What has happened to human fertility?

Michael Joffe

Department of Epidemiology and Public Health, Imperial College, London W2 1PF, UK

1Correspondence address. E-mail: m.joffe@imperial.ac.uk

Semen quality appears to have declined in recent decades in some populations, e.g. north-western Europe. At the same time, couple fertility may have increased. Hypotheses are suggested for this apparent inconsistency. Alongside the deterioration of spermatogenesis there is clear evidence of an increase in other related problems, notably testicular cancer. The sharply rising trend in this condition started a century ago—decades earlier than sometimes thought. This and other evidence clearly indicates an environmental origin, but there is also a definite genetic component. The relationship of genetics and environment is discussed in the context of the puzzle that infertility is inherited, which appears to be impossible from an evolutionary standpoint. Poor semen quality is related not only to testicular cancer but also to zygote development, in which cancer-like disruption of the genetic apparatus is observed, with serious implications for offspring health. This needs to be seen in the context that human reproduction is prone to a higher degree of impairment than that of other mammalian species, in relation to spermatogenesis, couple fertility, early pregnancy loss and embryonic aneuploidy; female- and male-mediated pathways are both implicated. It is unclear whether such human specificity originated on an evolutionary/genetic or a historico-social timescale, which is important in relation to pathogenesis. The evidence clearly indicates that the currently most popular explanation for male reproductive system impairment, the endocrine disruption hypothesis, cannot explain the main features of the descriptive epidemiology. An alternative pathogenesis is outlined, and some possible exposures considered that could be responsible.

Key words: infertility / spermatogenesis / testicular cancer / embryo development / early pregnancy loss

Introduction

The publication of a article by Carlsen et al. (1992) in the BMJ in the early 1990s led to public concern over the possibility that human fertility has declined. Whether or not the findings in that particular report were true, the issue of human fertility is clearly important. (In this article, the term ‘fertility’ is taken to indicate biological fertility, the capacity to conceive given unprotected intercourse, in contrast to demographic fertility, the actual number of children. Sometimes the alternative term ‘fecundity’ is used for this purpose, but others use these terms the other way around.)

Fertility is difficult to study in humans. It can be looked at from a functional perspective, by the use of biomarkers, or from a mechanistic viewpoint. They are complementary, and ideally each should inform the others. Functional fertility refers to how easy or difficult a couple find it to conceive, given that they are having unprotected intercourse, and tends to be assessed by looking at how long this takes, since more fertile couples tend to conceive more quickly (see below).

The other two approaches are at the individual level, not the couple, and relate to the biological processes operating in each sex separately. Biomarkers can be used for descriptive purposes, e.g. to study trends, or to make other comparisons. In the case of trends, it is necessary to have repeated measures on reasonably large samples that are representative of the underlying population, which rules out invasive procedures. This means that in practice such research is restricted to the use of easily collectible biomarkers, which are not available for women, so that studies of trends in female reproductive function are confined to such things as age at menarche and menopause. As techniques improve, better information on female fertility at the population level may become available, for example blood FSH and anti-mullerian hormone concentration.

In men, the situation is more fortunate, because semen is relatively easily obtained. The biomarkers of semen quality that are available over a period of time long enough to be useful in assessing trends are, in practice, limited to density (concentration), volume, number, motility and morphology (WHO, in press). The focus in much of the debate has been on density, expressed as the mean, which is unfortunate because the highly skewed distribution in humans makes it far less suitable than the median for making such a comparison. A particularly useful way of analysing semen, Sperm Chromatin Structure Assay (SCSA) (Evenson and Wixon, 2005), has unfortunately not been widely available on a population basis for a long enough period.

The mechanistic approach is necessary so as to make biological sense of whatever trends or other patterns are identified: what underly- ing chemical and physical processes bring about the observations. This could for example involve cellular, molecular and regulatory mechanisms. It is important to recognize that a particular clinical condition, such as impaired spermatogenesis, could result from one of
several different types of pathogenesis; and any particular pathogenesis can typically be initiated by different exogenous and/or endogenous agents.

A problem in this literature, as so often, is that scientists’ specialization means that particular issues are looked at in isolation (sometimes called ‘silos’), rather than as an integrated whole. And yet the complete picture can only be understood by looking at the entire range of evidence—epidemiology, genetic epidemiology, clinical research e.g. with infertile patients, experimental research on animals with a more mechanistic focus, as well as toxicology.

This article is mainly concerned with examining trends in human fertility, and the related topic of species differences between humans and other mammals; it does not cover the related issue of infertility and toxic chemicals (Joffe, 2003), or the relevant genetic evidence (Joffe, 2007), which have been published elsewhere. Most of the available information relates to impairment of male fertility and associated conditions, so that women are comparatively neglected. The article also suggests a mechanistic basis for deteriorating male reproductive health. In this review, the focus will be on the central processes of reproduction, involving the gonads and germ cells, omitting impairments or diseases in accessory sex organs including the breast.

### Trends in Biomarkers of Fertility

A large part of the literature on fertility trends has related to sperm concentration: this has been the focus not only of scientific but also of public concern—the ‘falling sperm counts’ story. The issue was first raised as far back as 1974 (Nelson and Bunge, 1974), but became more firmly established with Carlsen et al. (1992) which attempted to overcome the methodological problems of the earlier literature. They reviewed 61 studies published between 1938 and 1990 which had analysed sperm concentration using a haemocytometer, from couples who were not consulting for subfertility. Their conclusion was that mean sperm concentration had fallen from 113 to 66 million/ml between 1938 and 1991.

This led to renewed methodological controversy, and was followed by further publications most notably that of Swan et al. (2000), who concluded that sperm concentration had declined sharply in Europe between the mid-1940s and 1996, from 170 to <60 million/ml, and less sharply in North America between the early 1930s and 1996s from 110 also to about 60 million/ml. The Carlsen and Swan meta-analysis articles still met with considerable scepticism on grounds of non-uniformity of laboratory method, statistical issues, etc., so there has not been universal agreement that they demonstrate a real decline (Jégou et al., 1999).

A particular problem is the question of selection, as the participation rate in studies of semen quality is so low—30% is regarded as ‘good’—that even if the sampling frame is a random sample of the target population, it is unclear how representative the participating men are. They could for example be anxious about their fertility status following their own or a family member’s experience, creating substantial bias. On the other hand, groups such as candidates for semen donation could plausibly be expected to be biased in the opposite direction; the problem is that one cannot know the direction or the extent of any selection bias. This means that comparisons between different populations are especially difficult to interpret, even if the measurements are carried out identically, which is not necessarily the case.

