

Perspectives

Semen quality analysis and the idea of normal fertility

Michael Joffe

Department of Epidemiology and Public Health, Imperial College London, St Mary's Campus, London W2 1PF, UK

Abstract

The World Health Organization (WHO) has extensively revised its manual for semen analysis, and seeks to provide reference limits for semen quality parameters. This raises the question of what is meant by ‘normal’ (excluding the use of this term to denote a Gaussian distribution). It could be taken in a purely statistical sense, using a biologically arbitrary cut-off point to denote an abnormal level, typically the extreme 5 percent of the population. Alternatively, ‘normal’ could be defined according to the biological concept of normality and abnormality, in terms of the point at which biological function becomes impaired. Either of these can be used in descriptive epidemiology, for example, to study trends, but in the case of fertility, both semen quality and functional fertility (time to pregnancy) are continuous variables with no clear threshold. The WHO manual uses the biological meaning of normal, in that it provides the semen parameter distributions for men who have recently fathered pregnancies that took 12 months or less to conceive. However, what is really needed is the same information the other way around: given a particular semen test result, what should be expected in terms of ability to conceive, and how long it is likely to take. In considering epidemiological research, the focus has been mainly on internal comparisons, rather than reference limits, but it would be beneficial if more attention were paid to the absolute levels and to what these mean in terms of function—in other words, if the data were better calibrated biologically.

Asian Journal of Andrology (2010) 12: 79–82. doi: 10.1038/aja.2009.3

Keywords: fecundability, fecundity, fertility, semen, semen analysis, subfecundity

1 The new WHO manual

The World Health Organization (WHO) has produced a fifth edition of its manual for semen analysis that has been extensively revised [1]. In addition to the greater detail provided on laboratory methods, including quality control, it seeks to provide reference limits for the parameters of semen quality based on research findings that have been published since the earlier edition. This is highly commendable, even if the available data have imperfections, for example, being derived from studies with low participation rates, and in some instances from pregnancy-based samples, thus excluding less fertile couples and possibly leading to bias. In addition, the manual seeks to

relate results from semen analysis to the achievement of a recent pregnancy that took 12 months or less to conceive, but it is unclear how the authors handled the issue of accidental pregnancies (admittedly, this would tend to affect mainly the more fertile couples—see below). The aim of this paper is to examine the concept of reference limits.

2 Reference limits and the idea of normality

There are two possible ways to interpret the term ‘normality’ (when using the term in its conventional meaning rather than to indicate a so-called ‘normal distribution’, which is better referred to as a ‘Gaussian distribution’). One, it could be regarded as ‘statistical normality’, which aims simply to describe the world as it is, without invoking a biological criterion to demarcate where a subpopulation should be regarded as abnormal. The distribution of values is regarded as a purely statistical entity, and the terms ‘normal’ and ‘abnormal’ merely denote a particular proportion at one or both extremes of this distribution, often 5%.

Correspondence to: Dr Michael Joffe, Department of Epidemiology and Public Health, Imperial College London, St Mary's Campus, Norfolk Place, London W2 1PF, UK.

Fax: +44-20-7607-2446 E-mail: m.joffe@imperial.ac.uk

Received: 1 January 2009 Revised: 3 January 2009

Accepted: 13 January 2009



The second meaning makes a normative judgement, in that it aims to describe the range of values that represent biologically normal reproductive capacity when not subject to any impairment, and could be called ‘prescriptive normality’. In some situations in medicine this is quite clear-cut—for example, below a certain minimum level of blood cortisol adrenal insufficiency occurs, but an excess is also detrimental: Addison’s disease and Cushing’s disease, respectively. Here the meaning of a ‘normal range’ is straightforward in principle (if not necessarily in practice—additional tests beyond a blood cortisol level are used to make a firm diagnosis in both cases), as it refers to a level beyond which adverse health consequences ensue; that is, biological function is disturbed. In such cases, pathology is defined not in statistical but in functional terms, so that the rather arbitrary criterion of defining normality in terms of being within two standard deviations and therefore including 95% of the distribution (assuming a Gaussian distribution) is not relevant.

