Contribution of Follow-up of Nonresponders to Prevalence and Risk Estimates: A Norwegian Respiratory Health Survey

Jan Brøgger¹, Per Bakke¹, Geir E. Eide²,³, and Amund Gulsvik¹

¹ Department of Thoracic Medicine, Institute of Medicine, University of Bergen, Bergen, Norway.
² Centre for Clinical Research, Haukeland University Hospital, Bergen, Norway.
³ Section for Medical Statistics, University of Bergen, Bergen, Norway.

Received for publication May 16, 2002; accepted for publication September 30, 2002.

Achieving a high response rate is often expensive and time consuming. Does an extensive survey effort change prevalence estimates and exposure-disease relations? In 1998–1999, the authors conducted a population-based respiratory health survey in two Norwegian counties (Oslo, n = 20,000; Hordaland, n = 5,000) of a random sample of the adult population aged 15–70 years. A postal questionnaire was mailed, with as many as two reminder letters. A 25% random sample of postal nonresponders was contacted by ordinary or cellular telephone. Cumulative response rates after the first mailing, first reminder, second reminder, and telephone follow-up were 42.7%, 60.7%, 68.3%, and 79.9%, respectively. Compared with initial postal responders, responders at later stages were younger, and more were male. Late responders had a lower educational level, were more often smokers, and were occupationally exposed to dusts or gases more frequently. After the authors adjusted for these factors, late responders were found to have less asthma, attacks of breathlessness, and hay fever. Hardly any changes in prevalences and odds ratios were noted when initial responders were compared with all responders. The additional contribution of sending reminder letters and conducting a telephone follow-up to prevalence estimates and the exposure-disease relation was small. A single letter would have produced nearly the same prevalence estimates and risk factor associations.

Nonresponse bias is a potential problem in most population surveys (1, 2). Definitive investigation of nonresponse bias would require a 100 percent response rate, which is costly and nearly impossible to achieve. The “willingness to respond” approach has been used instead (3–5). Early responders or responders who are easier to reach are compared with later responders or responders who are more difficult to reach. In essence, survey effort is used as a proxy for nonresponse bias. This approach has been applied to prevalence estimates in cross-sectional population surveys on a wide range of topics (6–20). We are aware of only two studies that have looked at whether exposure-disease relations are in fact modified by a more intensive survey effort, and these studies did not have high statistical power (20–23). Additionally, approaching those nonresponders who do not have an ordinary telephone through their cellular phones may be a new way of improving response rates. To our knowledge, a systematic review of nonresponse bias in respiratory epidemiology has not been performed before.

To investigate the effect of increasing survey efforts on results, we performed a large-scale, population-based survey on respiratory health by using a postal questionnaire with a telephone and cellular phone follow-up. We also conducted a systematic literature review of studies of nonresponse bias in respiratory epidemiology.

MATERIALS AND METHODS

This cross-sectional study of the adult population was conducted in two regions of Norway, the capital county of Oslo and the county of Hordaland. The sampling frame was inhabitants in the Central Population Register aged 15–70 years on December 31, 1997. Random samples of 20,000 (Oslo) and 5,000 (Hordaland) inhabitants were selected.
Impact of Nonresponse Follow-up

Fieldwork

A postal questionnaire was mailed to each subject. After 3 weeks and 8 weeks, nonresponders were sent a reminder letter. Responders were classified as initial responders and responders to the first and second reminder letters.

A telephone follow-up was initiated for a 25 percent random sample of postal nonresponders. Telephone numbers for the follow-up sample were located by using a direct mail firm (DM Huset, Oslo, Norway), requiring a match on name, date of birth, and address. Six months after the first mailing, computer-assisted telephone interviews were performed by trained operators. Nine months after the study started, nonresponders without a telephone directory entry but with a cellular phone were located from the customer databases of the cellular phone providers. A similar cellular telephone interview was conducted.

Questionnaires were mailed, responses were processed, and telephone interviews were performed by the Interview Office of Statistics Norway.

