Does the potential for selection bias in semen quality studies depend on study design? Experience from a study conducted within an infertility clinic

Russ Hauser1,2,3, Linda Godfrey-Bailey1 and Zuying Chen2

1Department of Environmental Health, Occupational Health Program, Harvard School of Public Health, Boston, MA 02115 and 2Vincent Memorial Obstetrics & Gynecology Service, Andrology Laboratory and In Vitro Fertilization Unit, Massachusetts General Hospital, Boston, MA 02114, USA
3To whom correspondence should be addressed at: Occupational Health Program, Department of Environmental Health, Harvard School of Public Health, Building 1, room 1405, 665 Huntington Avenue, Boston, MA 02115, USA. E-mail: rhauser@hohp.harvard.edu

BACKGROUND: The low participation rates in human semen quality studies raises concern for the potential of differential participation based on semen quality (or a surrogate). To explore the potential for differential participation, we compared semen analysis results from study subjects with those of non-study subjects. METHODS: We obtained semen analysis results from 235 study subjects and retrospectively obtained results from a subset of 235 infertility clinic patients that were not study subjects but met the same eligibility criteria. The study was conducted at the Massachusetts General Hospital Infertility Clinic. All semen samples (study subjects and non-study subjects) were analysed for sperm concentration and motility by computer-aided semen analysis (CASA), and morphology was assessed using strict criteria. Semen analysis parameters for the non-study subjects were compared with the semen analysis results from study subjects. RESULTS: For all semen characteristics (sperm concentration, total sperm count, sperm motility and morphology), there were only marginal (non-significant) differences between study subjects and non-study subjects. CONCLUSIONS: Among men from an infertility clinic, we found no strong evidence of differential participation based on semen quality. This is reassuring since the potential for selection bias is of concern in semen quality studies. However, the potential for selection bias in other study designs remains unclear.

Key words: infertility/selection bias/semen quality

Introduction
A recurring difficulty in conducting studies on human semen quality is the very low participation rates among eligible subjects (Cohn et al., 2002; Tielemans et al., 2002). This has heightened concern for the potential of differential participation of subjects based on semen quality. Selection bias may occur if participation is differentially based on semen quality, or a surrogate thereof, such as history of infertility, and also based on the exposure of interest. The exposure of interest can be the level of an environmental chemical, a lifestyle factor or even a birth history variable such as birth weight. It is important to emphasize that for selection bias to occur, participation must be based on both semen quality (or a surrogate) and the exposure of interest (or surrogate).

To explore the potential for selection bias, one would ideally collect data on semen quality from both study subjects and those that refuse to participate. Since this is usually not possible, we generally obtain data on characteristics that are considered surrogates for semen quality and compare participants with non-participants. Characteristics that are frequently measured include age, smoking history, educational level, and medical and reproductive history (Eustache et al., 2004; Muller et al., 2004).

Several studies have explored whether participants in semen quality studies differ from non-participants. Since, by definition, semen analysis results on non-participants are generally not available, innovative designs have been used to explore the potential for differential participation. In a French study on male partners of pregnant women, information was obtained from a group of men agreeing to collect a semen sample and complete a questionnaire (group A), a group of men completing only the questionnaire (group B) and from men refusing to participate altogether (group C) (Eustache et al., 2004). The participation rate for men in group A was only 15.8%. There were no differences in the ages and socio-professional status across the three groups. However, a higher proportion of men giving a semen sample had a high educational level.
Time to pregnancy was also not significantly different across the groups, although there was a higher proportion of couples in group A taking longer than 12 months to conceive than in group B. Pregnancy outcomes were also similar in the three groups. Interestingly, although not statistically significant, the study found a higher proportion of group A volunteers with a history of urogenital diseases (such as cryptorchidism, infection and varicocele). However, semen characteristics were not significantly different between men having or not having a history of urogenital disease. This suggests that participation may not relate to semen characteristics. Overall, the authors concluded that despite the low participation rates, there did not appear to be major differences in characteristics between the three groups of subjects. There were minor differences in education and urogenital history, therefore data on these characteristics need to be collected and adjusted for in statistical analyses as potential confounders.

