

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

- Swan SH, Elkin EP, Fenster L. Have sperm densities declined? A reanalysis of global trend data. *Environ Health Perspect* 105:1228-1232 (1997).
- Carlsen E, Giwercman A, Keiding N, Skakkebaek E. Evidence for decreasing quality of semen during past 50 years. *Br Med J* 305:609-613 (1992).
- Becker S, Berhane K. A meta-analysis of 61 sperm count studies revisited. *Fertil Steril* 67:1103-1108 (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

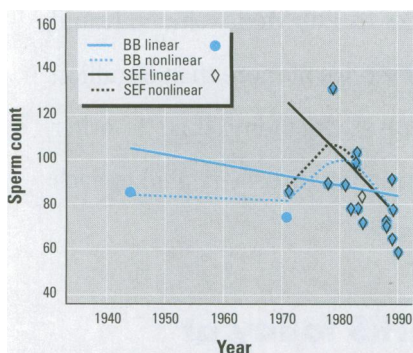


Figure 1. Regression fits for European data. Abbreviations: BB, Becker and Berhane (3); SEF, Swan et al. (1).

Table 1. Results of model fits to year of study^a for European sperm count studies subset from Carlsen et al. (2)

Model	Estimated β	t-Value	p-Value for overall model fit
Linear regression			
Subset by BB ($n = 17$)	-0.47	-1.02	0.3198
Subset by SEF ($n = 16$)	-2.59	-3.54	0.0033
Quadratic regression			
Subset by BB ($n = 16$) ^b			
Year	28.97	3.76	0.0018
Year squared	-0.29	-3.92	
Subset by SEF ($n = 16$) ^c			
Year	21.37	1.68	0.0032
Year squared	-0.22	-1.89	

^a1930 was subtracted from the year variable.

^bIncluding Australia but excluding two studies written in non-English languages.

^cExcluding the earliest study (see text).