

## Falling sperm quality: fact or fiction?

*Answering even simple questions is difficult*

See paper on p 19

Epidemiologists should be able to tell us whether sperm quality has changed over time, or at least whether the quantity has changed. In 1992 Carlsen and colleagues concluded that sperm concentration per unit volume had fallen by 40% over the previous 50 years.<sup>1</sup> This finding led to much speculation about the cause: oestrogens or pesticides in meat or water were the popular culprits (*Horizon*, "Assault on the Male: a Horizon Special," BBC, 1993 Oct 31). In an article in this week's journal Bromwich and colleagues argue that Carlsen *et al* applied the wrong form of analysis and that an artefact explains nearly all of the putative "fall" (p 19).<sup>2</sup>

Before we enter the modern debate it is worth being reminded that the hypothesis of falling sperm quality attracted most attention in the 1970s, but it was debated in the peaceful obscurity of the specialist journals.<sup>3</sup> Macleod and Wang temporarily silenced that debate; on the basis of a 10 year study of over 15 000 men they concluded that there had been no decline.<sup>4</sup> They considered all the large scale studies and compared them with their unique series from New York, where men had been analysed in the same laboratory for 30 years, with the same selection criteria and the same analytical methods being used.

This time around, Carlsen and colleagues' paper in the *BMJ* attracted the attention of the non-medical media. The question, nevertheless, raises difficult issues for the armchair epidemiologist, not least because of the problem of gaining access to the specialist literature. A typical postgraduate medical library in Britain is likely to have less than 10% of the original 61 papers quoted by Carlsen *et al*. Few reviewers have the time, energy, or resources to retrieve the full list. Those of us who have read most of the papers find several grounds for criticising Carlsen and colleagues' paper.

The authors began by searching MEDLINE and Index Medicus. This method itself produces its own publication selection bias, frequently fails to identify relevant articles in the scientific literature, and leaves out books and reports and other grey literature. The next problem was that of "patient" selection bias. Some men were examined before vasectomy; some were captured while their partners were attending antenatal clinics; some were volunteer donors participating in artificial insemination programmes (often medical students); and some were recruited as part of an occupational study in which a suspected hazard was being investigated. Others were recruited from infertility clinics but were included only if their partners subsequently became pregnant. Although Carlsen *et al* discussed selection bias, it would have been more

helpful if the groups had been kept separate and analysed separately.

Bromwich *et al* are right to point out the theoretical problem of selection bias resulting from the changing definition of "normal" sperm counts, which may have led to some men being excluded from some of the earlier studies, particularly those of healthy "normal" volunteers. Given the number of such studies, however, this is unlikely to have substantially biased the results of Carlsen and colleagues' analysis.

Carlsen *et al* included studies irrespective of their sample size; many of them were so small that they would not normally be considered to be admissible as evidence. One study was of seven men<sup>5</sup>; 11 studies were of fewer than 20 men and 29 studies of fewer than 50 men. Carlsen *et al* weighted these studies according to the logarithm of the sample size so that studies with small sample sizes were given greater weight than they deserved.

Although reference was made to the difference between the date of the study and the date of publication, no allowance was made when there were substantial discrepancies. For example, the largest study, with a sample size of 4435, was based on data collected between 11 and five years before publication.

Given the nature of the skewed distribution of the data, use of a geometric or logarithmic mean is clearly preferable to use of an arithmetic mean—a point that was made as early as 1979.<sup>4</sup> Bromwich *et al* make much of this skewness but from a theoretical perspective.<sup>2</sup> Given the plethora of research data, it is surprising that they chose to speculate rather than cite the literature.

Perhaps most puzzling of all is why Carlsen *et al* fitted a simple regression line to the data for sperm concentration and time. Many different ways of analysing time data exist, but rarely would a biological variable lend itself to simple regression. Why not try an exponential or logarithmic curve or even something cyclical? Indeed, several of the papers that Carlsen *et al* quoted applied cosine waves to several years of consecutive data with the conclusion that sperm concentration had an annual rhythm with peaks in the spring (March) and troughs in the autumn (September). Before deciding on which curve to apply it is useful to have some underlying hypothesis. Did an event in the past "cause" the decline or is some factor still operating? It is also useful to have a purpose—for example, to explain the past or predict the future.

Several lessons should be learnt from the *BMJ*'s saga of

vanishing sperm. Researchers need to exercise great care in how they collect data and describe their methods and results in journals. This is not a call for a return to simple empiricism, in which data are seen as being supreme. We would do well to remember that the word "datum" means "given" and provides us with information based on what we choose to collect and what we choose to leave out. There is inherent bias in how we define the problem in the first place. Once measurement begins, bias and error are fundamental.

The misapplication of increasingly sophisticated statistical tests (of which regression is one of the simpler) is becoming commonplace. More rigorous application of legitimate methods of analysis is required, particularly when time series are being analysed. When inferences are being drawn over time we deserve more than simple analysis: we need to follow McKeown's example and seek corroborating evidence from a wide range of disciplines.<sup>6</sup> Why not refer to the extensive data from veterinary research?