Questionnaire

The questions regarding symptoms were taken from a modified translation of the British Medical Research Council on Chronic Bronchitis (24, 25). A one-sheet questionnaire designed for self-completion was used; it contained 45 questions to be answered by all subjects on 18 respiratory symptoms, seven respiratory and cardiac diagnoses, and 20 risk factors. Smokers were defined as daily smokers at the time of the study. Former smokers were persons who had smoked daily but had stopped (26). The study was approved by the Norwegian Data Inspectorate.

Statistical methods

Responders were categorized as initial responders, responders to the first reminder letter, responders to the second reminder letter, and responders to the telephone-based follow-up. These categories represented “survey effort.” Prevalences and cumulative prevalences of symptoms or diagnoses by survey effort were computed, with 95 percent confidence intervals. A two-sided \( p \) value of less than 5 percent was taken as statistically significant. Equality between groups was investigated by using analysis of variance, ordered logistic regression, or logistic regression, as applicable. All statistical tests were of heterogeneity, not trend. To control for possible confounding of prevalence estimates, multiple logistic regression was performed (27).

Age was entered as a cubic spline with knots at 15, 30, 50, and 70 years (28). Smoking habits was modeled as pack-years. Pack-years was modeled as a linear term for current smokers but had stopped (26). The study was approved by the Norwegian Data Inspectorate.

those significant at \( p < 0.10 \) were considered further. In addition, the odds ratios of the risk factors for all responders were computed. A missing response was defined as a selection of “don’t know” or no response to a given question. Subjects for whom answers were missing in an analysis were excluded from that analysis. All calculations accounted for the sampling in the follow-up by using weighting and robust standard errors. Adjusted Wald tests were used to decide on the significance of terms. All data were analyzed by using Stata 7.0 software (30).

Literature search

A systematic literature search was performed to find studies of nonresponse bias in respiratory epidemiology. The following databases were searched in early 2002: MEDLINE (Bethesda, Maryland), CINAHL (Glendale, California), and the Cochrane databases through Ovid (New York, New York; http://www.ovid.com/). We also searched the database Science Citation Index Expanded (http://wos.isiglobalnet.com/) by using the phrase ("nonresponse" or “response” or “nonresponder” or “responder”) and (“epidemiology” or “health surveys”) and (pulm* or respirat* or lung*)). Studies that cited Drane et al. (31, 32) were identified by searching the citation database Science Citation Index Expanded. Matching references were retrieved and were screened on the basis of title and abstract. Only those studies of the general population in a defined geographic area on respiratory symptoms and/or diagnoses with at least a 10-year age span in the interval 15–70 years were included. Reference lists of retrieved articles were also searched.

For simplicity purposes and to enable a conservative comparison of response rates across studies, response rates were recomputed as the ratio of the number of answers to the number of subjects who had not yet answered at that stage, regardless of whether they had received the letter, had a telephone, were dead, or had moved. Some numbers were calculated from percentages or read from graphs. The total response rate was recalculated with weighting for possible subsampling in the follow-up, when available.

RESULTS

The response rates for the first mailing, the first reminder letter, and the second reminder letter were 42.7 percent, 31.4 percent, and 19.5 percent, respectively. There were 7,917 postal nonresponders, of whom 864 had moved and had an unknown address, 271 had actively refused, and 88 were excluded for other reasons. Of the remaining 6,694 nonresponders, 1,669 were randomly selected for follow-up. Telephone numbers were identified for 1,176 (70.5 percent). Of the 493 persons for whom no telephone number was found, a cellular phone number was available for 175 (35.5 percent). The response rate for the ordinary telephone interview (n = 660) was 56.1 percent, while that for the cellular phone interview (n = 61) was 34.9 percent.

