Is human fecundity declining in Western countries?

Egbert te Velde\textsuperscript{1,\ast}, Alex Burdorf\textsuperscript{2}, Eberhard Nieschlag\textsuperscript{3}, René Eijkemans\textsuperscript{4}, Jan A.M. Kremer\textsuperscript{5}, Nel Roeleveld\textsuperscript{6}, and Dik Habbema\textsuperscript{7}

\textsuperscript{1}Department of Public Health, Erasmus University MC, Rotterdam, The Netherlands \textsuperscript{2}Department of Public Health, Erasmus MC, Rotterdam, The Netherlands \textsuperscript{3}Centre for Reproductive Medicine and Andrology, University of Münster, Münster, Germany \textsuperscript{4}Mathematician and Statistician, Department of Public Health, Erasmus MC, Rotterdam, The Netherlands \textsuperscript{5}Department of Obstetrics and Gynaecology Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands \textsuperscript{6}Department of Epidemiology, Biostatistics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands \textsuperscript{7}Department of Public Health, Erasmus MC, Rotterdam, The Netherlands

\textsuperscript{\ast}Correspondence address. E-mail: e.r.tevelde@wanadoo.nl

\textbf{ABSTRACT:} Since Carlsen and co-workers reported in 1992 that sperm counts have decreased during the second half of the last century in Western societies, there has been widespread anxiety about the adverse effects of environmental pollutants on human fecundity. The Carlsen report was followed by several re-analyses of their data set and by many studies on time trends in sperm quality and on secular trends in fecundity. However, the results of these studies were diverse, complex, difficult to interpret and, therefore, less straightforward than the Carlsen report suggested. The claims that population fecundity is declining and that environmental pollutants are involved, can neither be confirmed nor rejected, in our opinion. However, it is of great importance to find out because the possible influence of widespread environmental pollution, which would adversely affect human reproduction, should be a matter of great concern triggering large-scale studies into its causes and possibilities for prevention. The fundamental reason we still do not know whether population fecundity is declining is the lack of an appropriate surveillance system. Is such a system possible? In our opinion, determining total sperm counts (as a measure of male reproductive health) in combination with time to pregnancy (as a measure of couple fecundity) in carefully selected populations is a feasible option for such a monitoring system. If we want to find out whether or not population fecundity will be declining within the following 20–30 years, we must start monitoring now.

\textbf{Key words:} environmental effects / sperm quality / epidemiology / male infertility

\textbf{Background}

Fecundity, the capacity of couples to conceive and have children, depends on numerous biological processes including spermatogenesis, oogenesis, transport of gametes, fertilisation of the oocyte, implantation of the embryo, and the development of the foetus thereafter (Habbema et al., 2004). Although these processes have an evolutionary-driven genetic base, intrauterine and postnatal influences related to the environment and our changing society, may adversely modulate the male and female reproductive potential. Since Carlsen et al. (1992) reported that mean sperm counts decreased by 50\% during the second half of the last century and suggested that this decline in sperm quality and the increasing prevalence of genitourinary abnormalities may have a common environmental aetiology, there has been widespread anxiety about the effects of environmental pollutants on human reproduction. Given the well-known adverse effects of occupational exposures, e.g. pesticides, on reproduction (Burdorf et al., 2006), extrapolation of the effects of environmental chemicals to the general population as suggested by Carlsen et al., did not come as a surprise. These concerns were fuelled by studies in wildlife in which feminization and sterility of male animals occurred in highly polluted areas (Edwards et al., 2006). The last author of the Carlsen paper elaborated on the explanation for their findings and launched the testicular dysgenesis syndrome (TDS) (Sharpe and Skakkebaek, 1993). This hypothesis postulates that in utero exposure of male fetuses to high levels of so-called ‘endocrine disrupting chemicals’, with estrogenic or anti-androgenic properties, may cause disturbances in the development of all the major cell types within the testis: the germ cells themselves, the sustaining Sertoli cells and the testosterone-secreting Leydig cells. In addition to impaired sperm quality, this would lead to increased occurrence of testicular cancer and malformations of the male genital tract, such as cryptorchidism and hypospadias. The TDS hypothesis is corroborated by the observation of increasing occurrences of testicular cancer (Purdue et al., 2005) and possibly also of malformations of the genital tract (Aitken et al., 2004) suggesting that the whole male reproductive system is under attack by environmental chemicals. The steep decline in birth rates in Western societies since the sixties (Lutz et al., 2003) further
contributed to the general anxiety. In addition to the well-known societal changes with couples wanting fewer or no children nowadays, it was argued that this downward trend might also be related to a general loss of fecundity (Skakkebaek et al., 2001). This would explain the growing demand for in vitro fertilization and related assisted reproductive technologies (ART) (Jensen et al., 2008).