The work of Carlsen et al. thus remained inconclusive, but it did stimulate the analysis of time trends by laboratories that had collected data on sperm concentration consistently over an appreciable period, starting in the early 1970s. These single-centre studies provided much more reliable evidence on whether or not there had been a decline, at least in their catchment population.

Thus, Auger et al. studied 1351 potential semen donors in 1973–1992 in the Paris area. Despite their initial scepticism of the ‘falling sperm counts’ hypothesis (Jégou et al., 1999), they demonstrated a convincing decline in mean sperm concentration from 89 to 60 million/ml in these 20 years, or nearly 2% a year, after adjustment for age and duration of abstinence (Auger et al., 1995). In addition, they found a fall in the proportion of motile spermatozoa of 0.6% a year, and of spermatozoa of normal morphology of 0.5% a year.

The findings were compatible either with a period effect, reflecting some (unknown) factor operating in 1973–1992 with a rapid impact, or with a birth cohort effect that permanently affected men born since about 1950. Although volunteers for semen donation are unlikely to be representative of the general population, selection bias would only affect the estimate of the trend if the degree of representativeness changed over the period of the study in such a way as to bring about the results; this seems unlikely, which is one reason why single-centre studies are preferable to the meta-analysis approach. In addition, the laboratory methods are likely to have been consistent over time in these single-centre studies.

Similar studies produced evidence of deterioration in semen quality in Gent (Van Waerebeek et al., 1996) and Edinburgh (Irvine et al., 1996), but not in Toulouse (Bujan et al., 1996), Finland (Vierula et al., 1996) or the five USA locations with published data (Wittmaack and Shapiro, 1992; Fisch et al., 1996; Paulsen et al., 1996). Where deterioration was observed, it tended to affect motility and morphology as well as density, as in the Paris study.

An odd feature of this literature is the way that it has been framed: the hypothesis has been of a global decline in sperm concentration, with an assumption that the level at any given time is the same in all parts of the world, or in the case of Swan et al., in the whole of North America and of Europe. If this were true it would be unusual, possibly unique, in all epidemiology. This way of posing it means that any item of evidence tends to be regarded either as confirming or refuting a global decline, rather than providing information on the population from which the data were drawn. Even high-quality reviews, such as Jégou et al. (1999), have tended to cast their judgments in terms of whether the global hypothesis has been confirmed or rejected—or more frequently, why it is inconclusive. And yet it is far more likely, from what is known about the behaviour of epidemiological variables in general, that levels would vary from place to place, and that if decline has occurred then it has not occurred everywhere simultaneously.

Such spatial variation in sperm concentration has been confirmed (Fisch et al., 1996; Fédération Française des CECOS et al., 1997; Jørgensen et al., 2002), and is considerable, e.g. 72.7 million/ml in California compared with 131.5 million/ml in New York (Fisch et al., 1996). There is now a considerable literature on semen quality in different parts of the world, for example in China, where the apparent spatial diversity is at least as great (Li et al., 2009).
The idea of a simultaneous decline, with similar levels across wide swathes of the globe, needs to be abandoned.

The best interpretation of this body of evidence would appear to be that during the period studied, deterioration has occurred in some places but not others. The alternative view is that the lack of decline in some studies means that the apparent deterioration in, for example, Paris must be illusory, perhaps due to random error or changes in laboratory technique, but this is undermined by the consistency of the finding that the different semen parameters have tended to decline together. This suggests that the observations result from something in the real world rather than a methodological issue, and is consistent with the known correlation between the various semen characteristics.

Spatial variation has so far proved less informative about the causes of any possible deterioration in the health of the male reproductive system. There is one notable exception, the contrast between Denmark and Finland discussed below.

The ‘falling sperm counts’ hypothesis has also been examined in domestic animal species, based on the idea that a pervasive environmental factor would have affected them in the same way as men. The species involved have been bulls (twice), boars, rams and stallions (Setchell, 1997; Van Os et al., 1997). In no case was a decline observed, and in sheep a slight but statistically significant increase was seen. It was only possible to study sperm concentration, so there is no information on the other semen quality parameters. These findings need to be interpreted with care, as factors such as nutrition and animal selection could have improved during this period, and the places where the animals were studied did not necessarily coincide with regions where human semen quality declined, in addition to possible technical problems with the particular analyses (Jégou et al., 1999).

If semen quality did decline at least in parts of Europe at some period in the second half of the 20th century, two questions then arise. First, any such decline did not necessarily begin at the point at which the evidence began to be collected. It is clearly impossible to know what happened before then, but at least it is important to be aware of the possibility that we are merely looking at a late stage in a longer process of decline. It is noteworthy that Swan’s meta-analysis found that in 1996, the end of her period of study, the mean semen density in North America and in Europe had become similar, ~60 million/ml (Swan et al., 2000). This raises the question whether the mean North American sperm concentration was at some time similar to the 170 million/ml observed in Europe in the mid-1940s, before the first observations were made in the early 1930s in North America (110 million/ml). This would mean that sperm concentration had already declined in an earlier period, and Europe was catching up with North America in the post-war period. In fact, more broadly it is unclear when such a decline might have begun, nor what the levels were before semen quality started to deteriorate. As human spermatogenesis is far inferior to that in other mammals (Sharpe, 1994; Setchell, 1997; França et al., 2005), it is possible that there is a long but invisible history to this decline, a possibility that is discussed further below.

Secondly, what impact is this likely to have had on fertility, in the sense of a man’s ability to father a pregnancy? It is likely to be small: it is true that Carlsen et al. showed that the proportion of the population with low sperm concentration increased during the period of their study, as the whole distribution shifted leftwards. However, calculations based on our (admittedly imperfect) knowledge on the relationship between sperm concentration and fertility show that the observed decline would have had rather a small impact on the length of time that a couple takes to conceive (Slama et al., 2004; Leridon and Slama, 2008), let alone affecting achieved family size.