In another recently published French study, Muller and co-workers (2004) explored differential participation among three groups of men that were partners of pregnant women. Group 1 men completed a questionnaire and produced a semen sample (13% participation rate), group 2 men completed only the questionnaire and group 3 men refused the questionnaire and did not produce a semen sample. Smokers and poorly educated men were more likely to refuse to participate than non-smokers and highly educated men. Semen providers were more likely to have reported unfavourable pregnancy outcomes (odds ratio 1.68, 95% confidence interval 1.14–2.49) as compared with men only completing the questionnaire. However, time to pregnancy (TTP) was similar across groups, suggesting that fecundity was similar across groups.

Muller and co-workers (2004) found that there were differences in socio-demographic characteristics by level of participation, and therefore men that agree to participate may not represent the population from which they originate. Although they found participation differed based on education and reproductive history, because they did not have semen analysis results on men in groups 2 and 3 they were unable to determine whether participation was also related to the other axis (i.e. semen quality). Based on differences of unfavourable pregnancy outcomes and maternal smoking among the men in group 1, they hypothesized that semen quality may be lower for group 1 men. The TTP results, however, did not support this, although TTP may not have been sensitive enough to detect changes in semen quality.

In a subset of subjects from the Child Health and Development Studies (CHDS), Cohn and co-workers (2002) explored potential participation bias. Between 1960 and 1963, while pregnant, the mothers of the sons targeted for study recruitment were enrolled in the CHDS. Sons were 36–39 years of age at the time of attempted recruitment by Cohn et al. The projected participation rate was 53% for semen donation. Sons who agreed to participate after the initial mailing (respondents) and sons who were subsequently traced and recruited after the initial mailing (initial non-respondents) differed in semen parameters, including sperm concentration and the percentage of sperm with normal morphology. In addition, race, marital history and fertility history differed between respondents to the initial mailing and recruited non-respondents. The design of the CHDS study was population based, and therefore participation patterns and motivations may differ from studies conducted within infertility clinics (our study) and studies conducted among male partners of pregnant women (Eustache et al., 2004; Muller et al., 2004).

Another design frequently used in semen quality studies is to conduct the study within an occupational cohort. These studies are also difficult to conduct due to traditionally low participation rates. Larsen and co-workers (1998) analysed data from three Danish occupational sperm studies and evaluated how age and subfertility were related to participation and providing a semen sample. Participation rates varied from a high of 62 and 74% among the exposed and comparison group, respectively, in the greenhouse workers study, to 28% among traditional farmers and 49% among organic farmers in the farmers’ study. In the metalworkers and electricians study, participation rates were 38% in the exposed group and 35% in the unexposed group. In these studies, age and subfertility were found to be strong predictors of participation. Participation was higher among men <40 years of age and among men who experienced an infertile period. The effect of infertility was modified by occupational exposure status, producing a tendency for participation to be based on both subfertility, a surrogate for semen quality, and occupational exposure. This may introduce selection bias in studies conducted within these occupational cohorts.

Several studies explored whether hormone levels, specifically testosterone, may be useful to distinguish between participants and non-participants and as a surrogate for semen quality (Mullard et al., 1999; Andersen et al., 2000). Participation rates in the Mullard and Andersen studies were 35 and 18%, respectively. Andersen and co-workers compared serum testosterone levels in semen donors and non-donors among military recruits in a study in Denmark. They used the same cut-off as Mullard et al. (testosterone level of 400 ng/dl) and found similar mean testosterone levels among donors and non-donors. Four percent of semen donors had testosterone levels <400 ng/dl compared with 7% of non-donors. Mullard also found a lower percentage of men with <400 ng/dl among semen donors (0%) compared with non-donors (7%).

Because it is difficult, and generally not possible, to measure semen quality among non-participants, the published studies described above relied on measuring factors that may be related to semen quality, such as difficulty having a child or medical reproductive history. Some studies relied on comparisons of demographic or lifestyle characteristics between participants and non-participants. Demographic characteristics included age and educational level, and lifestyle factors included smoking history.