These concerns are contradicted by the results of various population studies in Europe and the USA on secular trends in fecundity, indicating that population fecundity has either improved (Joffe, 2000; Jensen et al., 2005; Stephen and Chandra, 2006; Scheike et al., 2008) or remained unchanged (Oakley et al., 2008) over the past 30–40 years. One of these authors concludes that ‘if one or more chemicals are relevant, either there has been no adverse impact in male fertility during this period, or any such effect has been more than compensated by a countervailing increase in couple fertility’ (Joffe, 2000). For example, behavioural changes may have resulted in more efficient use of the coital act during the fertile period of the menstrual cycle or may have encouraged couples to ‘try harder’ (Scheike et al., 2008), thus obscuring adverse trends in fecundity. Alternatively, because of the stricter regulations on reproduction-toxic chemicals, such as pesticides and organic solvents, in most European countries, our environment may have improved during the last decades (Mülnner et al., 2007). Moreover, changes in lifestyle, such as a decrease in smoking, increased fertility awareness and a decline in sexually transmitted diseases, may have changed population fecundity in a favourable direction (Graham, 1996; Stephen and Chandra, 2006; Scheike et al., 2008). However, the growing epidemic of obesity (Homan et al., 2007) and the postponement of parenthood to ages when not only women but also men become less fertile (te Velde and Pearson, 2002; Sartorius and Nieschlag, 2009) must have a negative impact on population fecundity.

Apparently, the evidence for an adverse trend in fecundity in the second half of the twentieth century is less straightforward than Carlsen et al. have suggested. Balancing all the evidence, we conclude that we do not know whether or not population fecundity is declining in Western countries. However, in our opinion, for the following reasons it is imperative to find out. First, the impact of a decline in fecundity is critical to the well being of individual couples who wish to have children. Secondly, the potential influence of widespread environmental pollution and of lifestyle factors adversely affecting reproduction, should be a matter of great concern triggering large-scale studies into its causes and possibilities for prevention. Finally, a decline in fecundity may result in adverse demographic trends in the future such as an increasingly unfavourable support ratio which is the ratio between the working and non-working fractions of the population (Lutz et al., 2003). Governments should be able to anticipate such trends and reverse them if possible. On the other hand, if population fecundity is not declining, the near-panic sometimes expressed in both the lay and medical press, and the claims that more ART is required to compensate for fecundity loss, are unfounded.

Is it possible to measure and monitor population fecundity?

For monitoring population fecundity, it is necessary to perform a baseline measurement and periodically thereafter, e.g. every 5 years. When no effective contraception was available, the mean number of children married women had during their reproductive life, was a fairly good reflection of couple fecundity. But in the present era of birth control and family planning most couples in Europe want only one or two children or no children at all. Hence, the level of fecundity of a population has to be determined by biological tests or through indirect approaches. Because they have been extensively studied in epidemiological research, the following two, in our opinion, are the best candidates: (i) measures of sperm quality including concentration, motility and morphology as a reflection of male reproductive health (Bonde et al., 2001), and (ii) measures of couple fecundity: time to pregnancy (TTP) and its derivatives 1-year fertility or permanent infertility rates (Joffe, 1997). Apart from those, the male-to-female sex ratio (James, 1996) and the incidence of spontaneous dizygotic (DZ) twinning (Ferrari et al., 2007) have also been proposed as potential markers of fecundity. Paternal occupational exposure to dibromochloropropane, for example, was found to reduce both male fecundity and the male-to-female sex ratio at birth (Potashnik and Porath, 1995) whereas in fruit growers exposed to high levels of pesticides a similar pattern was found (Smits et al., 2005). However, at the population level such an association could not be demonstrated (Joffe et al., 2007). Spontaneous DZ twinning rates, a potential measure of couple fecundity, have declined since the 1950s in most developed countries and have been related to adverse effects of environmental agents such as stilboestrol used in the cattle industry and pesticides (Rachootin and Olsen, 1980). However, due to advancing maternal age and the use of fertility drugs for ART, DZ twinning rates have actually steeply increased since the 1980s. Whether the age-adjusted and ART-adjusted spontaneous DZ twinning rate is a marker of fecundity in the general population, is unknown at present.

