Few empirical data exist on the characteristics of subjects who provide semen specimens in epidemiologic studies. The objective of this investigation was to determine participation rates and potential biases in a contemporary study of human semen quality. Subjects (n = 268) are a subset of the Child Health and Development Studies. Their mothers enrolled between 1960 and 1963 during pregnancy. Archived prenatal serum samples, prenatal and birth records, placental examinations, and follow-up for growth and development through adulthood are available. Sons were aged 36–39 years at the time of this study. Respondents to the initial mailing and nonrespondents, who were subsequently traced and recruited, differed in semen parameters, including sperm concentration (78.3 × 10^6/ml for respondents vs. 37.2; p = 0.003) and the percentage of normal morphology according to the 1987 criteria of the World Health Organization (58.7% for respondents vs. 53.3%; p = 0.04). The authors conclude that researchers designing population-based studies of semen parameters should expect nonrepresentative samples. Adaptation of the design to anticipate and mitigate bias and to maximize efficiency can yield scientifically sound information. Recommendations for study designs are discussed. Am J Epidemiol 2002;155:664–71.

Valid interpretation of studies of environmental exposures on semen parameters depends on characterizing potential response bias in studies where subjects donate semen. However, few empirical data exist on the characteristics of subjects who provide semen specimens. Response rates for prior semen studies suggest that response bias may be of concern. As reviewed by Eskenazi et al. (1), studies of the semen quality of workers report response rates below 50 percent. A recent, large-scale investigation of semen quality in the general population of several European cities reported response rates of 43 percent in Copenhagen, Denmark, 19 percent in Turku, Finland, and 15 percent in Paris, France (2). A Danish study of first pregnancy planners reported an average participation rate of 16 percent (3). In a study of an unselected population of men aged 18 years based on conscripts in Denmark, 16–19 percent of the eligible men agreed to physical examination, blood sample collection, and semen specimen donation (4). In contrast, when the protocol consisted of only a blood donation, 79 percent agreed to participate (4).

Here we report on the feasibility of collecting semen specimens and time-to-pregnancy information in a unique cohort, the Child Health and Development Studies (CHDS) (5). More than 40 years of follow-up have already been completed for CHDS offspring, beginning with the prenatal period. Interviews and clinical observations during pregnancy and delivery are available, as are archived serum samples from each trimester of pregnancy.

The objective of this study was to determine participation rates and potential biases in a contemporary study of semen quality in a cohort of CHDS offspring. Our findings have implications for the design and interpretation of epidemiologic studies of human semen quality.

MATERIALS AND METHODS

Study population

The study population comprised men who were members of the CHDS cohort (5). Their mothers enrolled in the CHDS during pregnancy (1960–1963). Offspring have been followed for a variety of outcomes from before birth to early adulthood. Subjects were members of the Kaiser Permanente
Health Plan who resided near Oakland, California. Nearly 100 percent of all the pregnancies during the target years were identified. Archived prenatal serum samples are available to assess prenatal exposures to endogenous and exogenous factors that may influence the reproductive function in sons, including endocrine disrupters.

We selected all CHDS sons, born between 1960 and 1963, who had been participants in a follow-up study when they were aged 15–17 years and who resided within 50 miles (80.47 km) of Santa Clara, California (n = 268). This location was chosen because of proximity to an andrology clinic, at Kaiser Permanente, Santa Clara, which was a site for semen evaluation in clinical trials conducted by the National Cooperative Reproductive Medicine Network (6). A trained technician and equipment for preparing videotapes of sperm motility were available at this site. Men were aged 36–39 years at the time of this data collection.

Design

All subjects in the study population received a letter inviting them to participate in the study and asking that they return a postcard providing a current phone number and a convenient time to receive a call from study recruiters. The most recent Department of Motor Vehicles’ linkage (September 1998) was used to obtain an address for each study subject. One interviewer contacted all study subjects. After subjects agreed to the study, their telephone number was forwarded to the computer-assisted telephone interview unit of the Public Health Institute to schedule the interview. Appointments for semen specimen collection were scheduled at the same time. There were separate consent procedures for the interview (over the telephone), for providing a blood sample, and for the semen donation (both written consent). All subjects were offered a standard incentive of $100 to complete the interview and to provide a blood sample and two semen samples. A subject could agree to the complete protocol or only part of the protocol. The institutional review boards of the Public Health Institute and Kaiser Permanente approved the protocol.

