EUROPEAN
JOURNAL
OF
EPIDEMIOLOGY

PREDICTING RISK AMONG NON-RESPONDENTS
IN PROSPECTIVE STUDIES

K. SHEIKH

Department of Epidemiology - University of Michigan School of Public Health
109 Observatory Street - Ann Arbor, Michigan 48109 - U.S.A.

Potential non-response bias was investigated in a follow-up study of 2,011 chronically disabled patients. 82.5% and 73.3% of the study subjects responded to self-administered mail questionnaires respectively at 6-month and 1-year follow-up. Information on employment status, the outcome of interest, of approximately 90% of the non-respondents was obtained from indirect sources. Employment rate was lower among the non-respondents than the respondents. Non-response was associated with age, social class, previous employment record, and the type of disability; but none of these characteristics were associated with the outcome. Out of the five known independent risk factors for unemployment, only one (incompletion of rehabilitation course) was associated with non-response. The employment rate among the respondents was also assessed according to the delay in response, that is the number of reminders sent to achieve response. The outcome among the late respondents was similar to that among the non-respondents. These data suggest that (a) risk estimates may be biased even when the response rate is greater than 80%, (b) the prevalence of risk factors among non-respondents may not indicate the presence or the degree of non-response bias, but (c) reliable estimates can be obtained from extrapolations of the rates among the respondents according to the delay in response.

INTRODUCTION

Non-response in prospective studies may be described as (a) non-participation by a section of a study sample, (b) loss among participants due to death (for example, in morbidity or social studies) or emigration, and (c) non-response to follow-up among study subjects available for the follow-up. Any one or all of these sources of selection process may be operating in a prospective study. Irrespective of the type, if non-response is associated with the risk factors and the outcome of interest, risk estimates are very likely to be biased (9). In any case, an estimate of the incidence of outcome among non-respondents is necessary in order to assess the presence and the extent of

possible non-response bias. Theoretically, however, non-response bias may exist even when the risk factor prevalence and outcome incidence among the non-respondents are the same as among the respondents (4, 7).

Despite the preventive measures, some degree of non-response is frequently found in prospective studies. Ideally, a separate follow-up of a representative sample of non-respondents should be carried out to ascertain the outcome. Alternatively, the required information may be obtained from an indirect source, such as hospitals and registeries (9). Depending on the outcome of interest, these procedures may not be possible or feasible. If the study sample is a homogenous group, some investigators assume no difference in outcome

between respondents and non-respondents (1). Such assumptions are not uncommonly made in cross-sectional surveys achieving response rates as low as 49% (17). Estimation of the « maximum or minimum possible incidence » in non-respondents has also been professed (10). Accordingly, a range of rates for the study sample is calculated, one extreme on the assumpion that the non-respondents have had a favorable outcome and the other on the opposite assumption; the true rate being somewhere between the two extremes. Alternatively, it might be assumed that the non-respondents have an incidence rate double or half that of the respondents (11). Unless the non-response rate is low and the outcome is a frequent event, a range of risk estimates is of limited use (9). Besides, the interpretation of the relative risk will be difficult if there is a weak association between risk factor(s) and outcome, that is, if one extreme of the range of risk ratios is not significantly different from 1.

It may be possible to obtain information on demographic characteristics of the non-participants from other sources. A comparison of the participants and the non-participants with respect to age, sex, education and employment status has been used to rule out non-response bias (1). It is assumed that groups with similar characteristics will have similar risks. But, these characteristics may not be risk factors for the outcome of interest; and even if they are, the two groups with similar demographic characteristics may differ from each other with respect to other risk factors. However, if there is non-response to follow-up, it may be possible to estimate the incidence of outcome among the non-respondents from baseline prevalence of risk factors in the non-respondents.

The influence of the event under investigation on health or psychosocial survey response behavior has been demonstrated in many studies. It appears that non-respondents resemble more with late respondents in terms of the prevalence of the event or characteristic of interest, than with early respondents (2, 6, 8, 12). Although an estimate for non-respondents is based on guesswork, it is an informed guess (2). This procedure has not been utilized in estimating the incidence of outcome in non-respondents in prospective studies.

An opportunity to study response behavior and the characteristics of non-respondents occurred during a follow-up study of gainful employment in a cohort of chronically disabled subjects (14, 15). This report presents summary data to demonstrate that the prevalence of known risk factors may give misleading estimates of outcome in non-respondents, and that response behavior of respondents gives an indication of the probable direction of non-response bias, if it exists.

