American Journal of Epidemiology

Vol. 157, No. 2
Printed in U.S.A.
DOI: 10.1093/aje/kwf178

Invited Commentary: Monitoring Fecundity over Time—If We Do It, Then Let's Do It Right

Jørn Olsen[1] and Pamela Rachootin[2]

[1] The Danish Epidemiology Science Centre, University of Aarhus, Aarhus, Denmark.
[2] Southern Division of General Practice, Brighton, South Australia, Australia.

Received for publication September 14, 2001; accepted for publication October 31, 2002.

A number of investigators have pointed to the possibility of a secular decline in human fecundity due to changes in sperm concentration. It is unlikely that any historical trends will be definitively quantified, but a good case can be made for more precise monitoring of this phenomenon in the future. Such monitoring would be justified on the grounds of the importance of early detection of environmental effects on the capacity of humans to reproduce. Establishing a surveillance system that will be sensitive enough to detect changes in fecundity over time is, however, a challenging enterprise because of methodological concerns. It may be impossible to obtain a quality of design that will pick up subtle changes in fecundity.

data collection; fertility

The possibility of a decline in human fecundity has been discussed for at least 20–30 years (1–9). This issue was first featured prominently in the public media when Carlsen et al. (10) compared historical sperm density measurements recorded 50–60 years ago with recent survey data. A plausible biologic hypothesis was offered to explain the findings. It was suggested that fetal exposure to high levels of estrogen during the period of development of Sertoli's cells might be linked to impaired semen production in adult life, as well as to increased risk of cryptorchism and testicular cancer (11). Possible increases in maternal and, hence, fetal estrogen levels could be mediated by exposure to chemicals with estrogenic effects. Hormonal medication and changes in diet and obesity over time could also have an influence.

The present evidence for secular changes in human fecundity is circumstantial. Because a systematic surveillance system for monitoring human fecundity has not been implemented, no data exist to examine changes in fecundity over time as a possible marker of either adverse environmental exposure or the interaction between environment and social or biologic factors.

The usefulness of monitoring the incidence of cancer, myocardial infarction, asthma, congenital abnormalities, diabetes, and many other diseases is well recognized. Changes in human fecundity over time are likewise important to document. Among international agencies, epidemiologic research on changes in human fecundity has been neglected as a priority funding issue. This may be due to the general concern about the dangers of overpopulation. However, we believe that the potential contribution of fecundity as a health indicator should not be ignored. The frequency, suffering, and cost of infertility to both the individual and society are also grounds for supporting this research direction. Subfecundity may in itself be a risk factor for breast cancer and other diseases.

METHODOLOGICAL ISSUES

Fecundity

The interpretation of measures of fecundity based upon conception rates in couples that practice unprotected intercourse is complicated by the widespread volitional regulation of reproduction to limit family size. Even in the Third World, we see an increasing availability of effective contraception, and only small, isolated population groups remain without use of modern methods of family planning. Long-term trends in social norms, particularly with regard to desired family size, will continue to influence contraceptive behavior and therefore our efforts to measure fecundity.

Fecundity refers to the probability of conceiving within a given menstrual cycle. It may simply be measured as the proportion of women that conceive within the first cycle of unprotected sexual intercourse. Usually the entire distribution of menstrual cycles or months of unprotected intercourse are used to model fecundity.

Correspondence to Dr. Jørn Olsen, Vennelyst Boulevard 6, 8000 Aarhus C, Denmark (e-mail: jo@soci.au.dk).

Episodes or periods of unprotected intercourse may be reported prospectively by means of diaries, but they will be difficult to recall retrospectively. For this reason, the time to pregnancy is often measured in planned pregnancies, and experience shows that this is quantifiable even for pregnancy attempts that took place years ago (12). Unplanned pregnancies, including pregnancies resulting from contraceptive failure, have no measurable time to pregnancy. Because contraceptive failures and unplanned pregnancies will be more frequent among the most fecund, exclusion of these couples introduces bias. Accepting the time of unprotected sexual intercourse for unplanned pregnancies may solve the problem of selection related to planning, but it will introduces new sources of bias; data may be of poor quality and they will be subject to selection bias directly related to fecundity. These couples are included only if they become pregnant.