We selected 103 of 268 CHDS sons with the shortest travel time to the clinic for intensive recruitment (group I). The geographic restriction resulted from funding limitations of this pilot study. It was not possible to set up clinical sites to receive semen donations at multiple, convenient sites, as might be done for a full-scale study. Subjects in group I were traced and recruited actively if they failed to return the postcard.

The remaining 165 CHDS sons were assigned to a second study group (group II). Group II received the same mailed invitation to participate as group I, but these subjects were not contacted again if they failed to return the postcard from the initial mailing. As for group I, group II subjects who returned postcards were scheduled for both an interview and for collection of the semen sample. The purpose of the group II mailing was to increase the sample size of the persons who responded to the initial mailing to establish the demographic and reproductive characteristics of the respondents. The respondents from groups I and II are compared with the nonrespondents from group I who were later recruited into the study. The similarity of parental characteristics for group I and group II (table 1) and the similar response rates to the initial mailing (18 percent in both groups) support our decision to combine the respondents to the initial group I and group II mailings. Comparison of the interview characteristics of group I and group II respondents (18 from group I and 23 from group II) is not very informative because of the small sample size. There were no remarkable or statistically significant differences in the sons’ characteristics between the two groups.

Location and intensive recruitment of group I subjects

If a subject in group I did not respond to the initial mailing, directory assistance was called to identify a telephone number, and/or his parent or sibling was contacted to obtain an updated address and telephone number. If there was no parent or sibling telephone number available, then a trace was done using various sources, including telephone directories and other residence directories. Second mailings were sent when new addresses were identified.

Several of the study subjects had a prior vasectomy, making them ineligible to provide a semen specimen. These respondents were not recruited for semen donation but were interviewed. Similarly, California residents who were found to live too far from the clinical site (outside the original area identified for the first mailing) were asked to complete the telephone interview. We identified a number of persons who were not accessible for the pilot study because of travel distances, but who could have participated in a full-scale study with more convenient facilities. These included eight subjects: Six had moved out of California, one was in jail, and one was out of state for the duration of the study.

Interview

The interview was based on the questionnaire from a previous study used to assess the time to pregnancy for partners of men exposed to diethylstilbestrol in utero (7). In that prior study, interviews with the partners of men whose mothers had taken diethylstilbestrol during pregnancy were used to confirm the time to pregnancy reported by the men themselves. We interviewed the men only to obtain reproductive history, occupational history and other demographics, and the time to pregnancy, because agreement between the partner’s and the men’s reports was very good in a prior study (Donna Day Baird, National Institute of Environmental Health Sciences, personal communication, 1999). The interview was a computer-assisted telephone survey, which required an average of 25 minutes to administer.

Semen evaluation

Subjects were asked to provide two semen specimens approximately 2 weeks apart, to ejaculate 2–5 days prior to their appointment, and to abstain from ejaculation for at least 48 hours before their appointment. At their first appointment, subjects were also asked for a blood sample, which was drawn after their semen was collected.
Because timing and temperature are critical to evaluation of the semen specimen, subjects were asked to collect the specimen at the clinic. A container was provided, and subjects were asked not to use lubricants or condoms and to report if the entire specimen was collected. At the time of the appointment, the subject was asked the date and time of his last ejaculation.

Three andrology laboratory technicians were trained at the University of California, Davis, to perform semen analysis according to a standard protocol (8) also used by the National Institute of Child Health and Human Development-funded National Cooperative Reproductive Medicine Network (6). Briefly, the semen analyses were performed within 1 hour of sample collection and included measurements of ejaculate volume, determinations of sperm concentration by the microcell technique, and counts of the percentage of motile sperm (8). Sperm morphology was assessed on Papanicolaou-stained seminal smears at the University of California, Davis. Sperm head, tail, and midpiece morphology was classified by a single technician according to the criteria published in 1987 by the World Health Organization (9). A second technician evaluated sperm head morphology on the same slides according to strict criteria (10).

