Re: “Have Sperm Densities Declined? A Reanalysis of Global Trend Data”

Last year Swan et al. (1) published a reanalysis of data from 61 studies originally compiled and analyzed by Carlsen et al. (2). Just prior to the appearance of the Swan et al. article, we published a reanalysis in another journal (3).

Regional differences were considered in both reanalyses, but we examined only the effect of year in the final models (fertility status was also considered in the initial model), whereas they included several additional indicators culled from each study. However, while the results in the two papers for the U.S. studies were very similar (coefficients for the effect of year of -1.3 and -1.5 in their paper and ours, respectively), Swan et al. (1) reported a significant decline in sperm counts over time for Europe, whereas we found a nonsignificant decline. We doubted that this difference was due to the confounding with the additional covariates that they included, so we decided to explore.

We found the reason for the difference to be that Swan et al. only did a reanalysis of a subset of studies from the Carlsen et al. compilation (2). While dropping “two studies that included men who conceived only after an infertility workup” (1) seems justified on scientific grounds, dropping three non-English language studies was arbitrary, inappropriate, and led to the different results.

Two of the three non-English papers were from Europe and were written in Danish and German in decades before English dominated the scientific literature as it does today. These two studies, contrary to an assertion of Swan et al. (1) in their discussion, have sperm count values that are low relative to later studies done in Europe, so the slope is nonsignificant when they are included (in our analyses). Swan et al. also included an Australian study with the European ones; this would make sense if one had a hypothesis that there was a genetic or cultural cause of differences in sperm counts, but would be inappropriate if counts were hypothesized to vary with climate or environmental factors. Actually, the inclusion or exclusion of the Australian study influences the fits only trivially.

Figure 1 shows the linear regression fits for the data used in the Becker and Berhane (3) and the Swan et al. (1) studies and those of the flexible nonlinear models, fit as described in our original paper. Note the two data points in the analysis of Becker and Berhane but not in that of Swan et al. (blue circles) and the one Australian study included by Swan et al. but not by Becker and Berhane (diamond). The nonlinear models do indicate a decline in Europe after 1980, which is what Swan et al. documented. However there is actually an increase in sperm counts before 1980. The flexible fit based on the data used by Becker and Berhane shows significant nonlinearity [approximate F-value = 9.3 (2 degrees of freedom); p<0.01]. For the data after 1970, the flexible fits trace a pattern of a quadratic curve. The point estimates and statistical tests for the quadratic effect of year are shown in Table 1. The value for 1944 was excluded because it is clearly not part of the quadratic pattern.

In conclusion, the significant and very marked decline that Swan et al. (1) found for Europe was an artifact of their inappropriate sampling from the original studies. If the two non-English studies from 1944 and 1971 are included, there is no significant decline over the entire period. However, a significant nonlinear pattern is found, with an increase until about 1980 followed by a decrease. Such a significant quadratic pattern was not found in either the United States or in the other regions combined (not shown). We lack an explanation for the observed pattern in Europe, but since the Carlsen paper appeared, a number of other papers with more recent data from Europe have been published [see references in Becker and Berhane (3)].

There are several methodological morals to this story. First, single data points can have considerable influence in linear regression, particularly when the total number of sample points is small. Only very careful inspection of residuals from the linear regression over the entire period would allow one to spot the nonlinearity in this case. Second, it is inappropriate and parochial to only accept English-language studies in scientific meta-analyses.

Stan Becker
Department of Population Dynamics
Johns Hopkins University
Baltimore, Maryland

Kiros Berhane
Department of Preventive Medicine
University of Southern California
Los Angeles, California

References and Notes
1. Swan SH, Elkin EP, Fenster L. Have sperm densities declined? A reanalysis of global trend data. Environ Health Perspect 105:1228–1232 (1997).
2. Carlsen E, Gwernerman A, Keiding N, Skakkebaek E. Evidence for decreasing quality of semen during past 50 years. Br Med J 305:609–613 (1992).
3. Becker S, Berhane K. A meta-analysis of 61 sperm count studies revisited. Fertil Steril 37:1103–1108 (1997).

Response: Sperm Density Declines

Becker and Berhane take issue with the exclusion of three non-English language studies (1–3) from our reanalysis of the 61 studies on sperm density (4) that were included by Carlsen et al. (5). This objection raises two issues.

First, could we have used these studies in our analysis? We would argue that we could not. Unlike Becker and Berhane, whose own reanalysis (6) did not require any data other than what was published in Carlsen et al. (5), our multivariate analysis (4) required that we read the underlying studies. Otherwise, we would not have been able to abstract the detailed information on variables, such as age, abstinence time, and method of sample collection, that we included in our multivariate analysis. Moreover, not being fluent in German, Spanish, and Danish, we were not able to ascertain the eligibility of these studies.

