to avoid truncation bias, as well as other potential problems that are beyond the scope of this paper (16, 23).

Experience shows that stable estimates of the time to pregnancy distribution can be achieved with 200–300 pregnancies (2); fewer are needed in the case of ordered data such as successive 5-year age groups (9). It is wise to focus on the first pregnancy (or first phase of unprotected intercourse not leading to conception). This avoids the need to adjust for parity, a procedure that may introduce bias (16, 23), and for past obstetric history, resulting in a shorter and simpler questionnaire; it would also be unnecessary to stratify on desired family size (24). An alternative strategy is to include all pregnancies and infertile phases, in which case statistical methods must be used that do not assume the independence of events.

The potential therefore exists to collect data on fecundity that could be used for descriptive epidemiologic purposes. These could include spatial and sociodemographic variation, as well as the monitoring of time trends. The questionnaire module could be incorporated into existing surveillance systems or other routine data sources, such as the multipurpose surveys that are carried out by government bodies in most developed countries and repeated periodically. In addition to assisting in dealing with response bias, this is also an

efficient use of resources. A decision would need to be made on the frequency of data collection.

Ideally, the target age would be set so as to encompass all women who have passed their first "starting time," that is, unprotected intercourse that could lead to conception, plus an interval allowing sufficient time for conception to take place. This would correspond to the censoring time, for example, 14 months. However, in practice, because this age cannot be predicted, sampling would cover women with a broad age range, some of whom would not yet have reached this age (9). The potentially biasing effects of this need to be explored.

The questionnaire module could also be incorporated into occupational health surveillance schemes, where these exist, especially for workforces who are exposed to agents that could affect reproductive potential in either sex. In the case of female workers, allowance would need to be made for the "infertile worker effect" (25). Data could thereby be obtained on people who are occupationally exposed to a variety of exposures.

REFERENCES

- Bonde JPE, Hjollund NHI, Kold Jensen T, et al. A follow-up study of environmental and biologic determinants of fertility among 430 Danish first-pregnancy planners: design and methods. *Reprod Toxicol* 1998;12:19-27.
- Joffe M. Time to pregnancy: a measure of reproductive function in either sex. *Asclepios Project. Occup Environ Med* 1997;54:289-94.
- Baird DD, Weinberg CR, Rowland AS. Reporting errors in time-to-pregnancy data collected with a short questionnaire. *Am J Epidemiol* 1991;133:1282-90.
- Zielhuis GA, Hulscher MEJL, Florack EIM. Validity and reliability of a questionnaire on fecundability. *Int J Epidemiol* 1992;21:1151-6.
- Joffe M, Villard L, Li Z, et al. Long-term recall of time-to-pregnancy. *Fertil Steril* 1993;60:99-104.
- Joffe M, Villard L, Li Z, et al. A time to pregnancy questionnaire designed for long term recall: validity in Oxford, England. *J Epidemiol Community Health* 1995;49:314-19.
- Joffe M. Feasibility of studying subfertility using retrospective self reports. *J Epidemiol Community Health* 1989;43:268-74.
- Joffe M, Li Z. Male and female factors in fertility. *Am J Epidemiol* 1994;140:921-9.
- Joffe M. Time trends in biological fertility in Britain. *Lancet* 2000;355:1961-5.
- Wilcox AJ, Horney LF. Accuracy of spontaneous abortion recall. *Am J Epidemiol* 1984;120:727-33.
- Axelsson G, Rylander R. Validation of questionnaire-reported miscarriage, malformation, and birth weight. *Int J Epidemiol* 1984;13:94-8.
- Apostoli P, Bellini A, Porru S, et al. The effect of lead on male fertility: a time to pregnancy (TTP) study. *Am J Ind Med* 2000;38:310-15.
- Jensen TK, Slama R, Ducot B, et al. Regional differences in waiting time to pregnancy among fertile couples from four European cities. *Hum Reprod* 2001;16:2697-704.
- Akre O, Cnattingius S, Bergström R, et al. Human fertility does not decline: evidence from Sweden. *Fertil Steril* 1999;71:1066-9.
- Jensen TK, Keiding N, Scheike T, et al. Declining human fertility? *Fertil Steril* 2000;73:421-3.
- Weinberg CR, Baird DD, Wilcox AJ. Sources of bias in studies of time to pregnancy. *Stat Med* 1994;13:671-81.
- Karmaus W, Juul S. Infertility and subfecundity in population-based samples from Denmark, Germany, Italy, Poland, and Spain. The European Infertility and Subfecundity Study Group. *Eur J Public Health* 1999;9:229-35.
- Basso O, Juul S, Olsen J. Time to pregnancy as a correlate of fecundity: differential persistence in trying to become pregnant as a source of bias. *Int J Epidemiol* 2000;29:856-61.
- Carlsen E, Giwercman A, Keiding N, et al. Evidence for decreasing quality of semen during past 50 years. *BMJ* 1992;305:609-13.
- Swan SH, Elkin EP, Fenster L. The question of declining sperm density revisited: an analysis of 101 studies published 1934-1996. *Environ Health Perspect* 2000;108:961-6.
- Olsen GW, Bodner KM, Ramlow JM. Have sperm counts been reduced 50 percent in 50 years? A statistical model revisited. *Fertil Steril* 1995;63:887-93.
- Fisch H, Goluboff ET. Geographic variations in sperm counts: a potential cause of bias in studies of semen quality. *Fertil Steril* 1996;65:1044-6.
- Weinberg CR, Baird DD, Rowland AS. Pitfalls inherent in retrospective time-to-event studies: the example of time to pregnancy. *Stat Med* 1993;12:867-79.
- Olsen J, Rachootin P. Invited commentary: monitoring fecundity over time—if we do it, then let's do it right. *Am J Epidemiol* 2003;157:94-7.
- Joffe M. Biases in research on reproduction and women's work. *Int J Epidemiol* 1985;14:118-23.