The cumulative response rates after the first mailing, the first reminder letter, the second reminder letter, and the telephone follow-up were 42.7 percent, 60.7 percent, 68.3
TABLE 1. Age, gender, educational level, smoking habits, and occupational exposure, by intensity of follow-up, of subjects in the Oslo-Hordaland respiratory disease survey, Norway, 1998–1999

<table>
<thead>
<tr>
<th>Education</th>
<th>All responders (n = 17,804)</th>
<th>Responders to the initial letter (n = 10,671)</th>
<th>Responders to the first reminder (n = 4,497)</th>
<th>Responders to the second reminder (n = 1,915)</th>
<th>Responders to the telephone follow-up (n = 721)</th>
<th>Value 95% CI</th>
<th>Value 95% CI</th>
<th>Value 95% CI</th>
<th>Value 95% CI</th>
<th>Value 95% CI</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>High school (%)</td>
<td>41.9</td>
<td>41.1, 42.8</td>
<td>40.8, 41.8</td>
<td>41.8, 42.2</td>
<td>43.0, 42.7</td>
<td>49.3, 47.1</td>
<td>49.5, 47.3</td>
<td>49.6, 48.0</td>
<td>49.5, 48.0</td>
<td>50.0, 49.0</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>University (%)</td>
<td>43.3</td>
<td>42.4, 44.1</td>
<td>46.6, 47.5</td>
<td>41.4, 42.9</td>
<td>40.3, 41.8</td>
<td>43.0, 41.9</td>
<td>42.8, 43.2</td>
<td>43.0, 44.8</td>
<td>43.0, 44.8</td>
<td>43.5, 45.0</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Occupational exposure (%)</td>
<td>42.3</td>
<td>41.5, 43.1</td>
<td>39.9, 40.8</td>
<td>42.2, 43.7</td>
<td>45.9, 44.8</td>
<td>48.9, 47.5</td>
<td>48.0, 48.6</td>
<td>48.0, 50.2</td>
<td>48.0, 50.2</td>
<td>49.3, 51.6</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

* Weighted n = 19,412.
† CI, confidence interval.
‡ p values for current smokers only.

percent, and 79.9 percent, respectively. The respective response rates in Oslo were 41.8 percent, 30.1 percent, 18.5 percent, and 34.6 percent; those in Hordaland were higher at 46.4 percent, 36.8 percent, 24.0 percent, and 46.6 percent for each successive phase of follow-up.

Willingness to respond was associated with age, gender, educational level, smoking habits, and occupational exposure to dusts or gases univariately (table 1). Men responded later than women. Age and educational level decreased with increasing contact effort. Current smoking and occupational exposure were more prevalent and pack-years increased with increasing contact effort.

The mean age of the sample was 40.0 years compared with 41.9 years for initial responders only and 41.0 years for all responders. The percentage of women in the sample was 50.0 percent. Women constituted 54.3 percent of the initial responders and 52.1 percent of all responders. Changes in prevalence by the end of the survey compared with prevalence for only the immediate responders were small (<2.5 percentage points), except that the prevalence of current smokers increased by 3.8 percentage points (from 30.1 percent to 33.9 percent). The mean number of pack-years of smoking increased by 0.3. Confidence intervals were only slightly narrower in the final estimate.

The crude prevalences of symptoms and diagnoses by survey effort are shown in table 2. All symptoms were associated with survey effort except for attacks of breathlessness, dyspnea when walking two stairs, and dyspnea at rest. There was no uniform tendency toward an increase or decrease in prevalence with survey effort. Note in particular the large change for morning cough in responders to the telephone follow-up. No association for diagnoses was found.

After adjustment for age, gender, county, educational level, pack-years, and occupational exposure, there was no longer any significant change in chronic cough and wheezing, whereas asthma was now associated with survey effort (table 3). No uniform tendency toward a decrease or increase in the odds ratios with survey effort was evident. Changes in the prevalence of symptoms or diagnoses by the end of the survey compared with the prevalence of only the immediate responders were small (<1 percentage point). The changes appeared to be stronger for the telephone responders.