Environmental pollutants and lifestyle determinants may also affect female fecundity (Bretveld et al., 2006). However, in contrast to sperm cells, oocytes are hardly accessible and measures of female fecundity such as abnormalities of the menstrual cycle are ill-defined. Therefore, appropriate female determinants for monitoring fecundity, are lacking.

Since the end-point of reproduction, the occurrence of a pregnancy resulting in the birth of a child, depends on both male and female factors, the unit of analysis is always the couple. Consequently, markers of sperm quality are to be considered as surrogate measures of couple fecundity.

In the following we will focus on sperm variables and TTP as measures of couple fecundity. In our opinion, they are the most appropriate methods available at present to monitor population fecundity.

Sperm quality measures

Although the three generally accepted sperm parameters, concentration, motility and morphology, initially seemed obvious and easily obtainable biomarkers of male fecundity, the aforementioned study by Carlsen et al., has triggered lively and still ongoing debate between supporters and opponents. This is related to the fact that the 61 studies included in their meta-analysis containing ~15 thousand subjects, are extremely heterogeneous in terms of period of abstinence, age, reasons for selection, socio-economic background, methods of semen analysis, lifestyle variables, demographic...
characteristics, seasonality, country/region of origin and methods of statistical analysis used (Lerchl and Nieschlag, 1996). Several re-analyses of the Carlsen report have been performed. A complete and independent re-analysis retrieving additional data (Swan et al., 1997) and an updated analysis including semen studies up to 1996 (Swan et al., 2000) confirmed the trends reported by Carlsen et al. In contrast, Olsen et al. (1995) argued that a 50-year trend analysis is not justified in view of the paucity of data in the first 30 years. Including only the last 20 years, containing 90% of the subjects and 80% of the studies, they found stable or even increasing sperm concentrations over time. In spite of the well-known right-skewed distribution of sperm variables, the Carlsen analysis is based on mean sperm counts. When using median values instead, available in the majority of studies, there was no significant change over time (Handelsman, 2001). Another major problem in the interpretation of the results is related to the large geographical differences in sperm counts. Before 1970, for example, 11 of the 13 studies in the Carlsen report came from the USA, seven of which were from New York (64%) where sperm counts are known to be consistently higher than elsewhere (Fisch and Goluboff, 1996). After 1970, only 17 of the 48 studies came from the USA, with three being from New York (6%). Re-analyses accounting for these geographical differences did not show a significant trend in the non-USA, mainly European studies. In contrast, in the USA studies a significant negative linear trend was present, comparable to the one in the Carlsen analysis (Becker and Berhane, 1997). For the New York studies and for the non-New York studies separately, however, no significant change over time could be identified (Saidi et al., 1999).

The Carlsen paper triggered 27 major studies on time trends of sperm parameters, each reporting on changes over time in the same area or country. There were six studies which demonstrated an unambiguous decline in one or more sperm parameters including sperm concentration, 16 studies did not find a decline or even saw an increase in sperm quality measures and five studies reported ambiguous results, e.g. improved sperm motility whereas sperm concentrations decreased, as reviewed by Fisch (2008). Again, large differences in sperm parameters between but even within countries appeared to be present in these, mostly European, studies. These contradictory findings raised the following questions: (i) Are these differences caused by differences in methodology and quality of sperm analysis tests or in subject recruitment? (ii) Do these differences reflect regional differences in fecundity? and (iii) Do the routinely performed sperm parameters provide a reliable reflection of the fraction of high quality sperm responsible for conception? These questions prompted an unprecedented, prospective and co-ordinated action at a European level, using standardized methods of subject recruitment and semen analysis, and taking account of many variables including lifestyle characteristics, educational levels and age of both partners. The results were published in three major reports (Jensen et al., 2001; Jorgensen et al., 2001; Slama et al., 2002). About an equal number of couples from four major cities in Finland, Denmark, Scotland and France, almost 1000 couples altogether, contributed semen samples. The women were three or more months pregnant after a spontaneous conception and the time it took them to become pregnant (TTP as a measure of fecundity) was recorded for each couple. The results demonstrated that with increasing sperm concentrations up to 55 million per millilitre, the average TTP became shorter corresponding with better fecundity. Sperm morphology independently contributed to the prediction of pregnancy. These results are in line with those of a major prospective study from Denmark in first pregnancy planners demonstrating that the probability of pregnancy per menstrual cycle steeply increased from no or a few sperm cells per ejaculate to a concentration of about 50 million/ml (Bonde et al., 1998). The results of the four-country-study also showed the paramount importance of regional differences. For example, the average sperm count of Finnish men appeared to be 34% higher than in Denmark, with Scotland and France in between. Nevertheless, Finish couples appeared to be the least fecund in comparison with their counterparts in the other three countries. These results demonstrate that differences between regions/countries, and consequently also changes over time, have to be interpreted with great caution. They may be caused by differences in in utero exposures of the male partners to endocrine disruptors some 20–30 years ago, the mechanism proposed by Carlsen et al., or by present exposures to environmental pollutants or by differences/changes in lifestyle related to smoking, stress, obesity, sexual behaviour and delay of childbearing or by genetic variations among populations. To include core information of these determinants seems imperative in studies on sperm quality in order to explain spatial and regional differences.

