Sperm output of healthy men in Australia: magnitude of bias due to self-selected volunteers

D.J. Handelsman
Andrology Unit, Royal Prince Alfred Hospital and Department of Medicine (D02), University of Sydney, Sydney NSW 2006, Australia

1To whom correspondence should be addressed

Controversial claims, based on a meta-analysis aggregating 61 heterogeneous observational studies, have been made that human sperm output has decreased by 50% over the last six decades and that this trend may be due to global pollution. If true, such effects should be evident in all areas of the globe; however, longitudinal studies within single centres in Europe and America have produced conflicting results and there are no reports from the southern hemisphere. We therefore reviewed semen analyses obtained from 1980–1995 from 689 healthy men volunteering for screening either as potential sperm donors for a donor insemination programme (n = 509) or to participate in five male contraception research studies (studies no. 1–5, n = 180). All were recruited through the Andrology Unit of the Royal Prince Alfred Hospital, Sydney, by the same doctors using standard methods of recruiting, screening and laboratory examination throughout the period 1980–1995. Recruitment was by advertising without regard to marital or fertility status except in two contraceptive efficacy studies (no. 1 and no. 3) where participants had to be in a stable relationship requiring contraception. Analysing the first semen sample individually or when grouped by year of ejaculation, there was no significant difference in sperm concentration over time or between years or according to year of birth. During the second half of this period, 180 consecutive volunteers were recruited by the same doctors and staff for five male contraception studies. The median sperm concentration for studies no. 1 (103 × 10⁶ ml) and no. 2 (142 × 10⁶ ml) were significantly (P < 0.05) higher than for studies no. 3–5 (84, 67 and 63 × 10⁶ ml, respectively) and for potential sperm donors (median 69 × 10⁶ ml). The inconsistency of these estimates illustrates the magnitude of bias (up to 100%) in sperm output that may occur in recruiting groups of self-referred volunteers within a single centre. This highlights the invalidity of extrapolating similar findings on sperm output of self-selected volunteers to the general male community or in using such study groups to characterize sperm output in supposedly ‘normal’ men.

Key words: sperm donors/sample bias/semen analysis

Introduction
Recently it has been claimed that sperm output (defined here as sperm production, encompassing both sperm concentration and total spermatozoa per ejaculate) of healthy men has fallen by as much as 50% over the last six decades (Carlsen et al., 1992). It was then further claimed that this alleged secular trend in human sperm output may be due to global pollution with synthetic chemicals with very weak oestrogenic (Sharpe and Skakkebaek, 1993) and/or anti-androgenic (Kelce et al., 1995) potency. This original claim, based on a meta-analysis of 61 heterogeneous observational studies of allegedly normal men (Carlsen et al., 1992), has been strongly disputed on various methodological grounds (Lerchl and Nieschlag, 1996) including notably inappropriate handling of skewed data (Bromwich et al., 1994) and statistical modelling (Olsen et al., 1995) as well as confounding due to changing inclusion criteria for ‘normal’ sperm output (Bromwich et al., 1994) and regional variation (Fisch and Goluboff, 1996). Subsequently, several investigators have reported analyses of sperm concentrations obtained serially within single centres over time (Wittmaack and Shapiro, 1992; Auger et al., 1995; Bujan et al., 1996; Fisch et al., 1996; Paulsen et al., 1996; van Waeleghem et al., 1996; Lemcke et al., 1997) which suggest that sperm output had increased (Fisch et al., 1996; Paulsen et al., 1996), decreased (Auger et al., 1995; Paulsen et al., 1996; van Waeleghem et al., 1996) or been unchanged (Wittmaack and Shapiro, 1992; Bujan et al., 1996; Irvine et al., 1996; Lemcke et al., 1997). Others have claimed regional variation between geographically separated groups of men studied in parallel (Bujan et al., 1996; Fisch and Goluboff, 1996; Fisch et al., 1996; Vierula et al., 1996; CECOS et al., 1997).