However, this clarity is not always present; a high level of plasma low-density lipoprotein (LDL) cholesterol is clearly related to the risk of ischemic heart disease, but there is no clear threshold beyond which it can be described as abnormal, as the association with increased risk applies across the entire existing distribution. Strictly speaking, plasma LDL cholesterol is a risk factor whereas semen quality is a biomarker, but the analogy is useful because both are used to predict future outcome in terms of biological function.

In descriptive epidemiology, for example, assessing trends or making spatial comparisons, statistical normality is highly suitable as it can be used to summarize and compare distributions. For example, Carlsen *et al.* [2] estimated the proportion of the population that had sperm concentrations below a particular cut-point at different times in order to focus on the lower tail of the distribution without using prescriptive normality. An alternative way is to present the quantiles of the distribution and how they change. If prescriptive normality is used in the context of descriptive epidemiology, for example, to describe trends, it is generally in terms of the proportion in the population that is in the abnormal range; for example, one could look at changes in the prevalence of Addison’s disease.

The WHO manual has opted for the second sense of normality, as the authors have used a biological criterion: recent fatherhood, with the additional restriction that conception had to have taken place within 12 months of the start of unprotected intercourse. The implicit concept of abnormality, if not of ‘disease’, is thus one of biological function. However, it is a little more complicated than this, because conception within 12 months is only an indirect way of assessing fertility. Some fertile men would be excluded merely because they have not been in a situation

of having unprotected heterosexual intercourse; that is, their fertility has not been put to the test. A second potential issue is that the relationship between a man’s degree of fertility and his likelihood of fathering a child is not completely direct, as it also depends on his female partner as well as on luck. A traditional criterion of regarding men who have fathered at least one child as ‘fertile’ has the drawback that even severely subfertile men sometimes father children; they just have more difficulty in doing so. It may be that the manual’s use of the additional 12-month criterion is better in this respect, as it restricts the set of men regarded as fertile, excluding the more subfertile, and is thus closer to the concept of normal biological function.

The next question is, what does this mean in the context of continuous measures? In the case of Addison’s or Cushing’s diseases, a discrete set of biological problems occurs below or above particular levels of blood cortisol, as is indicated by use of the word ‘disease’. Even in the case of plasma LDL cholesterol, despite the continuum of risk (absence of a threshold) across the range of concentrations, the risk is at least for something discrete, namely, ischemic heart disease. With semen quality, not only is each parameter a continuous measurement with no obvious biological threshold point, but the criterion variable of biological function—the length of time it takes to conceive (time to pregnancy)—is also continuous. It is therefore difficult to assess where ‘healthy’ ends and ‘abnormal’ begins.

An additional complication is that human male fertility may itself be abnormal in a sense. When compared with other mammals, humans have greatly inferior semen quality [3]. The distribution is also highly skewed to the right [4], compared with, for example, bulls and rams, which have a Gaussian distribution [5] similar to that seen with most biological variables. It therefore appears that it is statistically normal to be biologically abnormal. The situation in this respect resembles that of plasma LDL cholesterol; at least in populations in the developed world, a majority of adults have a level that puts them at increased risk of heart disease; so it is ‘normal’ to have an LDL cholesterol concentration that is abnormal in the sense that it puts one at risk.

Moreover, the human species is less fertile in a functional sense than other mammals. Human couples have a fecundability, defined as the probability of conceiving in any particular menstrual cycle, of 0.3 or lower [6], whereas in other mammals this value is far higher.

3 What a test result means for the individual

A diagnostic test can fulfil various functions. One is to inform the patient whether he is ‘normal’ or ‘abnormal’. As there is no clear threshold, but instead only degrees of

impairment of semen quality and reduced functional fertility, this would appear to have no definite meaning except in the case of azoospermia.