Fertility treatments reduce the number of unaided cycles of attempts at conception that should be taken into consideration when modeling the time to pregnancy distribution. Infertility treatment has advanced during recent years and is expected to continue to provide new treatment modalities. The trend for earlier intervention for treatment of infertility is expected to continue.

A monitoring system will need to be based upon representative samples from well-defined, stable populations, where sampling could be repeated over time. It is difficult to imagine that any "convenience" sampling could provide comparable data over time or between populations.

Although previous studies have utilized "convenience" sampling of pregnant women, there are several problems with such a design. First, sterile couples are not represented at all, and highly infertile couples are underrepresented. Measurement of the time to pregnancy in such a sample could not provide an estimate of fecundity unless it were combined with data from women within the catchment area who had attempted but failed to become pregnant (13). Furthermore, any interpretation of such data requires the assumption that couples were equally determined to achieve conception. This is unlikely to be valid across different populations and over long time periods (14).

The program should monitor couples attempting their first conception to reduce bias related to persistency in trying to conceive, the impact of changes in desired family size, and differences in the quality of data, which may be lower in women with shorter interpregnancy intervals.

Selection bias due to nonrespondents is especially troublesome. Good response rates have been achieved among pregnant women recruited from antenatal care centers. A high response rate from women selected at random from the general population is much more difficult to achieve. Such rates rarely reach over 80 percent and are often much lower. There is reason to believe that respondents have different fecundity than do nonrespondents (15). These and other methodological issues have been previously discussed (16–21).

A standardized, validated questionnaire will need to be developed (22) through a central agency. The instrument should be concise and include questions that define the time to pregnancy, assisted fertility treatments, contraceptive use, sexual habits, reproductive history, lifestyle, and diseases affecting reproductive capacity. We do not believe it is possible to set up a monitoring system with a high detection level of smaller changes in fecundity unless data are available for comprehensive adjustments. "Quick and dirty" surveys based on short questionnaires are not recommended.

Sperm quality

Sperm quality is an indirect determinant of fecundity. Previously, reports have suggested a secular decline in sperm count or sperm density. It is reasonable to continue this line of investigation, collecting and analyzing semen samples from well-defined populations under standard protocols that include rigorous quality-control measures and high compliance rates. This process may be facilitated in the future by computerized laboratory technology. It is, however, usually much more difficult to obtain high participation rates for semen studies compared with time to pregnancy studies. Response rates are often less than 40 percent in the studies we do. A high compliance is needed to reduce the risk of selection bias, and this may be encouraged by providing a generous financial inducement, which, in our experience, works in some populations. Stable estimates of semen parameters may be obtained from any sample at a given stage of data collection, but they are worse than unstable estimates if the estimates are biased.

In the selection of men for studying semen quality, it is important to control for seasonal variation, time from last ejaculation, collection method, delay in analysis of the sample, and method of analysis.

In addition, semen data have to be analyzed together with self-reported information on factors that may influence semen quality. A standard questionnaire should be developed for this purpose and made available in several languages.

Proposed outline of model

A monitoring system should aim to use at least two measures of fecundity. First, it should estimate the time to pregnancy and the time of unprotected sexual contacts. Second, it should measure sperm quality. The protocol must be robust enough to be replicated unchanged over decades. High compliance must be achievable. Surveys must be based upon population-based data and, if sampling is applied, valid sampling principles should be used. Finally, every effort must be made to standardize data on the intention to conceive, for example, by stratifying on the desired family size. Study subjects should be restricted to couples desiring their first child for the reasons that 1) use of a partial reproductive history in pregnancy planning may create confounding that is difficult to adjust (23), and 2) they may more often limit fecundity-reducing behavior like smoking if they expect their fecundity to be low.