Statistical methods

Analyses of proportions were based on contingency tables. Statistical significance was tested using chi-square tests or Fisher’s two-tailed tests when expected cell counts were less than five. Differences between the means for continuous semen parameters were tested with a t test for the case of equal variances, because the hypothesis that variances are equal was not rejected at p < 0.05 for any of the semen parameters.

RESULTS

The demographic characteristics of the parents of our study population are shown in table 1. Subjects residing closer to the andrology clinic (group I) were from families with somewhat lower unadjusted incomes and younger mothers than were subjects residing farther from the clinic (group II). Despite this difference, the positive response to the initial mailed invitation to participate in this study was identical for both groups (18 percent).

Completion rates for group I (intensive recruitment group)

We were able to locate all but 11 group I subjects by the end of the study for a location rate of 89 percent. Five subjects could not be located (4.9 percent), and six were still being traced when the study ended. Fourteen subjects refused participation (17 percent of those eligible). At the end of the study, the recruitment of 20 subjects was still in progress. We project a response rate of 53 percent among those eligible for semen donation and 70 percent among those eligible for the time-to-pregnancy interview, for a full-scale study with multiple clinic sites and flexible hours (figure 1).

Of the 13 men in group I who completed the interview but did not provide a semen specimen, three refused the semen
evaluation because of personal discomfort, and most cited inconvenient laboratory hours or inconvenient laboratory location \((n = 10)\). Participants in group I who provided a semen sample were significantly more likely to be college graduates (56 percent vs. 25 percent; \(p = 0.05)\).

**Respondents to initial mailing compared with nonrespondents**

A large proportion of project resources was spent in recruiting group I subjects who did not respond to the initial mailing. Completion rates for these nonrespondents were much lower than for respondents. All respondents to the initial mailing completed the interview compared with 46 percent of the eligible nonrespondents. Moreover, 83 percent of respondents provided at least one semen sample, compared with 20 percent of the eligible nonrespondents (men who were located and eligible). The mothers of respondents and nonrespondents were similar in age, race, parity, and income during their pregnancies.

Thirty-one men in group I who were nonrespondents to the initial mailing were eventually recruited to participate in the study. Using responses from their interviews, we are able to compare the characteristics of respondents to the initial mailing with the characteristics of men who did not respond to the initial mailing.

Race, marital history, and fertility history differed significantly between respondents to the initial mailing and recruited nonrespondents (table 2). Respondents were significantly more likely to be White, to have never married, and to have never fathered a pregnancy. Exclusion of four men who reported only or mostly male sexual partners throughout their adult life (all in the respondents category) decreased the differences in reproductive history between respondents and nonrespondents. However, among heterosexual men, respondents were still more likely to be never married (43.2 percent among respondents vs. 16.1 percent among nonrespondents; \(p = 0.02)\). Respondents to the initial mailing were more likely to provide a semen specimen than were nonrespondents (75.6 percent vs. 38.7 percent; \(p < 0.01)\).

Results of the semen evaluations for respondents to the initial mailing and nonrespondents are displayed in table 3. Only semen samples that were given in compliance with the study protocol are included in table 3. Respondents were more likely than nonrespondents to comply with the abstinence instructions before collecting both of the requested semen samples. However, all participants provided at least one sample according to the abstinence protocol.
As shown in table 3, respondents to the initial mailing had a higher mean number of sperm in the ejaculate (204.6 × 10^6 vs. 112.9; \( p < 0.02 \)), a higher mean sperm concentration (78.3 × 10^6/ml vs. 37.2; \( p = 0.003 \)), and a higher mean percentage of normal morphology according to 1987 World Health Organization criteria (58.7 percent vs. 53.3 percent; \( p < 0.04 \)). These differences were not due to better reported compliance with the abstinence protocol among the respondents. Semen samples that were given less than 2 days or greater than 5 days after the most recent ejaculation are excluded from table 3.