MATERIAL AND METHODS

The study population consisted of all economically active disabled patients with employment problems attributable to their disability discharged from three rehabilitation centres in Greater London, England, during the years 1973 to 1975 (two centres) and 1975 to 1976 (one centre). Details of demographic and social characteristics, and medical and employment histories were obtained from the case notes at the three centres. The results of functional and psychological assessments for each participant, and general unemployment rates in his/her home area at the time of discharge from the centre and during follow-up were also recorded. These data are routinely collected and recorded in a standard form at the centres (15).

Members of the original cohort who for any reason (death, emigration, or confinement to prison) were not at risk of unemployment in the United Kingdom were excluded from follow-up. Data on a group of 398 patients, who were attending a recommended full-time vocational training course during the follow-up period, were analyzed separately since they could not be counted as « unemployed ». Follow-up was conducted by self-administered mail questionnaires 6 months and 1 year after discharge. Subsets of the study population were followed for 2, 3, or 5 years. At each follow-up, the questionnaire inquired in confidence about employment, vocational training, and the reasons for unemployment where applicable. Up to four reminders were mailed to those who failed to return the completed questionnaires. Telephone contacts were also made with those available on the telephone. Information about the employment and/or vocational training status of the non-respondents was obtained from other sources; these included disablement resettlement officers, government employment offices, social workers, family physicians, hospitals, employers, relatives and friends.

The choice of study variables (risk factors for unemployment in disabled people) was based on our earlier observations and those of other investigators (15). The data presented here are restricted to the results of 6-month and 1-year follow-up of those at risk of unemployment (excluding trainees) during the observation period. The study outcome is defined as full-time gainful employment during the follow-up period. Further details of the study population, follow-up and analytical methods are given elsewhere (13-15). The χ^2 test was used to assess the significance of differences between proportions. The level of statistical significance of all associations described here was set at 5% or less.

RESULTS

The results of the 6-month and 1-year follow-up are given in Table 1. Outcome was not known for 29 (1.4%) patients at the 6-month follow-up and 57 (2.9%) patients at the 1-year follow-up. Hereafter, « non-respondents » will be defined as those subjects who did not respond to follow-up and information about their employment status was obtained from other sources. At the 6-month follow-up, 44.8% of the 1659 respondents and 34.4% of the 323 non-respondents had been employed. At the 1-year follow-up, 51.2% of the 1436 respondents and 43.8% of the 465 non-respondents had been employed. The differences between employment rates among the respondents and non-respondents are statically significant (p < 0.01).

TABLE 1.

Response to follow-up of chronically disabled patients discharged from three English rehabilitation centres in 1973-76.

	6-month follow-up		1-year follow-up	
	n.	%	n.	%
Total followed	2011	100.0	1958	100.0
Respondents	1659	82.5	1436	73.3
Non-respondents:				
Outcome known	323	16.1	465	23.8
Outcome unknown	29	1.4	57	2.9

The study variables are listed in Table 2. Non-response was positively associated with lower social class, poor past employment record, and certain types of disability; it was inversely associated with age, and the completion of a rehabilitation course. Response rates according to the level of each of these independent variables have been published elsewhere (13). For example, 84.3% of the respondents to the 1-year follow-up and 76.2% of the non-respondents had completed their rehabilitation course. In a multiple regression analysis of the data on the respondents and non-respondents, independent risk-factors for an adverse outcome were a low level of motivation for work, a high level of physical disability, long duration of unemployment before rehabilitation, failure to complete a rehabilitation course, and a high level of general unemployment in their home area. (Details of these analyses, regression coefficients for each variable, and multiple correlation coefficients are given in previous reports (14, 15)). Hence, only one variable, the completion of a rehabilitation course, was associated with both non-response and the outcome. The association, however, was rather weak. The r² for the

TABLE 2.

Independent variables considered in a follow-up study of return to work among chronically disabled patients discharged from three English rehabilitation centres in 1973-76.

Age *	Disability in climbing
Sex	Working at heights
Social class *	Duration of unemployment ●
Education	Premorbid work record
Qualifications	Employment record *
Intelligence	Income
Motivation ●	Mobility for employment
Compensation	Completion of course *●
Duration of disability	Premature discharge reasons
Disability in walking ●	Home area unemployment rate ●
Disability in lifting	Type of disability *

* Associated with response.
● Associated with outcome.

independent relationship between this risk factor and the outcome was only 0.02, as compared to r² of 0.16 for all the five risk factors (15).