**Trends in Functional Fertility**

The debate on a possible decline in semen quality stimulated research aimed at ascertaining whether couple fertility has declined. The first such study took place in Britain, and was a cross-sectional survey of a representative population sample, in which participants aged 16–59 years were asked about their first pregnancy (Joffe, 2000). In particular, they were asked how long it took to conceive—the Time to Pregnancy (TTP)—using the question, ‘How many months or years did it take you to become pregnant?’ for women, and substituting ‘your partner’ for ‘you’ in the case of men. This retrospective method of ascertaining TTP had previously been used in other contexts, e.g. to investigate possible reproductive hazards in the workplace. It had already been established that retrospective recall of TTP can be used in this way: the TTP distribution obtained gives a good estimate of the true distribution, even though the individual values are often inaccurate by a month or two (Baird et al., 1991; Zielhuis et al., 1992; Joffe et al., 1993, 1995), and this remains true even with a recall period of up to at least 20 years (Joffe et al., 1995; Jensen et al., 2005). [A small recent study has called the latter assertion into question (Cooney et al., 2008), but in fact the inaccuracy they describe at the individual level is very similar to that found by Joffe et al. (1993) and Joffe (in press); the issue is whether the TTP distribution obtained is biased, as validity at the group level is all that has ever been claimed for TTP studies, and Cooney et al. do not report any such bias. Such measurement error introduces imprecision not bias, increasing the necessary sample size.] Such surveys are also highly acceptable to the participants. However, there are several other methodological problems that can arise in such studies, and great care is needed in their design and analysis (Weinberg et al., 1994; Joffe et al., 2005; Key et al., 2009). It is important not to exaggerate these: the allegation that ‘we may never know’ whether TTP has altered over time (Salmén et al., 2005) has been shown to be due to a series of misunderstandings (Joffe et al., 2006; Key et al., 2009).

The findings of the British study (Joffe, 2000) were contrary to expectations: fertility had increased rather than decreased during the period 1961–1993—in other words, couples had taken less time to conceive in the later period than earlier on, especially after about 1980. Similar results were obtained from the groups of female and male participants, which were independent, lending additional support to the robustness of the findings. A subsequent study among Danish twins born between 1931 and 1952, covering the period 1948–1997, also found no evidence for a decline in fertility (Jensen et al., 2005). It was less clear-cut than the British study about a possible increase: from the reports of the female participants fertility appeared to have increased slightly, and in the male group severe infertility (defined as not having any children despite having tried to conceive) appeared to have decreased over time.
A study from the USA found a 30% decline in subfertility defined as TTP ≥ 12 months during the period 1982–2002 (Stephen and Chandra, 2006). Like the previous two cited studies this was based on a nationally representative sample, in this case of 15 303 women, and included couples who remained unable to conceive despite unprotected intercourse. This design is superior to pregnancy based samples (Joffe et al., 2005).

A large pregnancy based study in Sweden examined trends in subfertility, defined as TTP ≥ 12 months, using data from the Swedish Medical Birth Registry from 1983 to 2002 (Scheike et al., 2008). A decline was observed, which was more apparent when analysed as a birth cohort effect than as a period effect: a decrease of about 35% from year of birth 1945–1975.

In contrast with the aforementioned articles, one study has found a decline in fertility, but this could not be evaluated by the present author because it is only available in Finnish [reported in Sallmén et al. (2005)].

We urgently need further studies of this issue using the best available methods, and taking a better account of behavioural factors that could influence fertility (Joffe et al., 2005). These behavioural factors include knowledge of the days of the menstrual cycle during which a woman is fertile, and persistence in trying to conceive (Basso et al., 2000). They could affect the probability of conception, and therefore the TTP distribution, for any given degree of biological impairment. In addition, the fertility of a couple may be affected by relational factors between the partners that affect biological capacity, for example inbreeding (Helgason et al., 2008).

### How Can the Observations on Fertility and on Semen Quality be Reconciled?

The evidence that we have, which is admittedly imperfect, strongly suggests that semen quality has deteriorated in some localities. On the other hand, most available studies suggest that couple fertility has increased. This combination could simply result from a coincidence, that the TTP studies took place in regions of increase, although the reverse was true of some of the semen quality studies. Otherwise, we have an apparent contradiction requiring some explanation. It may be more apparent than real: we have already noted that the impact of the observed decline in sperm concentration on TTP is calculated to be small (Slama et al., 2004). This would mean that a relatively small influence in the opposite direction would more than counteract it.

The nature of such an influence is unknown. It is not necessarily biological: for example, it could be that knowledge of the days of the menstrual cycle during which a woman is fertile has increased, and diffused through the population during this period. A possible type of biological explanation, raised in two of the above-mentioned articles (Stephen and Chandra, 2006; Scheike et al., 2008), is that the impact of sexually transmitted infections (gonorrhoea and chlamydia) on fertility has lessened. However, it appears likely that the prevalence of cases of infection severe enough to reduce fertility is too small to have had a population-wide impact on TTP (Guzick and Swan, 2006; Ochsendorf, 2008). Alternatively, female fertility could have increased, e.g. as a late result of the 20th-century decline in food scarcity in developed societies.

Another possible biological explanation relates to population mixing: during the 20th century in the developed world, with urbanization and improvements in transport, it became increasingly likely that couples would be from distant populations both spatially and genetically, rather than from the same or a nearby village or urban neighbourhood. This would be expected to result in hybrid vigour. Both this hypothesis and the effect of diffusing knowledge could readily be studied by incorporating them into the type of TTP study described above.

### Fertility in the Context of Reproductive System Health and Offspring Development

In addition to examining the descriptive epidemiological data, it is important also to take a mechanistic approach. This is because a change in fertility could signal something about the underlying biology, implying that parallel changes are occurring in other reproductive system end-points in either sex, and/or could affect the offspring. In particular, abnormal gametes could underlie impairment both in fertility and in embryonic, fetal, and child development and health via genetic and/or epigenetic pathways. It is more difficult to detect trends or variations in these offspring-related outcomes, because typically a particular underlying lesion can have a number of different phenotypic manifestations: e.g. exposure to a mutagen can result in a multiplicity of different mutations; and genomic disruption could lead to many different forms of aneuploidy, e.g. various trisomies or monosomies, often with mosaicism. In either case it is a scattergun effect, a single cause giving rise to multiple outcomes (Joffe, in press), each of which is uncommon and therefore difficult to study directly. Subfertility can therefore be regarded as a sentinel indicator of impaired gametes, and it would be difficult to overstate the importance of gamete quality for offspring development and health.