Second, should we have used these three studies in our analysis even if we were able to
read them? We would argue that they should not have been included because these few studies are unlikely to represent all eligible non-English language studies published between 1938 and 1990. To determine the volume of non-English language articles in this field, we reviewed the Medline listing for 1989 publications obtained by Carlsen et al (5). We selected 1989 for this review because this was the last complete year included by Carlsen et al. and was therefore likely to have the least non-English publications during the study period if, as stated by Becker and Berhane, “English dominated the scientific literature” in recent decades. Of the 244 studies included, 58 (24%) were in languages other than English, with sixteen languages represented. Our Medline review suggested that Becker and Berhane’s perceived dominance of the scientific literature by the English language may be the “parochial” view, rather than ours. This review also suggested that it is unlikely that the three non-English language studies included by Carlsen et al. (all published before 1972) represented all eligible non-English studies; thus, there was no reason that these three alone should have been included. This application of our exclusionary criteria appears better justified than Becker and Berhane’s post hoc exclusion of the study by Varnek (1) simply because “It is clearly not part of the quadratic pattern.”

Finally, as noted in our paper, data from additional European studies suggested that sperm densities in Europe tended to be high early in the study period. Davidson (7), not included by Carlsen et al. (5) although eligible, reported a mean density of 143 × 10^6/ml in 1949. Further, the mean sperm density from five studies published in 1944–1962, which included 2,456 infertile European men (8–12) was 98.5 × 10^6/ml. It is reasonable to assume that sperm densities from fertile European men would have been at least as high and therefore would not support the quadratic model with low sperm counts in Europe prior to 1975, as proposed by Becker and Berhane in their letter [although not in their own analysis (6)].

Since its publication in 1992, the analysis by Carlsen et al. (5) has been widely discussed; our recent Medline search found it cited 231 times. It is unlikely that further discussion will resolve all remaining disagreements. Nevertheless, the conclusion of a mean decline in sperm density of about 1% per year is quite robust and is the same whether the analysis is based on the original 61 studies or only the 56 studies we included (4). Therefore, we suggest that at this point, efforts might best be spent elsewhere. Studies to rigorously estimate cross-sectional differences in semen quality are currently ongoing in several countries; these should provide reliable information about geographic variation in semen quality. Comparable data on temporal variation must await the results of prospective longitudinal studies.

Corrections and Clarifications

In the review by Hernandez-Boussard and Hainaut (A Specific Spectrum of p53 Mutations in Lung Cancer from Smokers: Review of Mutations Compiled in the IARC p53 Database) published in EHP [106:385–391 (1998)], the keys for Figures 6 and 7 were incorrect. The corrected tables appear below:

**Figure 6.** Comparison of the mutation spectra of p53 mutations in lung cancer of smokers and nonsmokers. Mutations are classified as transitions or transversions affecting G:C base pairs or A:T base pairs. Abbreviations: C to T, GC to AT transitions; G to T, GC to TA transversions; A to C, AT to CG transversions; A to G, AT to GC transitions; A to T, AT to TA transversions; Other, insertions, deletions, and complex mutations.

**Figure 7.** Comparison of the mutation spectra of p53 in lung cancer between male and female in the IARC p53 database. Mutations are classified as transitions or transversions affecting G:C base pairs or A:T base pairs. Abbreviations: C to T, GC to AT transitions; G to T, GC to TA transversions; A to C, AT to CG transversions; A to G, AT to GC transitions; A to T, AT to TA transversions; Other, insertions, deletions, and complex mutations.

**Shanna H. Swan**  
**Eric P. Elkin**  
**Laura Fenster**  
California Department of Health Services  
Berkeley CA

**REFERENCES AND NOTES**

1. Varnek J. Spermmaundersogser ved sterilitet: Med specielten hankl pa spermierens morfologi [Ph.D. dissertation]. Universitetsforlaget i Aarhus, Aarhus, Denmark, 1944.

2. Sturde H-C, Gloovania HJ, Bohm K. Vergleichende ejeculatuntersuchungen bei marnern aus stripen und ferteilen ehen. Arch Derm Forsch 241:426–433 (1971).

3. Robles G. Estudio del liquido spermatico. Arch Peruano Patol Clin 1:615–61 (1947).

4. Swan SH, Elkin EP, Fenster L. Have sperm densities declined? A reanalysis of global trend data. Environ Health Perspect 105:1228–1232 (1997).

5. Carlsen E, Hvermancen A, Keiding N, Skakkebaek NE. Evidence for decreasing quality of semen during past 50 years. Br Med J 305:609–613 (1992).

6. Becker S, Berhane K. A meta-analysis of 61 sperm count studies revisited. Fertil Steril 67:1103–1108 (1997).

7. Davidson HA. Male subfertility: interim report of 3,182 cases. Br Med J 2:1328–1332 (1949).

8. Hammer R. Studies on Impaired Fertility in Man: With Special Reference to the Male. Copenhagen: Einar Munkgaard, 1944.

9. Bostoffe J, Sarup J, Rebbe H. Has the fertility of Danish men declined through the years in terms of semen quality? A comparison of semen qualities between 1952 and 1972. Int J Fertil 26:91–95 (1983).

10. Oser S, Liedholm P, Rastam J. Depressed semen quality: a study over two decades. Arch Androl 12:113–116 (1984).

11. Bendoveld E, Gottlieb C, Bygdemann M, Eneroth P. Depressed semen quality in Swedish men from barren couples: a study over three decades. Arch Androl 26:189–194 (1989).

12. Bendoveld E. Semen quality in Norwegian men over a 20-year period. Int J Fertil 34:401–404 (1989).