Second, a test can provide a diagnosis. Semen analysis can show the type of abnormality if one is present, for example, teratozoospermia, asthenozoospermia and/or oligozoospermia. Quantification is useful here because it allows identification of the most important abnormality or abnormalities affecting a man's semen quality. However, this analysis is limited, because it gives an understanding of neither the pathogenesis (mechanism) nor the initiating cause. The former is important because it indicates what treatment is likely to be effective, and the latter is needed for prevention. This test therefore only provides a first-stage diagnosis, rather like Addison's disease, which indicates a shortage of cortisol but not the underlying cause of the deficiency, which could be, for example, auto-immune disease or tuberculosis. A deeper diagnosis is needed, but unfortunately, in the present state of knowledge, this is rarely possible for abnormalities of spermatogenesis.

A third function is to provide a prognosis. It is here that the reference limits are probably most relevant. From the patient's viewpoint, what he and his partner are likely to want to know is whether they are going to be able to conceive naturally and how long this is likely to take. This information can play a vital role in management, as the semen test can show whether a medical intervention such as Intracytoplasmic sperm injection (ICSI) is likely to be needed, or possibly whether other courses of action should be considered, such as donor insemination or adoption.

The desired type of information is thus similar to the predictive value of a diagnostic test, as it would link a given test result to a general prognosis. The WHO manual gives related information in a different format: if a man has recently been able to father a child with a time to pregnancy of no more than 12 months, what is the probability that he will have a test with a certain degree of abnormality? For example, only 5% of such men will have a sperm concentration below 15 million per mL, and only 2.5% will have below 9 million per mL. This is more similar to the specificity of a screening test for a disease than to its predictive value, specificity being defined as the proportion of people without the disease who are correctly identified as such by the test. In this instance, 'people without the disease' corresponds to men who are fertile enough to fulfil the above criterion. This implies that 5% of such men have fewer than 15 million spermatozoa per ml of semen, but it does not specify what proportion of men with fewer than 15 million spermatozoa per ml are fertile in the sense of being capable of fathering a pregnancy within 12 months.

As is well known, the predictive value of a test is closely related to its specificity and sensitivity. The other

piece of information needed is the equivalent for semen quality of the prevalence of the disease. This would require an accurate statistical description of the distribution of the different semen parameters in each specific population—the general, unselected population, not one that is recruited as a result of success in conceiving (for example, through a maternity unit), nor indeed one that is derived from a fertility clinic. The sample needs to be large enough to enable the lower end of the distribution of each semen parameter to be adequately characterized. Given how difficult it is to study semen quality, especially with participation rates being low, it is unlikely that the perfect study can be done. However, existing studies of relatively representative population samples, for example, military conscripts, could be used instead of the more biased studies described above. This information could be combined with the information already provided in the WHO manual to generate predictive value-like information.

Until now, the best available evidence has been the prospective study by Bonde *et al.* [7], who found that below a threshold of about 40 million spermatozoa per mL the probability of conception was approximately proportional to the sperm concentration. Associations with motility and morphology were also shown. However, this is only one study, albeit a pioneering one, and it is never advisable to rely on a single study. Moreover, its participation rate was about 2% of potential volunteers, and its design excluded accidental pregnancies—although this would affect the lower end of the fertility/semen quality spectrum far less than it would affect the more fertile couples/men, as the latter are more likely to have accidental pregnancies [8]. It would be valuable to know whether the studies considered in the WHO report come to a similar conclusion when analysed in the suggested way.

4 Epidemiology: description and causal explanation

Another important role for accurate semen analysis is in epidemiological research. In principle, this can provide evidence of aetiological factors that could be targeted in prevention.