Most recently it became evident that total sperm count, the product of ejaculate volume and sperm concentration, better reflects male reproductive potential than sperm concentration alone provided that abstinence times are carefully taken into account (Amann, 2009; Cooper et al., 2009). As the determination of total sperm counts is a matter of relatively simple counting whereas sperm morphology and motility are more sensitive to subjective judgement, we think the former is more suitable for monitoring purposes. Whether strict adherence to the WHO guidelines on sperm analysis sufficiently guarantees the obligatory reproducibility of total sperm counts between centres and over time or whether automated sperm analysers are more suitable for monitoring population fecundity, still has to be decided.

**Time to pregnancy**

The degree of fecundity of a couple is determined by the chance per menstrual cycle of a conception leading to live birth, given unprotected intercourse i.e. fecundability in demography. The distribution of individual couple chances is extremely heterogeneous, varying from zero = sterile to an estimated upper limit of 60% per menstrual cycle = ‘super fertile’ (Bongaarts, 1975; Leridon and Spira, 1984). Since the most fecund couples will conceive first, progressively less fecund couples remain in the pool of couples who have not (yet) achieved a pregnancy as time goes by. Hence, the time it takes to become pregnant since actively trying to conceive, time to pregnancy (TTP), is a measure of couple fecundity. TTP as outcome has been extensively used in epidemiological studies to detect the effects of occupational exposures, but also to determine differences between regions and trends over time (Baird et al., 1986; Joffe, 1997; Joffe et al., 2005; Bonde et al., 2006). A prospective study design has definite advantages but is extremely expensive and time-consuming (Baird et al., 1986; Tingen et al., 2004; Bonde et al., 2006; Key et al., 2009). Interpretation of the results of studies with a retrospective design are hampered by potential biases related to recruitment, fertility
treatment, accidental pregnancies, the degree of planning and the persistency in trying, social background, sexual behaviour, female age and non-response (Baird et al., 1986; Weinberg et al., 1994b; Juul et al., 1999; Jensen et al., 2000; Basso et al., 2000; Tingen et al., 2004; Bonde et al., 2006). Most studies have been limited to fertile couples who sooner or later achieve a pregnancy, thus excluding sterile cases. Opinions differ about the reliability of retrospectively collected TTP data, especially when the pregnancies occurred long ago (Baird et al., 1991; Zielhuis et al., 1992; Cooney et al., 2009).

Despite its susceptibility to bias, the retrospective design may be well-advised for ongoing population surveillance aimed to monitor trends in population fecundity (Joffe, 2003b; Olsen and Rachootin, 2002). It seems feasible to analyse the TTP distribution of a well-defined, cross-sectional sample of the population and repeat these measurements at regular time intervals. However, the questionnaire to be developed should not be restricted to couples who sooner or later conceive but also should collect information from couples who attempted but failed to have a live birth (Olsen and Andersen, 1999). In order to reduce various sources of bias, only information from first pregnancies are to be collected (Olsen, 1994; Weinberg et al., 1994a). Given the fact that miscarriages are heavily underreported, it is advisable only to include pregnancies leading to live births (Axelson and Rylander, 1984). Because teenage pregnancies should also be included into the data base, information from girls of about 13 or 14 years of age onwards should be obtained as well.