Because our study was conducted in an infertility clinic, we had the unique opportunity to explore directly whether semen analysis results of study subjects differed from the infertility clinic population from which the study subjects were recruited. The infertility clinic population includes three
sets of subjects: (i) subjects that were approached and agreed to participate in the study (study subjects); (ii) subjects that were approached and refused to participate (refusals); and (iii) subjects not approached because of scheduling conflicts or other reasons. Those not approached represent a heterogeneous group consisting of those who would have refused if given the opportunity and those who would have participated if given the opportunity. Despite our ability to compare semen quality results of study subjects with the clinic population from which they arose, we were not able to identify a pure group of non-participants (i.e. subset number ii as described above). We were therefore limited to a comparison of study subjects and the population of infertility clinic patients that included both refusals and subjects not approached.

Materials and methods

Because all men visiting the Andrology Laboratory provide a semen sample as part of their clinical care at Massachusetts General Hospital (MGH), we had access to semen analysis results from both men who agreed to participate in our study and all other men presenting to the infertility clinic. Men were eligible for study recruitment if they met the following criteria: (i) between the ages of 18 and 55 years; (ii) referred by a study-approved physician; (iii) did not have a vasectomy; and (iv) were not receiving hormonal treatments.

Men who agreed to participate were consented by the research nurse and asked to provide a urine sample, to allow us to draw a blood sample and to allow us to have access to the results of their semen analysis from their clinic visit. Approximately 65% of men approached agreed to participate fully. Fewer than 5% agreed to partial participation, which included not agreeing to provide a blood sample. We determined the participation rate of the study by tracking recruitment contacts that the research nurse had with men in the infertility clinic. Men who verbally declined to participate in the study were noted as declining. An additional category of men who were not ready to make a decision about study participation or who had not been approached by the nurse were not included in the total participation rate. Men who had not decided about the study after a first meeting with the research nurse were eligible to be approached again about the study at subsequent clinic visits. If this second contact led to either participation or declining to participate, the participation rate was adjusted accordingly.

To explore the potential for differential participation based on semen analysis results, we retrospectively and anonymously obtained semen analysis results from a subset of infertility clinic patients that were not study subjects. The MGH and Harvard School of Public Health (HSPH) Human Subject Committees approved this study. Since semen analysis results were collected anonymously, we were unable to confirm abstinence time.

All semen samples (study subjects and non-study subjects) were analysed by Z.C. (or her trained technician) for sperm concentration and motion parameters by computer-aided semen analysis (CASA; Hamilton-Thorn Version 10HTM-IVOS, Beverly, MA). Setting parameters and the definition of measured sperm motion parameters for CASA were established by Hamilton-Thorn Company [frames acquired, 30; frame rate, 60 Hz; straightness (STR) threshold, 80.0%; medium VAP cut-off, 25.0 μm/s; and the duration of the tracking time, 0.38 s]. To measure both sperm concentration and motility, aliquots of semen samples (5 μl) were placed into a pre-warmed (37°C) Makler counting chamber (Sefi-Medical Instruments, Haifa, Israel). A minimum of 200 sperm from at least four different fields were analysed from each specimen. The percentage of motile sperm was defined as WHO grade ‘a’ sperm (rapidly progressive with a velocity ≥25 μm/s at 37°C) plus ‘b’ grade sperm (slow/sluggish progressive with a velocity ≥5 μm/s but <25 μm/s).

Using the ‘feathering’ method (World Health Organization, 1999), at least two slides were made for each fresh semen sample. The resulting thin smear was allowed to air dry for 1 h before staining with a Diff-Quik staining kit (Dade Behring AG, Düdingen, Switzerland). Morphological assessment was performed with a Nikon microscope using an oil immersion 100x objective (Nikon Company, Tokyo, Japan). Spermatozoa were assessed and scored as normal or abnormal using the strict criteria of Kruger et al. (1988). Results were expressed as the percentage of normal spermatozoa.