It is well known that, due to the highly personal and potentially embarrassing manner of collection, men are reluctant to provide semen samples unless currently concerned about their fertility. As a result, all available studies of men providing semen samples involve self-selected volunteers with various non-neutral motivations. While studies constituted in this fashion may have satisfactory internal scientific validity (based on within-study randomization, for example), aggregate data from such study populations cannot be legitimately extrapolated to their apparent source populations unless they originate or otherwise constitute a representative sample of that reference population. In standard survey sampling methodology, such inference would be based on random, probabilistic samples of the source population with or without stratification to ensure a representative sample at each stratum. With regard to semen analyses, however, the fertility of self-selected volunteers willing to provide semen samples is highly likely to be biased with respect to their source populations making extrapolation of aggregate study findings a priori scientifically dubious, if not invalid. Selection/participation bias based on samples of
infertile men is well recognized (Carlsen et al., 1992) and similar bias might apply to other selected subsets of apparently normal men based on sperm concentrations above a threshold (Irvine et al., 1996) or only married, once-fertile men (Auger et al., 1995; Bujan et al., 1996; CECOS et al., 1997). Similar reservations apply to longitudinal observation studies within single centres which, despite removing between-centre method variability, still consist of self-selected volunteers lacking a defined source population of which they constitute a representative sample.

If the secular trend in sperm output was genuine and due to global pollution, similar effects might be expected around the world but so far there have been no comparable data from the southern hemisphere. More importantly, despite the exclusive use of self-referred volunteers in such studies, there has been no analysis of the nature and influence of selection/participation bias on such observational studies of sperm concentration. In updating our analysis of men living in Sydney, screened as potential sperm donors (Handelsman et al., 1984) together with those men screened for participation in research studies (WHO, 1990, 1996; Handelsman et al., 1992, 1996) where semen analysis from healthy men was required, we have identified major participation bias in sperm output.

Materials and methods

Study subjects

Healthy men over the age of 18 years who volunteered as potential sperm donors or for research studies were recruited through the Andrology Unit, Royal Prince Alfred Hospital, Sydney, by the same doctors using standard methods of recruiting, screening and laboratory examination throughout the period 1980–1995. Volunteers were obtained by general poster-board advertising, word-of-mouth, and publicity via electronic and print media. All volunteers were screened by the same medical and nursing staff in the same location. A standardized medical history, physical examination and routine clinical chemistry (biochemical profile, full blood count) was obtained from each volunteer. Potential sperm donors represented a wide cross-section of society recruited without regard to marital or fertility status and students constituted only a small minority. After screening, only a minority of potential sperm donors were eventually selected to be used as sperm donors for the donor insemination programme. Volunteers for research studies were recruited by the same staff using similar means except that the inclusion criteria for two WHO studies (WHO, 1990, 1996) required that volunteers be in a stable relationship requiring contraception. None of these studies required men to have proven prior fertility.

Semen samples

Volunteers were routinely advised to have ≥2 days sexual abstinence prior to semen collection by masturbation. Semen samples were analysed by standard approved methods (WHO, 1980, 1987, 1992).

Data analysis

Only the first semen samples, individually or grouped by year of ejaculation, were analysed as some men provided only a single semen sample before exclusion or rejection. Use of multiple semen samples would have introduced a further bias according to whether the subjects were accepted either as sperm donors or as research study participants who (by design) generally have higher sperm output. Sperm concentrations are presented as median and interquartile range and analysed by suitable parametric methods after cube-root transformation. Box-and-whiskers plots illustrate the quartile ranges (including median) with 10th and 90th confidence limits and outliers as closed symbols. A two-tailed P value of 0.05 was considered statistically significant.

Results

We analysed semen samples from 689 men consisting of 509 healthy men (age 33.0 ± 0.4 years; 260/509 (51%) married; 226/509 (44%) with proven prior fertility) screened consecutively as potential sperm donors during the period 1980–1995 and 180 healthy men (age 30.1 ± 0.5 years; 142/175 (86%) married; 87/162 (54%) with proven prior fertility) volunteering to participate in five male contraception research studies during the period 1987–1994. All men were considered healthy on the basis of no significant abnormalities in their medical history, physical examination and routine clinical chemistry. All men included in this study provided a semen sample and no individuals were deleted or rejected from this analysis; however, only some of the men screened eventually participated in either programme and, for a few men who did not complete full screening, some clinical data (e.g. marital or fertility status) was lacking.