The effect of risk factors on symptoms might vary with survey effort. Figure 1 shows the association of risk factors with survey effort of those symptoms or diagnoses for which the p value of the test of a change in one or more risk factors with survey effort was less than 0.10. Risk factors evaluated were gender, occupational exposure to dusts or gases, and pack-years. Of 16 symptoms or diagnoses, only chronic cough (joint-test \( p = 0.06 \)) and ever wheezing (joint-test \( p = 0.10 \)) reached the threshold. For chronic cough, the effect of gender was found to vary by survey effort (\( p = 0.05 \)), probably driven by the large effect of gender among telephone responders. For wheezing, the effect of pack-years was significantly different (\( p = 0.03 \)), probably because of the lesser effect of pack-years among telephone responders. These changes had little impact, since changes in odds ratios by the end of the survey compared with the odds ratios of only the immediate responders were small.

Analyses were rerun separately for Oslo and Hordaland. Analyses of unadjusted and adjusted prevalences, and associations with risk factors, by county showed the same trends and magnitudes of associations.

Cost and time were substantial for the follow-up. The second reminder letter increased the cost of the survey by 25 percent. The telephone follow-up increased the cost of the survey by an additional 14 percent and netted 741 responses.
TABLE 3. Adjusted* odds ratios of symptoms or diagnoses, by intensity of follow-up, among subjects in the Oslo-Hordaland respiratory disease survey,† Norway, 1998–1999

<table>
<thead>
<tr>
<th></th>
<th>First reminder</th>
<th>Second reminder</th>
<th>Telephone follow-up</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR†</td>
<td>95% CI†</td>
<td>OR</td>
<td>95% CI</td>
</tr>
<tr>
<td>Morning cough</td>
<td>1.07</td>
<td>0.98, 1.17</td>
<td>0.98</td>
<td>0.86, 1.11</td>
</tr>
<tr>
<td>Day cough</td>
<td>1.08</td>
<td>0.97, 1.21</td>
<td>1.25</td>
<td>1.08, 1.45</td>
</tr>
<tr>
<td>Phlegm when coughing</td>
<td>1.04</td>
<td>0.94, 1.14</td>
<td>1.04</td>
<td>0.91, 1.18</td>
</tr>
<tr>
<td>Chronic cough</td>
<td>1.08</td>
<td>0.96, 1.21</td>
<td>1.17</td>
<td>1.00, 1.36</td>
</tr>
<tr>
<td>1 episode of cough and phlegm</td>
<td>1.04</td>
<td>0.94, 1.15</td>
<td>0.96</td>
<td>0.83, 1.11</td>
</tr>
<tr>
<td>≥2 episodes of cough and phlegm</td>
<td>1.05</td>
<td>0.95, 1.16</td>
<td>1.03</td>
<td>0.90, 1.18</td>
</tr>
<tr>
<td>Dyspnea while walking uphill</td>
<td>1.15</td>
<td>1.05, 1.27</td>
<td>1.08</td>
<td>0.94, 1.23</td>
</tr>
<tr>
<td>Dyspnea while walking two stairs</td>
<td>1.10</td>
<td>0.98, 1.23</td>
<td>1.08</td>
<td>0.92, 1.27</td>
</tr>
<tr>
<td>Dyspnea while walking on level ground</td>
<td>1.00</td>
<td>0.81, 1.24</td>
<td>1.47</td>
<td>1.13, 1.92</td>
</tr>
<tr>
<td>Dyspnea at rest</td>
<td>1.00</td>
<td>0.74, 1.34</td>
<td>1.31</td>
<td>0.90, 1.89</td>
</tr>
<tr>
<td>Attacks of breathlessness</td>
<td>1.06</td>
<td>0.96, 1.18</td>
<td>0.88</td>
<td>0.75, 1.02</td>
</tr>
<tr>
<td>Wheezing ever</td>
<td>1.07</td>
<td>0.98, 1.17</td>
<td>1.12</td>
<td>0.99, 1.26</td>
</tr>
<tr>
<td>Hay fever ever</td>
<td>0.97</td>
<td>0.88, 1.06</td>
<td>0.95</td>
<td>0.83, 1.09</td>
</tr>
<tr>
<td>Asthma, physician's diagnosis</td>
<td>0.90</td>
<td>0.79, 1.03</td>
<td>0.74</td>
<td>0.60, 0.91</td>
</tr>
<tr>
<td>Bronchitis, physician's diagnosis</td>
<td>1.07</td>
<td>0.98, 1.18</td>
<td>0.97</td>
<td>0.85, 1.11</td>
</tr>
<tr>
<td>Emphysema, physician's diagnosis</td>
<td>0.64</td>
<td>0.38, 1.09</td>
<td>0.89</td>
<td>0.44, 1.82</td>
</tr>
</tbody>
</table>

* Adjusted, by using logistic regression, for age, gender, county, educational level, pack-years of smoking, and occupational exposure.
† Initial postal responders are the reference.
‡ OR, odds ratio; CI, confidence interval.