### DISCUSSION

We examined semen quality for respondents and recruited nonrespondents in a pilot study designed to determine whether the sons of a long-term pregnancy cohort, the CHDS, could be located and would agree to participate. We found that the target population could be located and that participation rates depended on the study protocol.

Projected completion rates for the interview (70 percent) were higher than for semen donation (53 percent). Of those men who provided a semen specimen, 88.6 percent (\( n = 39 \)) provided two, as requested. Of 83 specimens, 94 percent (\( n = 78 \)) were provided according to protocol (as reported by the subjects), with abstention from ejaculation for at least 48 hours and up to 5 days prior to the semen sample donation.

A number of limitations of this pilot study impacted completion rates. Most serious were the limited hours for the clinic where semen specimens were collected and the single location for specimen collection. Funds were not sufficient to support evening or weekend hours or the multiple locations more convenient to study subjects. The level of funding available in a full-scale study might still be insufficient to address these limitations.

The differences in semen quality that we observed between early respondents and the heavily recruited participants may be related to fertility concerns and/or different compliance with the abstinence protocol. However, demographic or reproductive characteristics did not explain

---

**TABLE 2. Response to invitational mailing and selected interview characteristics of all participants (groups I and II), Men’s Reproductive Health Study, Child Health and Development Studies, 1999*,†**

<table>
<thead>
<tr>
<th></th>
<th>Respondents (( n = 41 ))</th>
<th>Recruited nonrespondents (( n = 31 ))</th>
<th>( p ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Race, non-White‡</td>
<td>5</td>
<td>10</td>
<td>0.04</td>
</tr>
<tr>
<td>Education, did not graduate from college</td>
<td>22</td>
<td>17</td>
<td>0.92</td>
</tr>
<tr>
<td>Gross family income (current) ($)/year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;50,000</td>
<td>14</td>
<td>9</td>
<td>0.86</td>
</tr>
<tr>
<td>50,000–74,999</td>
<td>11</td>
<td>8</td>
<td>0.003</td>
</tr>
<tr>
<td>≥75,000</td>
<td>16</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>Marital status§</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never married‡</td>
<td>20</td>
<td>5</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Currently unmarried‡</td>
<td>22</td>
<td>9</td>
<td>0.04</td>
</tr>
<tr>
<td>Ever had children? Yes</td>
<td>17</td>
<td>18</td>
<td>0.16</td>
</tr>
<tr>
<td>Ever achieved a pregnancy? Yes‡</td>
<td>21</td>
<td>23</td>
<td>0.05</td>
</tr>
<tr>
<td>Ever had a pregnancy loss? Yes¶</td>
<td>4</td>
<td>4</td>
<td>0.72</td>
</tr>
<tr>
<td>Ever had a problem or been concerned about a problem with your fertility? Yes</td>
<td>7</td>
<td>2</td>
<td>0.28</td>
</tr>
<tr>
<td>Ever had intercourse without contraception for a period of at least 1 year? Yes</td>
<td>14</td>
<td>10</td>
<td>0.67</td>
</tr>
<tr>
<td>Provision of at least one semen specimen? Yes‡, #</td>
<td>31</td>
<td>12</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

* Respondents are subjects who responded to the invitational mailing and completed the interview and/or provided a semen specimen. One participant gave a semen sample but was not interviewed and therefore is not included in this table. Recruited nonrespondents are subjects who did not respond to the invitational mailing but who were eventually recruited for an interview and/or semen specimen.

† Subjects in group I were actively recruited to maximize participation. Subjects in group II received an invitational mailing only.

‡ Statistically significant differences (\( p < 0.05 \)); \( p \) values are determined using chi-square tests or Fisher’s two-tailed tests when expected cell counts are less than five.

§ Categories overlap and do not add up to 100%.

¶ Pregnancy losses include miscarriages, blighted ova, stillbirths, tubal pregnancies, and molar pregnancies.