The upper age should account for the large variation in the end of the fertile period of women and for the long durations of time after which natural conceptions may still occur (te Velde and Pearson, 2002). The availability of a standardized, concise and robust questionnaire which is validated in different populations and remains applicable over long periods of time, is essential (Olsen, 1998; Bonde et al., 2006). The key questions ‘how long did it take you to conceive?’ and ‘how long did you attempt but failed to conceive’ have also to include information on periods when couples temporarily stop trying to become pregnant and questions on contraceptive use and fertility treatment.

For the same reasons as mentioned for sperm quality parameters, questions on lifestyle, ages of both partners and, if possible, on environmental and occupational exposures, should be added. Embedding the questionnaire into a multipurpose periodical health survey, will reduce selection bias due to non-response (Joffe, 2003a). Monitoring population fecundity by using a TTP questionnaire seems feasible but requires large numbers (see Discussion). Therefore, building a collection of high quality data for several decades, an essential condition for monitoring of fecundity, is a large-scale and costly undertaking for which the protracted motivation by governments and authorities is required.

**Discussion**

The significance of monitoring the incidence of diseases, such as cancer, is well recognized in epidemiological research. Although equally important in our opinion, why is a system for monitoring human fecundity lacking? Undoubtedly, because there is no unambiguous ‘gold standard’ for fecundity or infecundity. Hence, monitoring fecundity is complex compared with the registration of disease cases over time which is appropriate for demonstrating trends in cancer, congenital abnormalities or other diseases with a clear-cut diagnosis. This is the main reason, we think, why we still do not know whether or not population fecundity is declining. The results of studies so far are ambiguous and difficult to interpret. The claim that environmental pollutants are involved in a decline of population fecundity can neither be confirmed nor rejected, even though their effects in specific occupational groups have been well-established.

Temporal trends in fecundity may also be determined by lifestyle attitudes related to behavioural and societal changes. The preponderance of data on sperm parameters and on time trends in fecundity seem to indicate that there has been no decline in sperm quality and couple fecundity so far. However, the results of some well conducted studies on sperm quality suggest that in some parts of the world environmental pollutants or life style determinants may already have a negative impact on population fecundity, as reviewed by Fisch (2008). Such trends may spread to other areas and may challenge the fecundity of future generations. Hence, in our opinion it is of great importance to monitor population fecundity.

But what makes a test or method appropriate for monitoring trends over time? Other conditions are required than in reproductive medicine, where the main focus is on individual patient care and most tests are meant to contribute to the diagnosis and prognosis of individual couples. These tests are continuously changed and adapted in order to improve their diagnostic and predictive accuracy. In contrast, tests appropriate for monitoring have to be stable, standardized, robust and reproducible in order to detect changes over long periods of time.

The aim of monitoring is to detect increasing or declining temporal trends in fecundity. Therefore, the study populations to be selected for monitoring, not necessarily have to be representative for the fecundity of the general population, provided they are always selected in the same manner. Moreover, because of the considerable and largely unexplained spatial differences, monitoring should be restricted to well-defined regions.

Large numbers of women are required for such TTP surveys. For example, to detect a decline in the mean pregnancy chance per menstrual cycle of 7% from time A to B, the recruitment of ~6000 women was estimated to be required at each time point (Slama et al., 2004). This would correspond with a seemingly small fecundability decline from 20 to 18.6%. To detect smaller declines more couples and to detect larger declines fewer couples are required. Essential information on an additional number of determinants, such as on environmental exposures in the past and present and on lifestyle variables probably requires an increase in the sample size.

Ideally, the partners of the women taking part in the TTP surveys should participate in the studies on sperm quality. However, donating sperm is often felt as embarrassing and hence the participation rate of volunteers contributing semen samples for research is known to be low (Muller et al., 2004).

What is the relation between total sperm counts and TTP? In a simulation study it was estimated that considerable declines in sperm concentration will result in relatively small decreases in fecundity (Slama et al., 2004). If true, sperm concentration and also total sperm count might prove to be a more sensitive marker than TTP for monitoring fecundity. Although complex and full of pitfalls, total sperm count and TTP, in our opinion, are potentially feasible means to monitor population fecundity.
Finally, a statement from a commentary on monitoring fecundity: “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” (Olsen and Rachootin, 2003). Indeed, let us stop arguing about human fecundity in the past, but instead let us look ahead and start monitoring now. There is enough reason for concern: the fecundity of future generations is at stake.

References


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;10:419–421.