DISCUSSION

We have shown an association between survey effort regarding prevalences and exposure-disease relations, but

Am J Epidemiol 2003;157:558–566
changes were small when we compared all responders with only immediate responders. The effect of gender on chronic cough varied with survey effort, and the effect of pack-years on wheezing varied with survey effort. Through a telephone-based follow-up, we managed to increase the response rate from 68.3 percent to 79.9 percent. Although we found little evidence of nonresponse bias by survey effort, average age improved toward the sample mean of 40.0 years from the initial responders only (41.9 years) through all responders (41.0 years).

Previous surveys using exactly the same postal survey methods had response rates of 89 percent in Oslo in 1972 (42) and 90 percent in Hordaland in 1985 (23). This decrease in response rates may reflect a temporal trend (43), although solid long-term data on response rates indicate that this trend varies with country and the circumstances of the survey (44). There may be reasons other than a simple decline in willingness to respond (4). The introductory text was longer because of regulatory and ethical concerns, and the questionnaire itself was slightly longer. The previous surveys were executed locally and were mailed in a University of Bergen envelope. The present survey was performed with the aid of Statistics Norway and was sent in a Statistics Norway envelope.

Oslo had a response rate 4–12 percent lower than that for Hordaland at each stage of survey follow-up. The survey methods were exactly the same in both counties. A number of sociodemographic differences between Oslo and Hordaland might explain the result. Oslo is the capital of Norway and is the undisputed economic and political center. It is a wholly urban area of 500,000 persons situated in a greater metropolitan area of about 1 million inhabitants (out of 4.5 million inhabitants in Norway as a whole). Hordaland is the center of coastal western Norway, a mixed urban and rural area with the city of Bergen (250,000 inhabitants) at its center (45). Gross income is 20 percent higher in Oslo than in Bergen (45).

Oslo has a large population of foreigners (foreign born or born to one or both foreign parents), constituting 19 percent of the population, whereas the figure is only 6.4 percent in Bergen (46). In the 2000 Norwegian census, the response rate was as low as 56 percent among those from Asia, Africa, and Latin America. When we reran the analyses separately by study area, the effects were of the same magnitude and direction.

We identified a telephone number for only 70 percent of the follow-up sample, whereas telephone coverage is universal in Norway. This figure was probably due to our conservative telephone book look-up algorithm. However, we then found a cellular telephone number for as many as 35 percent of those for whom we could not identify an ordinary telephone number.

For both the adjusted and unadjusted estimates, it appeared that the changes among telephone responders were driving most of the associations. Telephone responders had a lower risk of symptoms. The use of a telephone interview instead of a postal questionnaire may explain this finding. However, in an earlier study (47), we found few differences between the postal and the telephone survey mode. A previous Norwegian study found that postal nonresponders were more likely to be smokers but had fewer symptoms (23). Telephone responders are probably less symptomatic and hence probably perceive little benefit from participating in a postal questionnaire. When questioned directly on the telephone, resistance to response may be overcome. Our findings support this theory.
TABLE 4. Results of systematic review of the initial sample size, number of responders, and response rates (%),* by intensity of follow-up, in studies of nonresponse bias in respiratory health surveys of the general adult population†