# One participant gave a semen sample but was not interviewed and therefore is not included in this table but is included in table 3. The difference in semen provision rates remains significant when men with vasectomies are excluded.
Finally, intense recruitment is also inefficient. Recruitment subjects who responded to our initial invitation to the study. comply with the semen collection protocol than were the among intensely recruited subjects, who were less likely to longer follow-up. The quality of participation was also poor with multiple sites for semen collection and resources for view and 36 percent for the semen sample in a large study could have been increased only to 62 percent for the inter-

Respondents are subjects who responded to the invitational mailing and who provided semen specimens. Recruited nonrespondents are subjects who did not respond to the invitational mailing but who were eventually recruited to provide a semen specimen.

† Subjects in group I were actively recruited to maximize participation. Subjects in group II received an invitational mailing only.

‡ Semen characteristics for both respondents and nonrespondents are based on semen samples that met the ejaculation criteria: within 2–5 days of last ejaculation.

§ SE, standard error.

¶ Semen samples with reported loss of semen are coded with missing values for total number of sperm in semen and total volume of semen. One subject reported semen loss for all samples and thus has no value of volume or total number of sperm.

# Statistically significant differences ($p < 0.05$); $p$ values are determined using Fisher’s two-tailed tests for proportions and by $t$ tests for continuous variables assuming equal variances, since the hypothesis that variances are equal for the two groups was not rejected for any of the semen parameters.

** The traditional method for determining morphology is described in the 1987 World Health Organization publication (WHO laboratory manual for the examination of human semen and semen-cervical mucus interaction. Cambridge, UK: The Press Syndicate of the University of Cambridge, 1987); the strict method is described in the 1999 publication by the same body (WHO laboratory manual for the examination of human semen and sperm-cervical mucus interaction. Cambridge, UK: Cambridge University Press, 1999).

**

TABLE 3. Response to invitational mailing and selected semen characteristics of all participants who provided semen specimen (groups I and II), Men’s Reproductive Health Study, Child Health and Development Studies, 1999*, †, ‡,

<table>
<thead>
<tr>
<th></th>
<th>Respondents (n = 31) (mean (SE§))</th>
<th>Recruited nonrespondents (n = 13) (mean (SE))</th>
<th>$p$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volume of ejaculate (ml)¶</td>
<td>2.79 (0.25)</td>
<td>3.3 (0.30)</td>
<td>0.26</td>
</tr>
<tr>
<td>No. of sperm ($\times 10^9$)¶#</td>
<td>204.64 (22.39)</td>
<td>112.88 (21.45)</td>
<td>0.02</td>
</tr>
<tr>
<td>Sperm concentration (10/ ml)#</td>
<td>78.31 (7.74)</td>
<td>37.21 (7.46)</td>
<td>0.003</td>
</tr>
<tr>
<td>Sperm motility (% motile)‡</td>
<td>47.21 (2.10)</td>
<td>49.29 (4.08)</td>
<td>0.62</td>
</tr>
<tr>
<td>Sperm morphology (% normal)**</td>
<td>58.65 (1.28)</td>
<td>53.26 (2.50)</td>
<td>0.04</td>
</tr>
<tr>
<td>Traditional method#</td>
<td>12.84 (0.93)</td>
<td>11.52 (1.35)</td>
<td>0.43</td>
</tr>
<tr>
<td>Strict method</td>
<td>12.84 (0.93)</td>
<td>11.52 (1.35)</td>
<td>0.43</td>
</tr>
</tbody>
</table>

No. of semen specimens provided

<table>
<thead>
<tr>
<th></th>
<th>No.</th>
<th>%</th>
<th>No.</th>
<th>%</th>
<th>$p$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>One</td>
<td>2</td>
<td>6.5</td>
<td>3</td>
<td>23.1</td>
<td>0.14</td>
</tr>
<tr>
<td>Two</td>
<td>29</td>
<td>93.6</td>
<td>10</td>
<td>76.9</td>
<td></td>
</tr>
</tbody>
</table>