The sperm concentrations of 509 consecutive men screened as potential sperm donors are illustrated in Figures 1 and 2. Within each year 9–76 men were studied (median 18/year). The overall median sperm concentration among potential sperm donors from 1980–1995 was 69 × 10^6/ml. There was no significant relationship between donor sequence number and year of semen sample (Figure 1) nor between years of collection (Figure 2) or year of birth (data not shown). There was no significant relationship between annual medians and year of sample or year of birth. There was no significant relationship between semen volume and total sperm output (million/ejaculate) and the year of observation or the year of birth (data not shown).

During the latter half of this period, 180 volunteers were recruited by the same doctors and staff for five consecutive male contraception research studies (Figure 3). There was a significant (P < 0.05) difference between the medians of studies no. 1 (103 × 10^6 ml) and no. 2 (142 × 10^6 ml) and the other studies (no. 3, 84 × 10^6 ml; no. 4, 67 × 10^6 ml; no. 5, 63 × 10^6 ml) as well as with the overall median among potential sperm donors (69 × 10^6 ml). Recruits for studies no. 1 and no. 3 were required to be in a stable relationship for at least 1 year (WHO male contraceptive efficacy studies) but for the other three studies and for potential sperm donors there were no marital status requirements. None of the studies required evidence of prior paternity.

Discussion

Over the six decades since quantitative analysis of human semen was first reported (Macomber and Sanders, 1929), there has never been any study reported in which semen samples were obtained from a large sample of healthy men representative of the general male population. Such a study would have included men sampled randomly and proportionately from the general
community who could be classified by their reproductive performance over a convenient interval (e.g. over last year) as recently fertile, once fertile, infertile or fertility unknown/undefined. In the absence of such data defining a genuine reference range for human sperm output in any population, studies have been restricted to convenient collections of self-selected volunteers who provide semen samples for various reasons. As providing a semen sample by masturbation under standardized conditions still represents an intimate, potentially embarrassing situation for most men, study populations of men willing to provide semen samples are skewed by those who have current curiosity or concerns about their fertility. Such skewed volunteers constitute a biased sample of the population with regard to their perceived (in)fertility and this must be considered likely to influence semen characteristics. This would apply to samples restricted to only infertile men or to married, once-fertile men. The net effects of multiple positive and negative biases on sperm concentration is, however, difficult to estimate. Psychological and attitudinal profiling of men volunteering as potential sperm donors has demonstrated distinctive characteristics compared with the general male population (Handelsman et al., 1985). As a consequence it is striking that even now semen analysis remains a unique laboratory test in its lack of genuine reference ranges based on empirical sampling of a population-based, and therefore representative, sample of men.