<table>
<thead>
<tr>
<th>Authors (reference no.)</th>
<th>Country</th>
<th>Start of study</th>
<th>Sample size (no.)</th>
<th>Initial response</th>
<th>One reminder</th>
<th>Two reminders</th>
<th>Telephone or home visit</th>
<th>Total response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Siemiatycki and Campbell (22)</td>
<td>Canada</td>
<td>1974</td>
<td>1,555</td>
<td>No. 731, % 47</td>
<td>No. 334, % 41</td>
<td>NA‡</td>
<td>No. 193, % 81</td>
<td></td>
</tr>
<tr>
<td>Siemiatycki and Campbell (22)§</td>
<td>Canada</td>
<td>1974</td>
<td>1,595</td>
<td>No. 1,160, % 73</td>
<td>NA</td>
<td>NA</td>
<td>247, % 57</td>
<td></td>
</tr>
<tr>
<td>Howard et al. (33)</td>
<td>United States</td>
<td>1979</td>
<td>2,981</td>
<td>No. 739, % 25</td>
<td>No. 194, % 39</td>
<td>NA†</td>
<td>789, % 58</td>
<td></td>
</tr>
<tr>
<td>Bakke et al. (23)</td>
<td>Norway</td>
<td>1985</td>
<td>5,000</td>
<td>No. 3,147, % 63</td>
<td>No. 1,030, % 56</td>
<td>No. 292, % 35</td>
<td>50, % 78</td>
<td></td>
</tr>
<tr>
<td>Chinn et al. (34)</td>
<td>United Kingdom</td>
<td>1990</td>
<td>15,000</td>
<td>No. 6,114, % 41</td>
<td>No. 1,699, % 19</td>
<td>No. 1,058, % 15</td>
<td>240, % 66</td>
<td></td>
</tr>
<tr>
<td>Crane et al. (35)#</td>
<td>New Zealand</td>
<td>1991</td>
<td>16,184</td>
<td>No. 11,111, % 69</td>
<td>No. 379, % 16</td>
<td>No. 545, % 26</td>
<td>205, % 71</td>
<td></td>
</tr>
<tr>
<td>Abramson et al. (36)</td>
<td>Australia</td>
<td>1992</td>
<td>4,500</td>
<td>No. 2,033, % 45</td>
<td>No. 397, % 16</td>
<td>No. 545, % 26</td>
<td>205, % 71</td>
<td></td>
</tr>
<tr>
<td>Ronnmark et al. (37)</td>
<td>Sweden</td>
<td>1992</td>
<td>9,132</td>
<td>No. 10,800, % 71</td>
<td>No. 1,558, % 37</td>
<td>No. 1,088, % 40</td>
<td>89, % 31</td>
<td></td>
</tr>
<tr>
<td>Bjornsson et al. (38)**</td>
<td>Sweden</td>
<td>1992</td>
<td>10,800</td>
<td>No. 6,555, % 61</td>
<td>No. 1,558, % 37</td>
<td>No. 1,088, % 40</td>
<td>89, % 31</td>
<td></td>
</tr>
<tr>
<td>de Marco et al. (39)††</td>
<td>Italy</td>
<td>1992</td>
<td>3,000</td>
<td>No. 1,470, % 49</td>
<td>No. 230, % 15</td>
<td>No. 169, % 13</td>
<td>147, % 13</td>
<td></td>
</tr>
<tr>
<td>de Marco et al. (39)‡‡</td>
<td>Italy</td>
<td>1992</td>
<td>4,000</td>
<td>No. 1,700, % 43</td>
<td>No. 310, % 72</td>
<td>No. 999, % 22</td>
<td>941, % 66</td>
<td></td>
</tr>
<tr>
<td>Hedman et al. (40)§§</td>
<td>Finland</td>
<td>1996</td>
<td>4,300</td>
<td>No. 1,700, % 43</td>
<td>No. 310, % 72</td>
<td>No. 999, % 22</td>
<td>941, % 66</td>
<td></td>
</tr>
<tr>
<td>Mono et al. (41)§§</td>
<td>Spain</td>
<td>1998</td>
<td>25,000</td>
<td>No. 10,671, % 43</td>
<td>No. 4,497, % 31</td>
<td>No. 1,915, % 20</td>
<td>741, % 44</td>
<td></td>
</tr>
</tbody>
</table>