Epidemiological research on semen quality has mainly focused on trends, and in particular the concern that it may have declined [9]. In the course of this debate, appreciable spatial variation in semen quality was also found [10, 11]. In the case of both trends and spatial variation, the evidence has been based on an internal comparison: the values at different times or in different places were compared with each other, rather than with a 'normal' standard. It is possible, however, that this may have led to inferences being incomplete: reviews of the evidence, including those by the present author, have missed an important observation. The 'meta-analysis' by Swan *et al.*



[9] found that sperm concentration had apparently fallen more steeply in Europe than in North America, but what appears to have been missed is that at the end of the period studied (1996), the values had reached similar levels—the European values had started off considerably higher. This raises additional issues, for example, the unanswerable question of whether North American sperm concentrations were at some time similar to the observed higher European values, before any observations were made, but declined in an earlier period. Europe would then have been catching up with North America. The more general point here is that the purely internal comparisons that have characterized this literature may have led to an overemphasis on contrasts and deflected attention from the relationship of the observations to absolute levels.

From the viewpoint of causation and therefore prevention, any contrast can provide a clue. But from the functional point of view, in order to assess the practical effect, this information needs to be calibrated biologically, and prescriptive normality is thereby introduced.

However, in evaluating the question of trends in semen quality, a ‘meta-analysis’ of the kind undertaken by Swan *et al.* (among others) [9] is highly misleading in that the assumption is made that any particular data point represents the whole of Europe or North America. The idea that semen quality levels are equal across a continent and decline in a synchronized manner conflicts with virtually everything that is known about the way in which epidemiological variables behave, yet the issue of trends in semen quality continues to be discussed in such terms [12]. It is more useful to focus on the well-conducted single-centre studies, in which case the conclusion is that semen quality has indeed deteriorated in some places, but not in others [12].

5 Conclusion

Research on semen quality has advanced considerably since the earlier edition of the WHO manual. The authors have made great strides in incorporating this informa-

tion into a form that will be useful for laboratory practice throughout the world. In turn, this will help to improve the quality of semen analysis, both in practice and in research.

References

- 1 World Health Organization. WHO Laboratory Manual for the Examination and Processing of Human Semen, 5th ed. Geneva: World Health Organization; 2010.
- 2 Carlsen E, Giwercman A, Keiding N, Skakkebaek NE. Evidence for decreasing quality of semen during past 50 years. *BMJ* 1992; 305: 609–13.
- 3 França LR, Avelar GF, Almeida FF. Spermatogenesis and sperm transit through the epididymis in mammals with emphasis on pigs. *Theriogenology* 2005; 63: 300–18.
- 4 Handelsman DJ. Optimal power transformations for analysis of sperm concentration and other semen variables. *J Androl* 2002; 23: 629–34.
- 5 Setchell BP. Sperm counts in semen of farm animals 1932–1995. *Int J Androl* 1997; 30: 209–14.
- 6 Wood JW. Dynamics of Human Reproduction: Biology, Biometry, Demography (Foundations of Human Behavior). New York: Aldine de Gruyter; 1994. p285–9.
- 7 Bonde JP, Ernst E, Jensen TK, Hjolland NH, Kolstad H, *et al.* Relation between semen quality and fertility: a population-based study of 430 first-pregnancy planners. *Lancet* 1998; 352: 1172–7.
- 8 Key J, Best N, Joffe M, Jensen TK, Keiding N. Methodological issues in analyzing time trends in biologic fertility: protection bias. *Am J Epidemiol* 2009; 169: 285–93.
- 9 Swan SH, Elkin EP, Fenster L. The question of declining sperm density revisited: an analysis of 101 studies published 1934–1996. *Environ Health Perspect* 2000; 108: 961–6.
- 10 Fisch H, Goluboff ET, Olson JH, Feldshuh J, Broder SJ, *et al.* Semen analyses in 1,283 men from the United States over a 25-year period: no decline in quality. *Fertil Steril* 1996; 65: 1009–14.
- 11 Jørgensen N, Carlsen E, Nermoen I, Punab M, Suominen J, *et al.* East-West gradient in semen quality in the Nordic–Baltic area: a study of men from the general population in Denmark, Norway, Estonia and Finland. *Hum Reprod* 2002; 17: 2199–208.
- 12 Fisch H. Declining worldwide sperm counts: disproving a myth. *Urol Clin N Am* 2008; 35: 137–46.