Unfortunately the research worlds of reproductive biology and of zygote/embryonic/fetal development tend to be separate. Scientists in one area are not necessarily aware of work in the other, nor mindful of the importance of their own work for the other area. However, some direct evidence for a link between male subfertility and zygote impairment does exist: embryos resulting from in vitro fertilization where the male partner had poor semen quality show several types of abnormality. These include multiple centrosomes, distorted cellular architecture, absent or asymmetric cytokinesis so that binucleate cells are common, and missegregation of chromosomes (Chatzimeletiou et al., 2005). Many embryos fail to develop beyond the early stages (Tesaryk et al., 2002), and mosaicism is common in the survivors. A central feature appears to be played by the zygotic centrosome, which is paternally inherited in humans and other mammals although not rodents (Chatzimeletiou et al., 2005, 2007).

More attention has been paid to the link between subfertility and reproductive system health, at least in the male case. The declining quality of spermatogenesis and the observed rise of young-onset germ cell testicular cancer during the 20th century (Davies, 1981; Adami et al., 1994; Bergström et al., 1996) may be connected (Skakkebaek et al., 2001). It is also possible that two important
congenital anomalies of male development, hypospadias and cryptorchidism, have increased during a similar period, but this is far more difficult to investigate as the rates reported by congenital anomaly registries depend on the level of ascertainment, and such data are hard to interpret (Toppari et al., 2001). This linkage in trends is plausible because these four outcomes tend to cluster in individuals, each of these features being associated with each of the others, for example reduced fertility is present in testicular cancer patients even some years before the cancer appears (Møller and Skakkebaek, 1999)—although it is rare for all four to be present in a single person. Such individuals also have certain characteristic features such as the presence of Sertoli-cell only tubules and microlithiasis (Skakkebaek et al., 2001). It is common practice to call this the ‘testicular dysgenesis syndrome’ (TDS), although it may be more precise to think in terms of one or more common risk factors and commonality of underlying pathogenesis than a syndrome in the usual sense.

There is good evidence that the abnormality arises in early life, possibly in utero, as testicular cancer is preceded by carcinoma-in-situ (CIS) in boys (Skakkebaek et al., 1987), and abnormalities of spermatogenesis are life-long and already present in very young men, e.g. as seen in studies of armed forces conscripts (Jensen et al., 2004a); hypospadias and cryptorchidism obviously arise in utero. An epidemiological reason to suppose that these end-points are linked is from a spatial comparison rather than one concerning trends: that high rates of all four are found in Denmark but much lower rates in Finland (Giwercman et al., 2006); Finnish couples also appear to have relatively high functional fertility (Joffe, 1996).

On the other hand testicular cancer has increased greatly among white men in the USA, but as we have seen the evidence shows little or no deterioration in semen quality there. And conversely, men in Japan and China have semen quality no better than those in Europe or North America (Iwamoto et al., 2006; Li et al., 2009), but very low testicular cancer rates (Parkin, 2005). Furthermore, spatial variations in functional fertility, for example the higher fertility in parts of southern compared with northern Europe (Karmaus and Juul, 1999), do not fit any pattern that is yet convincing. Clearly the factor(s) responsible for the rise of testicular cancer in the 20th century do(es) not explain all of the observed variation in male reproductive system impairment.

The descriptive epidemiology of testicular cancer has been well characterized, and if the semen quality issue is indeed linked it provides useful information. Unlike semen quality and functional fertility, testicular cancer is easy to study epidemiologically. This is because it is a disease of young men, typically occurring between the ages of 20 and 45 years with a median of 30, with unmistakeable clinical features. Cancer registry data are therefore highly reliable. In the early20th century it was also invariably fatal, so that mortality data can be used as a proxy for incidence in that period (Davies, 1981); in recent decades a high cure rate has been achieved, so that incidence data must be used.

The main feature of testicular cancer epidemiology is that the incidence has risen dramatically in about five decades, at least in populations of European ancestry and in certain other groups as well, notably Polynesian peoples including Maoris in New Zealand (Davies, 1981; Adami et al., 1994; Bergström et al., 1996; Parkin, 2005). Other populations have maintained low rates, notably China and Japan, and black men in the USA (Parkin, 2005). The typical rise has been 3–4-fold, sometimes more. A similar pattern has been seen in the two major subdivisions of testicular cancer, seminoma and non-seminoma, and in general they share all the major epidemiological features. The increase started in about 1920 in England and Wales (Davies, 1981), and the mid-20th century in the Nordic countries, Germany and Poland (Bergström et al., 1996). But these dates are misleading, as they represent the period during which incidence occurred, whereas the clear evidence on early-life onset implies that a birth cohort analysis should be used. This is reinforced by the observation that a pause in the inexorable rise of testicular cancer occurred among men born during 1935–1945 in Denmark, Norway and Sweden—although not in Finland, East Germany or Poland (Bergström et al., 1996)—indicating that birth cohort analysis is the appropriate method. This would place the start of the increase in the late nineteenth century in England and Wales, around 1905 in Denmark, Norway and Sweden, and around 1920 in East Germany, Finland and Poland (Bergström et al., 1996).

It is important to note, therefore, that the strong evidence for early-life origin together with the incidence data place the factor(s) responsible for the rise of testicular cancer—or at least its early part—clearly in the early 20th century (a few decades earlier in England and Wales according to the mortality data). This apparently elementary point has continued to elude scientists in this area, who persist in referring to a rise beginning in the mid-20th century, for example Sonne et al. (2008): 'The incidences of male reproductive disorders including impaired spermatogenesis, hypospadias, cryptorchidism and testicular cancer have increased during the last 50 years' [emphasis added].

An intriguing aspect of the four end-points that have been grouped under the ‘TDS’ heading is that each of them has a tendency to cluster in families, with evidence that this has a genetic basis rather than being due to shared environment at least for two of the outcomes. There is strong evidence for the inheritance of testicular cancer risk, although the heritable form is rather uncommon (reviewed in Lutke Holzik et al., 2004). There is also a genetic component in impaired semen quality and male fertility (Czyglik et al., 1986; Lilford et al., 1994; Meschede et al., 2000; Christensen et al., 2003; Storgaard et al., 2006a; Van Golde et al., 2004; Gianotten et al., 2004; reviewed in Joffe, 2007), with some evidence for increased testicular cancer risk in families with male infertility (Richiardi et al., 2004; Richiardi and Akre, 2005; Aschim et al., 2008). For both outcomes, the results suggest a highly heterogeneous picture, not implicating any specific locus, and with no single clear mode of inheritance. Cryptorchidism and hypospadias also appear to have familial aggregation, and boys with hypospadias tend to be born to fathers with impaired fertility (Czeizel and Toth, 1990; Fritz and Czeizel, 1996; Weidner et al., 1999; Fredell et al., 2002; reviewed in Joffe, 2007). This common pattern provides a further indication that these four conditions share a common risk factor and pathogenic mechanism.