* Response rates were recomputed; refer to the text for details.
† All studies except the second one included a postal survey with telephone follow-up.
‡ NA, not applicable.
§ Telephone survey first with telephone or home visit follow-up. Data given as initial telephone responders, with subsequent telephone or home follow-up (left to right).
¶ Data not given.
# Did not give the sampling fraction of the telephone follow-up. True response rate >74% since data for telephone responders should be weighted up.
** Data given as early responders (<4 weeks), intermediate responders (5–8 weeks), late responders (>8 weeks), and telephone responders (left to right).
†† Verona only.
‡‡ Pavia and Turin.
§§ Only the cumulative response rate after one reminder letter was given.

For differences in prevalence, our study had substantial power. The tests of changes in risk factors with survey effort would have smaller power. There was also some concern about multiple statistical tests, especially for the interactions of risk factors with survey effort. However, we tested a limited number of plausible interactions in a hierarchic fashion designed to reduce the number of false positives (48). There may have been differences in exposure-disease associations for other symptoms, but our data could not provide enough precision to pinpoint them. Differences for other risk factors may also have been possible. These we did not test, to avoid multiple comparisons.

The systematic literature search found 11 respiratory epidemiology articles that contained some data on nonresponse bias (22, 23, 33–41). The response rates to the first postal questionnaire varied from less than 22 percent to 69 percent. A number of reasons for this variation are possible. Mono et al. (41) had a very low response rate, which, even after one reminder, was only 22 percent. These authors used a 1991 census sample frame, and it appears that the study was initiated in or about 1996. The study by Howard et al. (33) was conducted in a bilingual Hispanic population and used a commercial household directory that was not as up-to-date. Clearly, a prerequisite for a high response rate is a good-quality, up-to-date sampling frame. Even so, the remaining initial response rates varied from less than 41 percent to 69 percent. All the Nordic studies (23, 37, 38, 40) except ours had initial postal response rates of 60 percent or more. These studies were all based on mandatory universal population registries. Both Chinn et al. (34) and de Marco et al. (39) used a population-based register, but ones that were used for health care purposes, so quality may have been less than optimal.

All studies that reported on smoking found that smokers were late responders (table 4) (22, 23, 37–39, 41). The relation of intensity of follow-up with symptoms and diagnoses was not universal; one study showed no relation (22), others found an increased prevalence of symptoms with intensity of follow-up (23, 37, 38), and others found a decreased prevalence of symptoms with follow-up (36, 39). None of these studies adjusted for potential confounders. However, follow-up had little effect on the final prevalence.

Only two studies are known to have investigated exposure-disease associations by survey effort (22, 23). One found a difference in an exposure-disease relation that caused occupational exposure to be no longer significant in a stepwise regression (23), but both reported little effect on final exposure-disease relations (22, 23).

The total sample size was moderate to large in most of these studies, but the number of respondents in follow-ups was relatively small (table 4). For example, the mean number of respondents in the telephone responder group was 389. If the prevalence of a symptom was 10 percent in a

Am J Epidemiol 2003;157:558–566
group with 389 respondents, the 95 percent confidence interval would be 7.2 percent to 13.4 percent. If this prevalence was compared with the prevalence in a larger group of initial responders, this confidence interval might obscure an important difference. In our study, the impact of follow-up on confidence intervals was small. Smaller studies may find it necessary to include a follow-up on the basis of study power alone. Most etiologic studies in respiratory epidemiology require adjustment for confounding, and doing so increases the required sample size. This adjustment is especially important for continuous variables for which sparseness of data is likely to be a problem.

Drané (31) proposed a linear or exponential extrapolation of trends in prevalence by reminder letter (32), which is attractive in its simplicity. This method has seen some use (34, 36, 39) but no thorough validation. There is little theoretical or empirical justification for selection of a linear or exponential trend. A what-if scenario with plausible alternatives might be just as reasonable. One plausible alternative is that the latest responders are most “equal” to the nonresponders and that prevalences for the nonresponders should be “set” to the prevalence of these latest responders (5). We did not explore this alternative, since neither of these techniques is sophisticated enough to cater to the multivariate situation, for example, when exposure-disease estimates are estimated and are controlled for confounding (49).