Careful consideration must be given to the source of semen samples. They could be obtained from either an independent male sample or males participating in the fecundity study. In the future it may be possible to replace semen samples with blood tests for biomarkers such as inhibin B. It

is known that inhibin B controls follicle-stimulating hormone secretion via a negative feedback mechanism, but its relation to the function of Sertoli's cells is not known (24). Blood samples should be easier to obtain than semen samples.

Ideally an international organization such as the World Health Organization should host the project to ensure the necessary political, financial, and technical support.

A note of caution

Despite rigorous quality-control measures, any monitoring system will remain highly sensitive to self-selection bias related to nonrespondents. Furthermore, the reason for declining invitations to participate may change over time and may differ between regions. We accept Joffe's point (25) that embedding time to pregnancy monitoring into a general health survey may reduce selection bias, but the price to pay may be too high, namely, limited data on other determinants of time to pregnancy measures.

The establishment of fecundity monitoring entails a raft of technical problems. Obtaining a valid picture of a large population will be exceedingly challenging. It is a challenge worth meeting only if we can navigate around sources of bias and confounding. However, many sources of bias are out of our control. An acceptable level of compliance to a surveillance program may be impossible to achieve if ethics committees do not accept active recruitment or use of incentives, and it is worth considering whether the route is worth traveling. Neither the journey nor the destination may compensate for the effort if we end up with data of insufficient quality. Just because we are technically capable of organizing a fecundity surveillance system clearly does not mean that we should necessarily do so.

Monitoring, like screening, has its own ethics. One must start with specific ends in mind. If there is not sufficient expectation of practical benefit, the gathering of deeply private data at public expense cannot be justified. There are, however, important benefits that flow from a valid fecundity monitoring system. The most important will be detecting ongoing harmful exposures. Good data may generate new hypotheses; a surveillance system may aid population planning.

Census data on fertility have been deemed sufficient for administrative planning. Simple counts of the population do not, of course, measure fecundity per se, but such counts have the compensating advantage of reflecting the large contribution of in- and outmigration, as well as natural growth. Given the common emphasis on short-term population goals, it seems unlikely that there will be increasing demand for true fecundity measures to aid population planning, unless fecundity reaches a level that threatens reproduction.

It is not clear that fecundity surveillance will supply vital niche data for which there are no available alternatives, and again, the problem to keep up the collection of high quality data for decades may be a naïve ambition.

CONCLUSION

A decline in human fecundity would represent a public health problem that deserves a major investment and a much higher priority than it now attracts. Such a decline would reflect external causes that need to be addressed. If these causes are general environmental pollutants, the decline in fecundity may be slow and impossible to detect over a chosen time span. If the causes are lifestyle factors such as smoking, which operates during fetal life and which may follow fashion, changes in fecundity could be of a magnitude that is detectable in the proposed monitoring system.

A decline in fecundity that exceeds 20 percent (e.g., from 30 to 24 percent) over a time period of less than 10 years or a reduction in semen concentration of a similar magnitude (e.g., from 70 million/ml to 56 million/ml) should be detectable in a well-designed and well-conducted surveillance system. The limit of sensitivity of the proposed monitoring system is probably along these lines. A decline of lesser magnitude may not be detectable because of bias or confounding. Confounding or bias may, of course, also produce a spurious change in fecundity in any direction. It may be impossible to establish a large monitoring system with the requirements mentioned either because of financial constraints or because of inadequate response rates. In this case, no data are a better alternative to insufficient data.

Monitoring is best conducted in industrialized countries, where the expected exposures of concern (11) are most prevalent and where the infrastructure for obtaining data is more likely to exist. Such monitoring would benefit from a population that is already clearly enumerated and where women are able to report their reproductive history.

Whether the sperm concentration and human fecundity have declined during the past 50 years is a question we will probably never be able to answer. A monitoring system could ensure that we have a better understanding of developments over the next 50 years. The issue is not solely of environmental concern. Infertility is a frequent problem that carries severe burdens for the individuals involved and for society.