By the nature of their work epidemiologists erect hypotheses

and invite others to test them to destruction. They run a constant occupational risk, that of being mistaken. Editors have their part to play in protecting this small occupational group from doing themselves damage. Alternatively, editors could consider printing the names of referees alongside articles. This would do nothing to allay anxieties, but it would certainly spread the blame.

STEPHEN FARROW  
Professor

Health Research Centre,  
Middlesex University,  
Enfield EN3 4SF

- 1 Carlsen E, Giwercman A, Keiding N, Skakkebaek NE. Evidence for the decreasing quality of semen during the past 50 years. *BMJ* 1992;305:609-12.
- 2 Bromwich P, Cohen J, Stewart I, Walker A. Decline in sperm counts: an artefact of changed reference range of normal? *BMJ* 1994;309:19-22.
- 3 Nelson CMK, Bunge RG. Semen analysis: evidence for changing parameters of male fertility potential. *Fertil Steril* 1974;25:503-7.
- 4 Macleod J, Wang Y. Male fertility potential in terms of semen quality: a review of the past, a study of the present. *Fertil Steril* 1979;31:103-16.
- 5 Polakoski KL, Zahler WL, Paulson JD. Demonstration of proacrosin and quantitation of acrosin in ejaculated human spermatozoa. *Fertil Steril* 1977;28:668-70.
- 6 McKeown T. *The origins of human disease*. Oxford: Basil Blackwell, 1988.

## Triglyceride concentration and coronary heart disease

### *Not yet proved of value as a screening test*

See paper on p 11

As a cardiac risk factor triglycerides have occupied an uneasy position for a long time: perpetually under a cloud of suspicion but never proved to cause or even have a definite association with the risk of coronary heart disease. There are good grounds for suspicion—biological plausibility and epidemiological data suggest that triglycerides might promote coronary heart disease and other forms of cardiovascular disease—but the epidemiological and clinical evidence is inconsistent and often flawed. Even recent recommendations calling for the use of triglyceride concentration in screening are qualified by two admissions: that the evidence for a causal link remains frail and that trials have not established any clinical benefit from measuring triglyceride concentration.<sup>1-3</sup>

Why should a final judgment about the importance of triglycerides in cardiovascular disease remain so elusive? One cause is the variability of measurements of triglyceride concentrations—much greater than for other common lipids. This is partly due to the laboratory and partly to "true" short term fluctuations. This variability helps explain why statistical techniques underestimate the association between triglyceride concentration and the risk of coronary heart disease.<sup>4,5</sup> Clinicians can reduce the influence of physiological variability in triglyceride values by averaging repeated measurements—just as they do when measuring cholesterol concentration or blood pressure. But blood for measurement of triglyceride concentrations should be drawn after a 12 hour fast, making repeated measurement inconvenient and unappealing.

Perhaps the greatest perplexity about the importance of triglycerides derives from their relation to high density lipoprotein cholesterol concentration. This has a strong inverse correlation with triglyceride concentration, and studies that include measurement of high density lipoprotein cholesterol usually find no independent association between triglycerides and coronary heart disease.<sup>6</sup> Nevertheless, some recent studies have still failed to control for confounding by high density lipoprotein cholesterol. For example, the study by Lindenstrom and colleagues in this issue of the *BMJ* suggests an association between serum triglyceride concentration<sup>7</sup> and stroke (p 11); two other recent studies have reported, once again, that triglycerides and coronary heart

disease are related.<sup>8</sup> Yet all three investigations failed to adjust for high density lipoprotein cholesterol concentration, thereby contributing little to solving the question of whether triglycerides have additional predictive power beyond that of high density lipoprotein cholesterol.

Yet even if triglycerides do not have an independent association with coronary heart disease in multivariate analyses, might they be used to refine the estimation of coronary risk in some people? The Helsinki heart study found an interaction between triglycerides and cholesterol: most of the risk of coronary heart disease and nearly all the benefits of gemfibrozil treatment were confined to people with high concentrations of both triglyceride and cholesterol (particularly those with a ratio of low to high density lipoprotein of more than 5).<sup>10</sup> Similar findings emerged from the prospective cardiovascular Munster study (PROCAM), an observational study.<sup>3</sup> These results suggest that if triglyceride concentration is included with other lipid measurements, people might be identified who would benefit most from treatment. But the gaps and contradictions in published research make it clear that any such conclusion is speculative. Previous attempts to identify interactions among the lipid subclasses have had mixed results, either failing to detect important interactions or even finding inconsistent relations. For example, some studies found that triglycerides were associated with coronary heart disease only in subjects with low concentrations of cholesterol.<sup>11,12</sup>

The cholesterol-triglyceride interaction in the Helsinki heart study, while intriguing, was discovered incidentally and was not statistically significant.<sup>13</sup> Even if the finding is valid its implications for treatment are uncertain. Does the interaction imply that patients with raised cholesterol and triglyceride concentrations should be given more aggressive treatment to lower cholesterol concentration, or does it imply that they should be treated specifically with a drug (such as gemfibrozil) that also lowers triglyceride concentration? Confirmation in other clinical trials will be needed before we learn whether and how measurement of triglyceride values is useful in guiding lipid lowering treatment in patients with raised cholesterol concentrations.