This leaves us with an intriguing combination, which has been called a ‘conundrum’ (Harland, 2000): the rapidity of the observed trends must be due to a change in one or more environmental factors, it could not have a genetic origin. This is reinforced by other evidence, e.g. from migrant studies, which show—as most migrant studies in epidemiology do—that the risk of testicular cancer in migrants moves from that of the country of origin towards the country of adoption, but with the unusual feature that this change is not seen until the next generation (Parkin and Iscovich, 1997; Montgomery et al., 2005).
This slowness of adaptation fits with the early life origin of the disease noted above. On the other hand, there is an undeniable element of heritability, even if it only affects a minority of cases. The question is, how to reconcile these two?

One possibility is simply to apportion the risk between ‘nature’ and ‘nurture’. A second is to analyse this in terms of gene-environment interaction (Skakkebaek et al., 2001). A third, which seems likely to apply in this case, is that the genetic defect is itself caused or made more frequent by an environmental factor (Joffe, in press), just as a mutagen can cause cancer in somatic tissues. How this might work is discussed below. These three possibilities are not mutually exclusive.

An example of an environmental agent leading to reproductive impairment in male offspring, in this case via an epigenetic rather than a genetic mechanism, is the work of Anway et al. (2005) on the anti-androgen vinclozolin using experimental rats. It is unclear how relevant these observations are to explaining the rise of TDS, partly because adult-onset disease is also induced, including prostate and kidney disease, which are not features of TDS (Anway and Skinner, 2008), and partly because population exposure to this agent (or to any other similar one) is more limited and recent than the descriptive epidemiological changes we are trying to explain.

**The Broader Context: Reproduction in the Human Species**

The heritability of a cluster of end-points including testicular cancer and subfertility presents two puzzles. One is that the idea of transmissible subfertility is apparently absurd: evolution is driven by the probability of passing one’s genes on to the next generation, so any gene that reduced fertility would be eliminated from a population within a relatively small number of generations [unless spermatogonial cells with mutations have a selective advantage (Goriely et al., 2003; Zollner et al., 2004; Choi et al., 2008)]. This could be resolved if the genetic predisposition towards TDS were constantly entering the population—which fits with the evidence of an environmental determinant, and also accords with the low proportion of familial cases.

This idea is complementary to Czeizel’s ‘relaxed selection’ hypothesis: that the decline of family size in the early 20th century meant that the most biologically fertile couples ceased to dominate the proportion of births in the population, and also that more recently the availability of assisted reproductive techniques has increased the reproductive presence of couples with impaired fertility (Czeizel and Rothman, 2002). There are two reasons why Czeizel’s hypothesis alone cannot account for the epidemiological observations. First, the calculations of Slama and Leridon show that the observed trend in low sperm concentration was too strong to be explained by this process, at least in France (Slama and Leridon, 2002); this would hold even more strongly for testicular cancer which has had a stronger trend but for which the selection process would largely be indirect through the associated subfertility. Secondly, the hypothesis depends on the existence of alleles for subfertility and the related conditions, but it does not explain how they came to be present in the population in the 19th century, before the relaxation in selection began. It thus requires the complementary hypothesis that postulates the generation of new damage.

The second puzzle is that there is no animal model for germ cell testicular cancer. This is odd because testicular biology is highly conserved through evolution, being essentially the same among all mammals.

This relates to a more general question: how does semen quality in humans compare with that in other mammals? Quantitatively, it is markedly inferior: according to Sharpe, daily sperm production is lower in man than in any other mammal: 4.4 million/gm of testis compared with 13 in the bull, 16 in the stallion and >20 in the six other species for which data were available (Sharpe, 1994). Furthermore, germ cell degeneration during meiosis, which is probably a quality control mechanism, is 39% in man compared with 4–6% in most other species apart from the bull (20%) and possibly the rabbit (Sharpe, 1994). This is reinforced by the observations of França et al. (2005) who report that the number of spermatids per type A1 spermatogonium is 3.2 in humans compared with 26–97 in 10 other mammalian species. Furthermore, the number of spermatids per primary spermatocyte is 1.3 in humans, the other species being in the range 2.8–3.6; and the loss of germ cells compared with the theoretical yield is 68% in man compared with 10–30% in the other species (França et al., 2005).

Not only semen quality but also functional fertility is lower in humans than in other mammals. Human couples have a fecundability, defined as the probability of conceiving in any particular menstrual cycle, of only 0.3 or lower, whereas in other mammals it is far higher (Wood, 1994; Zinaman et al., 1996).

These are not the only features of reproductive biology in which humans fare markedly worse. Of conceptions that successfully implant, over 30% (Wilcox et al., 1988) or more (Norwitz et al., 2001; Coonen et al., 2004) are lost in humans, far higher than other mammalian species. In studies of fertilized oocytes matured in vitro, 3.0–42.8% become aneuploid in humans, with an estimated overall frequency of about 13–14%, compared with usually less than 10% in other species—4.9, 9.0 or 14.2 in pigs, 2.2–7.1 in cattle, 5.8 in horses and in rabbits and 2.5–3.5 in mice (Sosnowski et al., 2003). Apart from semen quality, these other outcomes—lower fecundity, higher rates of early loss and aneuploidy—can in principle be either male- or female-mediated. In practice the majority are female-mediated at least for aneuploidy, excepting the sex chromosome aneuploidies such as 47,XXY (Klinefelter syndrome) (Delhanty, 2001). Maternally-mediated anomalies are difficult to study, and good evidence is restricted to the in-vitro fertilization context which is clearly a selected population. Nevertheless, the high incidence of chromosomal abnormality and mosaicism is striking (Vanneste et al., 2009a, b), and is specific to meiosis and/or embryogenesis; nothing comparable is observed in cultured somatic cells (Delhanty, 2001). Chromosomal anomalies of maternal origin are very important in practical terms because the aneuploidies that survive to birth have profound consequences on the health and life chances of the offspring, the commonest being trisomy-21, the most frequent anomaly underlying Down syndrome.