The present study illustrates that the magnitude of volunteer participation bias arising from use of self-selected volunteers may be up to 100%, despite standardizing of staff as well as recruiting and laboratory methods within a single centre. It is notable that this large bias effect size within a single centre is comparable with the magnitude of the alleged effect size for decrease in sperm concentration over six decades, according to the Carlsen meta-analysis (Carlsen et al., 1992). One potential interpretation of our data is that sperm output of men living in Sydney has not changed over the last 15 years. We reject this interpretation as unjustified as our data arose from self-selected volunteers and not from a representative population sample. A sample such as ours may permit scientifically valid internal comparisons based on appropriate analyses within the sample (Handelsman et al., 1984) but generalization about the overall sample characteristics at any one time or serially over time to the geographical region where the study was performed is unjustified without reasonable evidence that the sample studied is representative of the regional population rather than a sample of convenience (Auger et al., 1995; Bujan et al., 1996; Fisch et al., 1996;
Irvine et al., 1996; Paulsen et al., 1996; van Waeleghem et al., 1996; Vierula et al., 1996; CECOS et al., 1997). The observed results with self-selected volunteers are likely to reflect the nature and efficacy of our volunteer recruitment programmes rather than the general male population in the Sydney region. For the same methodological reason it appears unjustified to make analogous interpretations of temporal changes in overall semen characteristics based on longitudinal observations (of different subjects at each time point) from other single centres (Auger, 1995; Bujan et al., 1996; Fisch et al., 1996; Irvine et al., 1996; Paulsen et al., 1996; van Waeleghem et al., 1996; Vierula et al., 1996; CECOS et al., 1997) where the overall study sample characteristics are extrapolated to the general male population but where equally the study sample is unlikely to represent a valid statistical sample of the source population. In addition to volunteer participation bias, other forms of selection bias are present explicitly in these other studies. For example, one study of men providing semen samples for laboratory sperm biology research only included men who had ‘normal’ sperm, defined as sperm concentration exceeding $20 \times 10^6$/ml (Irvine et al., 1996) and others only included married, once fertile men (Auger et al., 1995; Bujan et al., 1996; CECOS et al., 1997). Restrictions such as these inevitably create bias if the aggregate data are used for extrapolation to characterize the general male population whereas it may not invalidate internally randomized comparisons. Furthermore a global environmental effect that reduces human sperm output is not consistent with regional discrepancies in whether sperm output is changing over the same time period between different cities in France (Auger et al., 1995; Bujan et al., 1996; CECOS et al., 1997) and the USA (Fisch et al., 1996; Paulsen et al., 1996) as well as between continents (Fisch and Goluboff, 1996). Under these circumstances it is difficult to accept claims that sperm concentrations are decreasing due to lack of reliable evidence. Finally, corollary evidence showing unchanged or increasing sperm output of rams, boars and bulls over the last six decades (Setchell, 1997) is more consistent with the absence of any global biological effect unless it is restricted to humans.

Independent evidence whether self-selected volunteers are representative of the general male community in this issue could arise from studies that characterize the types of volunteers providing source data. The small amount of quantitative research available to characterize volunteers for either sperm donation (Handelsman et al., 1985) or participation in research studies (Farkas et al., 1978; Strassberg and Lowe, 1995) indicates that they differ systematically from the general population in psychological characteristics including ones such as sexuality and risk-taking behaviours which could ultimately influence semen analysis results.

Observational studies (including meta-analysis) are subject to numerous biases which are difficult to estimate and control (Sackett, 1979). In the absence of prospective randomization, with its unique property of balancing the unknown as well as known predictors of outcome, the identification of unrecognized bias is an open-ended task. Two other important biases relevant to this topic include the selection bias in using aggregate data from infertility clinics to characterize sperm production in a region and publication bias. The former has been repeatedly used as a major source of semen analysis data (MacLeod and Wang, 1979; Adamopoulos et al., 1996; Vierula et al., 1996). Such data, however, are manifestly biased according to the proportion of cases with infertility due mainly or partly to male factors in whom sperm output is lower. Quite apart from any changes in spermatogenesis, this population proportion may vary over time according to numerous extraneous factors including changes in diagnostic, referral and even insurance reimbursement practice. For this reason alone even very large aggregates of data from infertility clinics or laboratories shed little useful light on whether human sperm output has changed in the general male community. Publication bias refers to the fact that negative data – in this case semen analysis data showing no apparent change with time – is less likely to traverse each stage of the life history of a scientific publication. The negative influence applies to the data being less likely to be analysed, or to the manuscript being written, submitted or accepted for publication and, if accepted, it is likely to appear in less prominent journals. Such negative publication bias has occurred in this area of controversy.

We concluded that the inconsistency of the estimates observed in this study illustrates the magnitude of the bias in sperm output that may arise from recruiting groups of self-referred volunteers even within a single centre using standardized recruiting and laboratory methods. This highlights the invalidity of extrapolating such findings on sperm output of self-selected volunteers to the general male community from which the volunteers originated. Such considerations dictate the need for major reservations about the external validity of studies using self-selected volunteers in attempting to characterize sperm output in supposedly ‘normal’ men who might be claimed to represent the populations from which they originated.

Acknowledgements

The author is grateful to Drs A. Conway, S. Wishart, L. Turner, L. Boylan, S. McGhee and C. Howe for their outstanding professionalism in clinical and technical skills over many years in conducting the sperm donor and research programmes.

References


*Received on March 14, 1997; accepted on September 9, 1997*