Because covariate information was not available for the non-study subjects, all results are unadjusted. Because distributions of sperm concentration and total sperm count were skewed, Wilcoxon signed rank tests were used to compare these semen characteristics across study and non-study subjects. t-tests were performed to compare sperm motility and morphology across study and non-study subjects. χ² tests were used to compare the percentage of study and non-study subjects who had semen characteristics below the reference range. All statistical analyses were conducted using Statistical Analysis Software (SAS), version 8.2 (SAS Institute Inc., Cary, NC).
Results

Semen analysis parameters for 235 non-study subjects were compared with the semen analysis results from 235 study subjects (Table I). For all semen characteristics (sperm concentration, total sperm count, sperm motility and morphology), there were only marginal non-significant differences between study subjects and non-study subjects. For instance, median (25th and 75th percentiles) sperm concentrations (×10⁶/ml) were 68.8 (28.0, 148.9) for study subjects and 72.8 (36.4, 134.6) for non-study subjects. The mean (SD) sperm motility (percentage motile) was 46.8 (25.6) for study subjects and 45.6 (24.6) for non-study subjects. The mean (SD) morphology (percentage normal) was 6.7 (4.5) for study subjects and 6.6 (4.2) for non-study subjects.

The percentage of subjects with below reference semen parameters was nearly identical for the study subjects and non-study subjects. This suggests that in analyses using categorical semen parameters, there would be no difference between study subjects and non-study subjects. With selection bias, the 65% participation rate in the present study is only moderately high and therefore does not ensure that selection bias is not a threat to the validity of the study. Because participation rates can only provide limited insight into the likelihood of selection bias, it is important to collect data that allow one to assess empirically whether there is differential participation based on both semen quality and the exposure of interest. In the present study, we were able to collect semen quality data on participants and non-participants. However, the present study was limited since it lacked exposure information on non-participants, as well as socio-demographic information on non-participants.

As described earlier, several studies have explored whether participants in semen quality studies differ from non-participants (Larsen et al., 1998; Mullard et al., 1999; Andersen et al., 2000; Cohn et al., 2002; Eustache et al., 2004; Muller et al., 2004). These studies compared demographic, medical or lifestyle factors of participants with those of non-participants.

Results across these studies were not entirely consistent; some studies found evidence of differential participation based on demographic or lifestyle characteristics, medical history and semen quality (Larsen et al., 1998; Cohn et al., 2002; Muller et al., 2004), while others did not (Eustache et al., 2004; this study). Upon closer inspection, the design of the study may be predictive of whether participation was likely to be differentially related to semen quality or the characteristics of the subjects.

Studies among men that were partners of pregnant women, referred to as ‘fertile male studies’ (Eustache et al., 2004; Muller et al., 2004), and studies among men who were partners of an infertile couple, referred to as ‘infertile male studies’ (this study), generally had lower potentials for differential participation. In these study designs, the men targeted for recruitment represent homogenous groups in terms of their fertility and also in terms of their concern with infertility.

### Table 1. Semen characteristics for study subjects (n = 235) and non-study subjects (n = 235)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Study subjects</th>
<th>Non-study subjects</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sperm concentration</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD), ×10⁶/ml</td>
<td>98.0 (100.3)</td>
<td>96.4 (95.8)</td>
<td>0.90‡</td>
</tr>
<tr>
<td>Median (25th, 75th), ×10⁶/ml</td>
<td>68.8 (28.0, 148.9)</td>
<td>72.8 (36.4, 134.6)</td>
<td></td>
</tr>
<tr>
<td>% men with &lt;20 × 10⁶/ml</td>
<td>19%</td>
<td>19%</td>
<td>1.0‡</td>
</tr>
<tr>
<td>Total sperm count</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD), ×10⁶</td>
<td>279.1 (269.4)</td>
<td>275.4 (278.1)</td>
<td>0.75‡</td>
</tr>
<tr>
<td>Median (25th, 75th), ×10⁶</td>
<td>198.0 (83.2, 378.9)</td>
<td>199.1 (78.4, 379.8)</td>
<td></td>
</tr>
<tr>
<td>% men with &lt;40 × 10⁶</td>
<td>16%</td>
<td>16%</td>
<td>1.0‡</td>
</tr>
<tr>
<td>Sperm motility</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD), % motile</td>
<td>46.8 (25.6)</td>
<td>45.6 (24.6)</td>
<td>0.60‡</td>
</tr>
<tr>
<td>Median (25th, 75th), % motile</td>
<td>50.0 (24.0, 69.0)</td>
<td>48.0 (26.0, 66.0)</td>
<td></td>
</tr>
<tr>
<td>% men with &lt;50% motile</td>
<td>50%</td>
<td>51%</td>
<td>0.78‡</td>
</tr>
<tr>
<td>Sperm morphology</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD), % normal</td>
<td>6.7 (4.5)</td>
<td>6.6 (4.2)</td>
<td>0.72‡</td>
</tr>
<tr>
<td>Median (25th, 75th), % normal</td>
<td>6.0 (3.0, 9.0)</td>
<td>6.0 (4.0, 9.0)</td>
<td></td>
</tr>
<tr>
<td>% men with &lt;4% normal</td>
<td>26%</td>
<td>24%</td>
<td>0.67‡</td>
</tr>
</tbody>
</table>