No. of semen specimens provided that meet abstinence criteria

<table>
<thead>
<tr>
<th></th>
<th>No.</th>
<th>%</th>
<th>No.</th>
<th>%</th>
<th>$p$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>One#</td>
<td>4</td>
<td>12.9</td>
<td>6</td>
<td>46.2</td>
<td>0.04</td>
</tr>
<tr>
<td>Two#</td>
<td>27</td>
<td>87.1</td>
<td>7</td>
<td>53.8</td>
<td></td>
</tr>
</tbody>
</table>

* Respondents are subjects who responded to the invitational mailing and who provided semen specimens.

† Subjects in group I were actively recruited to maximize participation. Subjects in group II received an invitational mailing only.

‡ Semen characteristics for both respondents and nonrespondents are based on semen samples that met the ejaculation criteria: within 2–5 days of last ejaculation.

§ SE, standard error.

¶ Semen samples with reported loss of semen are coded with missing values for total number of sperm in semen and total volume of semen. One subject reported semen loss for all samples and thus has no value of volume or total number of sperm.

# Statistically significant differences ($p < 0.05$); $p$ values are determined using Fisher’s two-tailed tests for proportions and by $t$ tests for continuous variables assuming equal variances, since the hypothesis that variances are equal for the two groups was not rejected for any of the semen parameters.

** The traditional method for determining morphology is described in the 1987 World Health Organization publication (WHO laboratory manual for the examination of human semen and semen-cervical mucus interaction. Cambridge, UK: The Press Syndicate of the University of Cambridge, 1987); the strict method is described in the 1999 publication by the same body (WHO laboratory manual for the examination of human semen and sperm-cervical mucus interaction. Cambridge, UK: Cambridge University Press, 1999).

Intensive recruitment of subjects to participate in a demanding semen collection protocol does not appear to solve the problem of potential bias due to nonresponse. Even after intensive effort, the “nonresponders” were not well represented. Only 46 percent completed the telephone interview and only 20 percent provided semen. We estimated that these could have been increased only to 62 percent for the interview and 36 percent for the semen sample in a large study with multiple sites for semen collection and resources for longer follow-up. The quality of participation was also poor among intensely recruited subjects, who were less likely to comply with the semen collection protocol than were the subjects who responded to our initial invitation to the study. Finally, intense recruitment is also inefficient. Recruitment had a low yield and associated high costs. On the other hand, our results demonstrate that great gains in efficiency are possible if semen donation is limited entirely to those subjects who responded to the initial mailing.

Considering our findings and response rates well below 50 percent reported for semen studies conducted around the world (1–4), we believe that it is unlikely that a representa-

Download the age sample of men can be recruited for a study of semen characteristics.

However, this limitation can be addressed by designs that quantify sources of bias to allow appropriate interpretation of study findings. We suggest that it may be advisable to conduct the semen evaluation component of a population-based study with those who respond positively to an initial mailing. We would not recruit the remaining men for the semen donation. Instead, we recommend collecting additional exposure and reproductive measures by interview from both semen donors (respondents to the invitation) and a high percentage of the subjects who do not respond, by actively recruiting for the interview only. These data can be used to assess potential bias in the semen evaluation component of the study.
The possibility of addressing bias in semen collection studies depends on the hypothesis under study. When the hypothesis of interest concerns an exposure that can be assessed in both semen donors and nondonors, bias may be more easily addressed. One example is a study of prenatal influences on semen quality in a pregnancy cohort with stored serum samples. Another example is an occupational study where exposures are known for a listed sample. In these cases, associations between the exposure of interest (e.g., prenatal organochlorines or years employed in a high-risk occupational category) can be measured in both semen donors and nondonors. This allows a direct estimation of nonresponse bias for the exposure of interest. If the exposure is unrelated to response status, response bias cannot explain study findings. However, if the exposure is related to response status, full characterization of response bias may still not be possible.