ACKNOWLEDGMENTS

The activities of the Danish Epidemiology Science Centre are financed by a grant from the Danish National Research Foundation.

The authors thank Ron Somers for valuable comments on an earlier version of the manuscript.

REFERENCES

1. Rachootin P, Olsen J. Prevalence and socio-economic correlates of subfecundity and spontaneous abortion in Denmark. Int J Epidemiol 1982;11:245–9.
2. Bostofte E, Serup J, Rebbe H. Has the fertility of Danish men declined through the years in terms of semen quality? A comparison of semen qualities between 1952 and 1972. Int J Fertil 1993;28:91–5.

3. Mosher WD, Pratt WF. Fecundity and infertility in the United States: incidence and trends. Fertil Steril 1991;56:192–3.
4. Mosher WD, Pratt WF. Fecundity and infertility in the United States, 1965–1988. Advance Data 1990;192:1–12.
5. Leto S, Frensilli FJ. Changing parameters of donor semen. Fertil Steril 1981;36:766–70.
6. Nelson CM, Bunge RG. Semen analysis: evidence for changing parameters of male fertility potential. Fertil Steril 1974;25:503–7.
7. MacLeod J, Wang Y. Male fertility potential in terms of semen quality: a review of the past, a study of the present. Fertil Steril 1979;31:103–16.
8. Rachootin P, Olsen J. Secular changes in the twinning rate in Denmark 1931 to 1977. Scand J Soc Med 1980;8:89–94.
9. Whorton MD, Krauss RM, Marshall S, et al. Infertility in male pesticide workers. Lancet 1977;2:1259–61.
10. Carlsen E, Giwercman A, Keiding N, et al. Evidence for decreasing quality of semen during the past 50 years. BMJ 1992;305:609–13.
11. Sharpe RM, Skakkebaek NE. Are oestrogens involved in falling sperm counts and disorder of the male reproductive tract? Lancet 1993;341:1392–5.
12. 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.
13. Olsen J, Andersen PK. We should monitor human fecundity, but how? A suggestion for a new method that may also be used to identify determinants of low fecundity. Epidemiology 1999; 10:419–21.
14. 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.
15. Larsen SB, Abell A, Bonde JP. Selection bias in occupational sperm studies. Am J Epidemiol 1998;147:681–5.
16. Baird DD, Wilcox AJ, Weinberg CR. Use of time to pregnancy to study environmental exposures. Am J Epidemiol 1986;124: 179–84.
17. Weinberg CR, Baird DD, Wilcox AJ. Sources of bias in studies of time to pregnancy. Stat Med 1994;13:671–81.
18. Juul S, Karmaus W, Olsen J. Regional differences in waiting time to pregnancy: pregnancy-based surveys from Denmark, France, Germany, Italy, and Sweden. The European Infertility and Subfecundity Study Group. Hum Reprod 1999;14:1250–4.
19. Keiding N. Analysis of time-to-pregnancy data. Scand J Work Environ Health 1999;25(suppl 1):10–11.
20. Olsen J. Is human fecundity declining—and does occupational exposures play a role in such a decline if it exists? Scand J Work Environ Health 1994;20:72–7.
21. Joffe M. Feasibility of studying subfertility using retrospective self reports. J Epidemiol Community Health 1989;43:268–74.
22. Olsen J. Epidemiology deserves better questionnaires. IEA European Questionnaire Group. Int J Epidemiol 1998;27:935.
23. Olsen J. Options in making use of pregnancy history in planning and analyzing studies of reproductive failure. J Epidemiol Community Health 1994;48:171–4.
24. Anderson RA, Sharpe RM. Regulation of inhibin production in the human male and its clinical applications. Int J Androl 2000; 23:136–44.
25. Joffe M. Invited commentary: the potential for monitoring of fecundity and the remaining challenges. Am J Epidemiol 2003; 157:89–93.