% men = percentage of men below the WHO reference values for sperm concentration <20 × 10⁶ sperm/ml, total sperm count <40 × 10⁶ sperm/ml, sperm motility <50% motile sperm and strict criteria for morphology <4% normal morphology.

aWilcoxon signed rank tests were performed for sperm concentration and total sperm count.

b-t-tests were performed for sperm motility and morphology.

χ² tests were performed to compare the percentage of men with below reference value semen parameters.
Concern with infertility is likely to be low for men in ‘fertile male studies’ and high for men in ‘infertile male studies’. Because a man’s concern with infertility is similar within the targeted group of ‘fertile men’ (or ‘infertile men’), this may minimize the likelihood of differential participation. For example, in the present study, because all men were visiting an infertility clinic as a partner in an infertile couple, they all had similar concerns with infertility. Thus participation was likely to be independent of the man’s concern with infertility. Because concern with infertility may be related to semen quality, participation would probably be independent of the man’s semen quality. We recognize that the level of concern will vary even among men visiting an infertility clinic; however, on average they are expected to have higher levels of concern than men not visiting an infertility clinic or men who have fathered a child.

Studies among men from the general population (Cohn et al., 2002) or from occupational groups (Larsen et al., 1998) generally had higher potentials for differential participation related to semen quality or demographic and lifestyle factors. This may be because these designs targeted a more heterogeneous group of men in terms of their fertility. In these designs, among the men targeted for recruitment, there may be large differences in the man’s concern with infertility since some men are fertile (low concern) and others are infertile (high concern). Therefore, there may be differential participation related to concerns with infertility. If concerns with infertility are related to poor reproductive history or poor semen quality, this may lead to the potential for bias.

In conclusion, in the present study within an infertility clinic population, there was little evidence of differential participation based on semen quality. However, it is unclear whether there may be differential participation in studies using other designs. When assessing the likelihood and magnitude of bias related to differential participation in semen quality studies, it is important to consider how the design impacts the potential for bias. Two studies may have equally low rates of participation but, because their design differs, the likelihood of bias may also differ. In addition to concerns with how the study design impacts the potential for differential participation and bias, the design also has implications for the external validity (generalizability) of the results. In studies on infertile or fertile men, generalizability may be more restricted than in studies on men from the general population. Although a detailed discussion of generalizability is beyond the scope of this study, it is important to consider both the internal validity (bias) and external validity (generalizability) when designing human semen quality studies. Further work in these areas will help identify more optimal designs to study predictors of semen quality.

Acknowledgements

The authors thank the staff of the Vincent Memorial Obstetrics and Gynecology Service Andrology Laboratory and In Vitro Fertilization Unit at Massachusetts General Hospital, Computer Programmers, Ms Lucille Potthier and Ms Janna Frelich, and Research Assistants Ms Ana Trisini and Mr Ramace Dadd. We thank Dr John Meeker for his thoughtful comments on the draft of the manuscript. This study was supported by grants ES09718 and ES00002 from the National Institute of Environmental Health Sciences, NIH.

References


Submitted on February 1, 2005; resubmitted on April 19, 2005; accepted on April 22, 2005