However, the vast majority of autosomal aneuploidies are non-viable, and conversely most early miscarriages have serious genetic/genomic anomalies, hence aneuploidy and early loss are closely connected. It is also highly plausible that the other types of impairment
are manifestations of the same underlying pathogenic process, and not just because they are all widespread in humans.

Low fertility and predisposition to early loss are linked: the genetic or genomic abnormality that makes an embryo non-viable is likely to interfere with ovum fertilization and/or with gamete survival and function. At least in the male case there is some direct evidence for a link, as already described, in the centrosome amplification and disturbance of cell structure, leading in severe cases to chromosome missegregation, which characterizes the zygotes from sperm of severely subfertile men (Chatzimeletiou et al., 2005, 2007). And it is not only poor semen quality, infertility, early loss and aneuploidy that are linked, but also cancer: Chatzimeletiou et al. (2005) report that under the microscope, the severely affected embryos display ‘genetic instability similar to that observed with human tumour cells’, and others have made similar observations (Delhanty, 2001; Vanneste, 2009b).

Humans are thus peculiarly prone to a range of reproduction-related impairments, suggesting the operation of a human-specific pathogenic process that involves genetic instability, and leading to cell-structure distortion in the more severe cases. This may explain why testicular germ cell cancer is a uniquely human disease.

The same process would be likely to occur in both sexes. Clearly the processes that have been observed in the male do not necessarily apply directly in the female case. Germ cell cancer also occurs in women, and ovarian cancer like testicular cancer tends to occur in younger age groups, but unlike the latter it has not shown a rising incidence (Goodman and Shvetsov, 2009), nor does it have such a clear association with infertility (Cetin et al., 2008). The biology is quite different in the two sexes, with the onset of meiosis early in the life of females, followed by a long quiescent period in metaphase II of meiosis. These differences in male- and female-mediated processes need more attention in the light of the issues discussed in this article.

If the pathogenic process that is relatively visible in the male line is relevant also, mutatis mutandis, to female-mediated pathology, this would include aneuploidies such as trisomy-21. Their high birth prevalence means that the practical consequences for health would then be very important. It is difficult to obtain good evidence that could throw light on the reason for the high human rate of aneuploidies. One observation is skeletal evidence indicating that Down syndrome was less frequent in Europe in the four millennia to 800 AD than in the populations we are familiar with (Czametzki et al., 2003). This suggests that some at least of these ‘human-specific’ characteristics have arisen on a historical/sociological rather than an evolutionary timescale, an idea which we now consider more fully.

Possible Explanations for these Observations

Why then do humans have so many reproductive and developmental problems? And why have some of them increased in recent decades in some populations?

To answer the first question, it is necessary to consider what is really meant here by ‘humans’. The evidence available for most of these end-points is mainly confined to people living in economically developed societies, and in most cases the assessments only started relatively recently. We can be confident that testicular cancer did not occur before its onset in late nineteenth century England and Wales: as argued above, testicular cancer epidemiology is reliable; Davies (1981) demonstrated that during this period, a type of testicular cancer confined to old men was replaced by the modern young-onset type. Semen quality testing is only available from the early 1930s in North America and later in Europe, with a few studies in other regions but none in populations that have been untouched by ‘civilization’—research that could prove highly informative. Fertility (TTP) research in the modern sense similarly only covers a recent period, but numerous studies of ‘natural fertility’ have been conducted by demographers, focusing on time from marriage to first birth, and it turns out that the estimates of subfertility are rather similar to those using the TTP methodology (Joffe, 1989). Studies of early loss and of aneuploidy appear to be confined to rather recent studies in developed countries. This means that apart from functional fertility, the term ‘humans’ here could refer to the species as a whole, or merely to a rather historically and socio-spatially select group, or somewhere in between. Even the natural fertility studies are confined to populations with a western lifestyle, albeit in an earlier historical period.

In principle, such observations could be compatible with a number of different pathogenic processes. In this section the hitherto dominant view, the endocrine disruption hypothesis, is considered along with a newly suggested way of synthesising the available evidence. Others are possible, and the different hypotheses are not necessarily mutually exclusive. For example, gestational genetic damage to primordial germ cells has been experimentally induced using ethynitrosourea, a highly potent mutagen (Brinkworth, 2000), and it is possible that pregnant women are exposed to similarly mutagenic agents e.g. in cigarette smoke. Maternal smoking during pregnancy is a possible risk factor for reduced semen quality (Jensen et al., 2004a, b), although this finding is not universal (Richtoff et al., 2007). However, seven studies agree that it does not predispose to testicular cancer (Pettersson et al., 2007).

The endocrine disruption hypothesis

One hypothesis to explain the deterioration in the aspects of male reproductive health grouped together as TDS is that it is due to some form of endocrine disruption or modulation. This has been highly influential since the early 1990s, to the extent that it has not only driven out discussion of other hypotheses, but has also dominated the debate about the male reproductive system—major research programmes on endocrine disruption were initiated in the USA and Europe, and in most countries it was only possible to carry out research on the epidemiology of male reproduction if the project could be presented as a test of the endocrine disruption hypothesis.

The earliest version was the suggestion that environmental agents with estrogenic properties were responsible for the manifestations of TDS (Sharpe and Skakkebaek, 1993). Such chemicals include bis-phenol A, which is present in numerous consumer products including those to which pregnant women and babies are exposed, the surfactants octylphenol and nonylphenol that are widely present in the environment, and estrogenic compounds in plants (phytoestrogens), especially in soya. This idea built on earlier epidemiological observations that the risk of testicular cancer (Henderson et al., 1979) and of cryptorchidism (Depue, 1984) was elevated in boys whose
mothers had had pregnancies characterized by excessive vomiting or high maternal weight, which are associated with high levels of endogenous maternal estrogens such as estradiol. But it immediately raised the problem that environmental exposures are minuscule (lower potency $\times 10^3$) compared with the mother’s estradiol (Safe, 1995), and the reply that the latter was 99% protein bound did not meet this objection. Also, soya consumption is high in Japan and China but their testicular cancer incidence is very low (Parkin, 2005). One tragic natural experiment throws light on this issue. Diethylstilbestrol (DES), a synthetic non-steroidal substance that has similar potency to estradiol and crosses the placenta, was prescribed for pregnant women in the 1940s, 1950s and 1960s, in the mistaken belief that it could prevent threatened miscarriage. Pharmacological doses were given in early pregnancy. It is known that girls exposed in utero developed genital anomalies and (fortunately rarely) vaginal cancer (Herbst et al., 1971). Follow-up of the boys born to women who had taken part in a clinical trial of DES revealed some genital anomalies, but not those characteristic of TDS; the men were also normal in functional fertility using a wide range of measures including TTP (Wilcox et al., 1995). Review of the evidence on DES and male offspring reveals none of the features of TDS, except testicular cancer (probably), and possibly lower sperm concentration in the highest dose groups (Storgaard et al., 2006b). Even in the case of testicular cancer, the possible increase was lower than the historical increases that have been observed (Davies, 1981; Adam et al., 1994; Bergström et al., 1996). [Intriguingly, an increased risk of hypospadias has been observed in boys whose mothers were exposed in utero to DES (Klip et al., 2002; Brouwers et al., 2006); the possible mechanism involved is obscure.] From this and other evidence it became belatedly accepted that the estrogen hypothesis was not the explanation for the rise of TDS (Sharpe, 2003).