A cautious conclusion is that, so far, no studies have found important nonresponse bias in respiratory epidemiology health surveys that used this method, in either prevalence estimates or exposure-disease associations. A number of studies in other fields of epidemiology have used the same “wave” or willingness-to-respond approach (6, 8–20). These studies cover a wide range of geographic areas, different populations, and a variety of subjects. Most have found trends in the prevalence of outcomes or risk factors by willingness to respond, but the effect on final estimates has been variable. Some have found strong differences (7, 18, 19), whereas most others have found only small differences (6, 8–17, 20, 22, 23). The effect on exposure-disease associations (e.g., odds ratios) has rarely been studied (20).

Surprisingly little is known regarding the consistent refusers, the residual nonresponders, which probably reflects the difficulty and cost involved in obtaining information on these persons. In our study, we had no information on the residual 20 percent of nonresponders, except for age, gender, and geographic location. In an extreme scenario, both prevalences and associations with risk factors could be quite different on a population level. If the trends observed for the telephone responder group persisted or increased, results on a population level might be somewhat different. Our own literature search (50–55) and various PubMed searches (National Library of Medicine, Bethesda, Maryland) revealed a single study (12) with a very high response rate (e.g., >90 percent). In a later commentary, Sheikh took a skeptical view of nonresponse (56), although he offered few solid references.

A limited amount of data is available from some studies that have compared nonresponders and responders regarding information external to the study. This use of external information to track nonresponse bias has shown that nonresponders are likely to under- and overreport previous illnesses (57, 58), have higher admission rates to the hospital (20, 59), and are more likely to have a history of psychiatric illness (60). They may also have higher mortality (20, 61–64) and differences in health care utilization (65)—possibly related to smoking (66). A number of mostly nonmedical studies are discussed by Wentland and Smith (51).

There is potentially a large nonresponse bias in most epidemiologic studies, or indeed in any survey. Outcomes are usually uncommon (e.g., <10 percent prevalence), and response rates are rarely above 90 percent. With a 90 percent response rate, a 5 percent prevalence among the responders may reflect a true prevalence of 4.5–14.5 percent in the whole population, given that the nonresponders have either a zero or a 100 percent prevalence of the outcome. This scenario may seem extreme but is probably the reason that response rates receive much attention, even given the fact that good examples of clear nonresponse bias are not abundant. In fact, nonresponse rates are sometimes used as the only indicator to judge the quality of a survey, which may be an oversimplification (67).

There are two possible interpretations of the present study. On the one hand, one may choose to use a larger sample size upfront and a less-intensive follow-up survey effort if it has already been established that nonresponse bias is not likely to be a large problem in the same population with the same survey topic. It would save time and money to use a large sample size and eliminate telephone follow-up to achieve a 50–60 percent response rate rather than to select a smaller sample size and aim for a response rate of 70–80 percent with follow-up. The larger survey would achieve about the same total number of responses and hence study power but in a shorter time and at a lower cost. In particular, the cellular telephone survey netted very few additional responses and required a substantial amount of time and work.

On the other hand, there is a large potential for nonresponse bias in all surveys, as noted above. A study may be greeted with skepticism if its response rate is as low as 43 percent, as it was in the present study. However, our study does show that pushing response rates from 60 percent to 80 percent would have been unlikely to materially affect conclusions since the additional reluctant responders were not very different from early responders. This finding is supported to some extent in smaller studies on other survey topics (6, 8, 9, 13, 15). If this finding can be replicated in other populations and for other survey topics, it could represent important savings of money and manpower.

In this study, the additional contribution of sending reminder letters and conducting a telephone follow-up to prevalence estimates and exposure-disease relations in respiratory epidemiology was small. In this population, a single letter would have been enough. The possibility that this finding can be generalized to other populations should be studied.

ACKNOWLEDGMENTS

This study was supported by the Norwegian Research Council.
REFERENCES