Even for listed samples, contemporary information on potential bias should also be collected. In this pilot study, we have shown that a time-to-pregnancy questionnaire, nearly identical to the instrument used in a recent study of sons of women who received diethylstilbestrol (7), is acceptable in a general population. We project a 70 percent response rate to this questionnaire, even when subjects were initially recruited with full knowledge that a semen sample would also be requested. There were virtually no refusals on questionnaire items, despite their personal nature, and no terminations of the interviews in progress.

For study designs where the exposure of interest is contemporary, that is, where the data collection is measuring both the exposure of interest as well as reproductive outcomes, addressing response bias is more difficult. However, one strategy is to assess alternative markers of reproductive function (such as time to pregnancy) in addition to semen parameters. Assessment and analysis of biomarkers for contemporary exposures as well as surrogate fertility measures in blood or urine from subjects who both do and do not provide semen specimens would also be useful for estimating response bias. For this reason, one could allow for and encourage participation in blood or urine collections for all targeted subjects, even those who do not participate in the semen evaluation. Biomarker data should be supplemented by questionnaire items that assess the subject’s reproductive history, socioeconomic status, and other variables that might be related to the exposure of interest, the outcome, and participation in the semen collection phase of the study. Every effort should be made to maximize response to an initial data collection that does not involve a semen collection. This strategy may not completely account for response bias in the semen study, but it will facilitate analysis critical to appropriate interpretation of findings. Analyses of potential bias should be published along with study findings.

It is possible that the semen protocol in our pilot study reduced the participation rate for the interview, because the entire study protocol was introduced to subjects in one step, including our intent to request a semen sample. In view of the importance of high response rates to the interview phase of the data collection, the study might be better conducted in two stages, interview first and semen request second.

Alternative protocols for semen collection and evaluation should also be considered for use in future studies. It is possible to obtain reliable data on semen parameters other than sperm motility from a specimen that is frozen and stored in a home refrigerator freezer (James Overstreet, University of California, Davis, unpublished data, 2001). The design of special containers for overnight shipment of semen samples is also being tested (11), although some semen parameters are impacted by delayed analysis, including the percentage of motile sperm (12). The value of sperm motility measurements needs to be carefully weighed against the increased participation rate that can be expected if subjects are given the option of collecting samples at the time and location of their choice. Priority should be given to the pilot testing of semen protocols for home collection and storage of the sample.

We suggest that it may not be possible to obtain semen samples representative of a general population in epidemiologic studies. However, we believe that even a nonrepresentative sample may still be of great value in investigating the hypotheses of adverse effects of environmental factors on male reproduction. Study designs should focus on multiple sources of information that will permit assessment of bias. Listed cohorts with existing biologic samples or exposure measures offer important advantages because they allow assessment of the relation between exposures of interest and study participation. Studies of contemporary adult exposures should include a stepped data collection that permits characterization of differences between participants in the semen collection and nonparticipants. This data collection should include characterization of exposures and measures of reproductive health other than semen quality, based on questionnaires or biomarkers. Studies designed to these standards will make responsible interpretation of findings possible. This may be particularly critical for the assessment of environmental hazards, where findings are likely to be the focus of politically and emotionally charged discussion.

ACKNOWLEDGMENTS

Supported by the National Institute of Environmental Health Sciences (R03 ES08612 and R01 ES08345) and the National Institute of Child Health and Human Development (N01-HD-6-3258).

The authors thank the late Dr. J. Yerushalmy for designing and assembling this unique cohort and his colleagues, Dr. Barbara J. van den Berg and Roberta Christianson, for maintaining the cohort and for reviewing early versions of the study design and the manuscript. They also thank Dr. Joseph D’Amico and Dr. Tsai Wu of the Kaiser Foundation Research Institute, Santa Clara, California, for clinical andrology studies and Dr. James Wiley, Public Health Institute, and Susan Burns, Survey Research Center, University of California, Berkeley, for implementation of subject recruitment. They want to thank Bonnie Davis, Public Health Institute, for supervising and implementing the interview and Angela Anderson, Public Health Institute, for administrative and clerical support.
REFERENCES