A more plausible version of the endocrine disruption hypothesis focused on anti-androgens such as dichlorodiphenyldichloroethylene (DDE), the stable breakdown product of the insecticide dichlorodiphenyltrichloroethane (DDT), and certain phthalates which are used as plastic softeners. Both of these date from the 1930s, so that one obvious objection is that at least the start of the rising trend in testicular cancer occurred before any of the known anti-androgens were introduced.

The evidence against DDE is strong: the striking contrast between Denmark and Finland cannot be explained by exposure to DDE, which is similar during this period (Ekblom et al., 1996) or only slightly different (Shen et al., 2008). Moreover, high levels of exposure to DDE in developing countries, in the course of attempts at malaria control, would by now have resulted in an epidemic of testicular cancer, and this has not occurred (Joffe, 2001).

The situation is different with phthalates. Experimental evidence shows that dibutyl phthalate in large enough doses can cause some features of TDS in rats—although not testicular cancer (Fisher, 2004). In humans it has been reported that reduced anogenital distance, a feature of the ‘phthalate syndrome’ in rats, is associated with phthalate exposure in newborn boys (Marsee et al., 2006). Thus, unrealistically large doses of phthalates can mimic some features of TDS; actually existing doses appear to have some impact on male reproductive development, but trivial compared with the magnitude of what we are trying to explain. In short, the effects of phthalate exposure are too little, too late.

It is clearly important to protect people from a possible reproductive toxin—but this is very different from saying that phthalate exposure has caused the rise of TDS, or even contributed to it in any important way—the descriptive epidemiology is far more dramatic than these findings (Joffe, in press). It is therefore clear that the rise of TDS is not mainly due to anti-androgens—unless there is a highly potent anti-androgen waiting to be discovered that was widely disseminated in the early 20th century.

Another way of looking at the endocrine disruption hypothesis is to ask what epidemiological changes would be expected a priori from exposure to endocrine disrupters of various types: which end-points would be affected, in which sex? One plausible expectation is of a coherent pattern in hormone-sensitive cancers, but there is no clear evidence for this (Joffe, 2001); more research on the topic would be welcome. A second is that endocrine agents would be expected to influence growth and development, secondary sexual characters and the timing of puberty. If the rise of TDS were attributable to endocrine disruption, a population-level change in e.g. age at puberty would be predicted. No such change has been reported among boys, either in Europe or America (Joffe, 2001). Furthermore, it is far from clear that endocrine disruption would affect non-quantitative aspects of semen quality, especially morphology. In summary, a thorough review of the evidence leads to the conclusion that the endocrine disruption hypothesis cannot explain the main features of the rise of testicular cancer or more broadly of TDS.

An alternative explanation
A different possibility is that the manifestations of TDS are due to environmentally induced genetic damage to the germ cells. This hypothesis is set out fully elsewhere (Joffe, in press); see also the Supplementary data appendix. It postulates an intergenerational process that starts with small duplications and deletions (D&D). The initial impairment is relatively minor, but the severity increases over time and across generations. In particular, in each generation the intricate process of meiosis is the stage at which the damage becomes more severe. This is because meiosis, unlike the simpler process of mitosis, starts with the pairing of homologous chromosomes, and the D&D mean that the DNA strands have become structurally different, leading to delay in completing meiosis—such delay has been observed in several clinical studies of infertile men, using different techniques (Vendrell et al., 1999; Guichaoua et al., 2005; Codina-Pascual et al., 2006). The resulting cell cycle delay leads to centrosome amplification and thus to loss of normal cellular architecture, and as the severity increases, to chromosome missegregation leading to aneuploidy. This could explain why intra-cytoplasmic sperm injection appears more likely to be successful if the selected spermatozoa has a smooth, symmetric, oval-shaped nucleus of normal size (Berkovitz et al., 2005).

At each cell division, especially later in the process, many abnormal cells are produced that are not viable, and the surviving cell lines are those able to pass through the cell checkpoints and other quality control mechanisms such as fertilization. Thus although the abnormality worsens as the process proceeds, there is an upper limit to severity in surviving clones, as the more affected ones simply disappear in a Darwinian-like process of natural selection.
The initial lesion in the first (F0) generation is postulated to be fragmentation of spermatozoid DNA. Abnormal repair in the F1 zygote alters the DNA structure, so that at F2 meiosis the cellular disruption begins. Transmission in the male line occurs because the human zygote’s centrosome is paternal in origin. In the F2 generation the severity increases, with the genomic apparatus now being so distorted in the more severe cases that even mitosis is affected. This can lead to mosaicism, with different clones showing different abnormalities—a scattergun effect, as previously noted. This feature, together with the persistence and propagation of existing defects, leads to each affected individual having a characteristic spectrum of abnormalities, or ‘fingerprint’ (Joffe, in press). In the most severe cases, chromosomal instability and loss lead to aneuploidy together with a block in differentiation, which are the characteristics of CIS cells, and some of these begin to develop into invasive testicular cancer after F2 meiosis exacerbates the abnormality further.

The process involves the genetic apparatus as a whole, rather than being a mutation of specific genes, so that alteration of gene expression is secondary to structural changes, at least in the early stages. However, as chromosomal instability increases, structural DNA damage leads to deletion and/or amplification of specific genes or chromosome regions in a rather random manner. If relevant loci are affected, they can play a major part in the development of the disease, e.g. chromosome region 12p has a major role in metastasis, introducing a positive feedback loop. This is similar to the malignant process that is thought to occur in numerous other forms of cancer (Nigg, 2002; Gibbs, 2003; Pihan et al., 2003; Sieber et al., 2003; Storchova and Pellman, 2004; Breivik, 2005; Kops et al., 2005).

Thus, TDS is seen as a transgenerational condition with relatively mild sub-malignant and more severe malignant forms. Only in the F2 generation is it recognizably TDS, before that it is the milder precursor form. This process explains why the zygote has abnormalities that also appear to be cancer-like. If transmitted to the F3 and future generations it becomes the familial form of the condition. Subfertility features in each generation, but with varying pathogenesis; only in the F2 generation does it become part of TDS, before that it could be considered a risk factor for TDS. Low sperm concentration results from loss of germ cells due to apoptosis, and impaired morphology and motility result from the abnormal cellular architecture. However, not all features of TDS are readily explained by this hypothesis, notably the occurrence of cryptorchidism and hypospadias. In addition, much male subfertility may be unrelated to TDS.

This process should not be confused with ‘genetic’ in the usual sense of ‘inherited’, rather the primary mechanism is genetic damage that is usually of environmental origin but may be inherited in a minority of cases. In addition this pathogenic process can be modified by truly inherited factors that might be manifest as, for example, ethnic variation (Giwercman et al., 2006; Joffe, in press). One candidate is the zygote’s repair capacity, which is maternally mediated, and exhibits variability of sufficient magnitude.

### Possible exposures

If this account is true, the responsible environmental factor(s) would need to be able to cause DNA damage to spermatozoids, as well as having exposure characteristics that correspond with the descriptive epidemiology. Whatever early-life exposure is responsible, the process could also be influenced by protective and exacerbating factors acting in adult life, for example antioxidant consumption (Eskelinen et al., 2005) or paternal occupation (Knight and Marrett, 1997).

Some features of the descriptive epidemiology are compatible with the idea that maternal cigarette smoking plays a role—which would have a different pathogenesis from that suggested above. Smokers’ sons do indeed have lower sperm concentration (Ramlau-Hansen et al., 2007), but maternal smoking is not implicated in testicular cancer risk, and may even be protective (Weir et al., 2000).

Numerous agents are able to damage spermatozoid DNA, sometimes via oxidative stress (Aitken et al., 2006). They include tobacco smoke; cis-unsaturated fatty acids; glycidamide, a metabolic product of acrylamide, which is present in many foods especially starchy food cooked at high temperature (including chips and crisps) (Joffe, in press); and heat. The latter is particularly interesting as sedentary occupations have increased over the last century, and are associated with raised intra-scrotal temperature (Bujan et al., 2000; Mieusset et al., 2007). Increasing the temperature, e.g. by scrotal insulation, has been shown to damage sperm DNA in experimental animals as observed using the SCSA technique (Karabinus et al., 1997; Sailer et al., 1997; Banks et al., 2005), but it is unclear whether similar effects would occur in sedentary men. It may be relevant that some studies of testicular cancer have shown a positive association with socioeconomic status (Akre et al., 1996), compatible with a role for non-manual occupations which tend to be more sedentary. More speculatively, the milder reproductive impairment seen in pre-industrial men could possibly be related to the wearing of underpants and trousers, which results in a raised intra-scrotal temperature even compared with men living in hot climates (Mieusset et al., 2007).

In addition, air pollution from burning soft brown coal is associated with sperm DNA fragmentation (Rubes et al., 2005). The observation that the rapid increase of testicular cancer in Denmark during the mid-20th century was for a time more marked in urban than in rural areas (Østerlind, 1986) is compatible with the heat and coal-combustion hypotheses, among others.

One implication of the proposed pathogenesis is that as D&D accumulate, mating between individuals who are genetically unalike would be associated with lower reproductive success as pairing at the start of meiosis would be more likely to be impaired. On the face of it, this contradicts the earlier observation that hybrid vigour would tend to increase fertility in unrelated individuals. Yet these two ideas may be compatible: both extremes—genetic similarity (inbreeding) and genetic distance (D&D accumulation)—could decrease fertility, so that an intermediate degree of relatedness would be associated with the highest degree of fertility. This could explain the evidence from Iceland that the greatest reproductive success, measured as the number of grandchildren, was observed in couples who were third or fourth cousins (Helgason et al., 2008). Comparable findings have been reported from elsewhere, for example Denmark (Labouriau and Amorim, 2008).

### Conclusion

Fertility is important in its own right and also because it is linked to other conditions, both of the individual herself/himself and also those affecting the offspring, and possibly later generations as well. Some of these conditions are severe. There is evidence that the
health of the male reproductive system has deteriorated during the 20th century at least in some populations, for reasons that remain unclear, with the rise of testicular cancer being a particularly serious trend. It is unclear what has happened to female fertility, but couples have apparently become more fertile in a functional sense in some places at least. In addition, the human tendencies to inferior semen quality, lower fertility, higher rates of aneuploidy and of early loss compared with other mammals may be linked; this appears to involve both male- and female-mediated pathways. The pathogenesis involved remains obscure but could involve an intergenerational process of environmental origin that induces genetic damage in germ cells.

Supplementary data

Supplementary data are available at http://humrep.oxfordjournals.org/.

Acknowledgements

I would like to thank Fernanda Almeida, Sue Barlow, Alan Boobis, Martin Brinkworth, Leendert Looijenga, Ian Morris, Eva Rajperts-de Meyts, Niels Skakkebaek, Frank Sullivan and Allen Wilcox for discussion of the issues dealt with in this paper. They do not necessarily agree with the views expressed in it.

References


Joffe M. Decreased fertility in Britain compared with Finland. Lancet 1996;347:1519–1522.


Shen H, Main KM, Andersson AM, Danquah IN, Virtanen HE, Skakkebaek NE, Toppan J, Schramm KW. Concentrations of persistent organochlorine compounds in human milk and placenta are higher in Denmark than in Finland. Hum Reprod 2008; 23:201–210.


Skakkebaek NE, Raijerts-de Meyts E, Main KM. Testicular dysgenesis syndrome: an increasingly common developmental disorder with environmental aspects. Hum Reprod 2001; 16:972–978.