At the 6-month follow-up, 73.8% of the 1659 respondents returned the first questionnaire mailed to them, the remaining 26.2% responded to reminders. The corresponding proportions at the 1-year follow-up were 66.9% and 33.1%. The employment status of the respondents according to the delay in response, i.e. the number of reminders mailed to achieve response, and of the non-respondents is given in Table 3. At each follow-up, the employment rate was highest among the respondents who responded without being reminded (early respondents), intermediate among the respondents who returned questionnaires after reminder(s) (late respondents), and lowest among the non-respondents.

TABLE 3.

Employment status according to response behaviour of chronically disabled patients discharged from three English rehabilitation centres in 1973-76 and followed for one year.

	Response after reminders			Non-respondents
	0	1	2-4	
At six months:				
No. of subjects	1225	334	100	323
Employed (%)	46.9	39.2	36.0	34.4
At one year:				
No. of subjects	961	394	81	465
Employed (%)	53.6	45.9	48.1	43.8

DISCUSSION

The results of this study provide some important methodological pointers. The outcome among the non-respondents was significantly different from that among the respondents even when the response rate was as high as 82.5% (Tables 1 and 3). Due to the unusual nature of the outcome of interest in this study, personal characteristics, such as age, sex, and social class, were not associated with the outcome (Table 2). An interesting finding was that the prevalence of four out of the five risk factors among the non-respondents was similar to that among the respondents (Table 2). The fifth risk factor, that was associated with non-response, explained a very small proportion of the variance in the outcome. Consequently, the rate of employment in the non-respondents predicted from the prevalence of risk factors (among the non-respondents) would have been grossly overestimated, and the magnitude of expected non-response bias would have been underestimated.

The lower employment rate among the non-respondents as compared to the respondents (Table 3) may to some extent be due to the effect of outcome on response behavior. It may also be that certain factors not considered in this study, that influence the risk of unemployment in the disabled, were more or less prevalent in the non-respondents than in the respondents. An alternative explanation may be that although the prevalence of known or unknown risk factors was similar among the two groups, the interaction between the risk factors and the outcome among the respondents was different from that among the non-respondents. This pattern of differential interaction may be analogous to the phenomenon of « the worried well » (4). A good example of « the worried well » were the participants of the Collaborative Lipid Research Program Study (5). In this study, a subset of non-participants were interviewed on the telephone. As compared to the participants, the prevalence of cardiovascular disease was higher among the non-participants, but the prevalence of a history of hyperlipidemia and family history of cardiovascular disease was lower.

There are several documented examples of the potential use of response behavior of respondents for estimating the prevalence of the event of interest among non-respondents. Cochrane (3) observed that among Welsh coal miners invited to a screening program, the prevalence of pulmonary tubercolosis was the lowest among those miners who participated early and the highest among late respondents. In a mail survey of men participating in the Normative Aging Study in the Veterans Administration Clinic in Boston (12), compared to late respondents, the prevalence of cigarette smoking was lower among early respondents and higher among non-respondents. A similar pattern was observed in a rather different context, a mortality follow-up of psychiatric patients. In this study, death rate among the group of patients traced with relative ease was lower than that among the group requiring intensive tracing (16).

The data presented in Table 3 support the viewpoint that late respondents are potential non-respondents, and that the outcome among non-respondents is very likely to be similar to that among late respondents. If this assumption is empirically tenable, should a direct extrapolation be made from potential non-respondents to actual non-respondents, or is a linear extrapolation dictated? A trend seen in Table 3 suggests that a linear extrapolation might be appropriate. Even if the precise magnitude of non-response bias can not be estimated from such extrapolations, an examination of response behavior should afford a clue to the presence and direction of non-response bias (6).

It should be realized that the objectives of this study were to assess the determinants of socio-economic consequences of chronic disability rather than the determinants of chronic disability. The study cohort were diseased or injured subjects with residual impairments, the outcome of interest was employment, mail questionnaires were used for follow-up, and the observation period was brief. Non-response rates were not very high. Consequently, the inclusion of the non-respondents made little difference to the overall employment rates (43.0% vs. 44.7% at 6 months, and 49.4% vs. 51.251.2% at 1 year), although the employment rates among the two groups were significantly different (Table 3). In studies with non-response rates higher than in this study or a greater difference in outcome among the respondents and the non-respondents, the inclusion of the non-respondents may substantially alter the overall rate. The findings of this study may not be replicable in other situations, but they should provide encouragement for exploiting prospectively collected data particularly in studies where response may be affected by the outcome.

In conclusion, risk estimates in a prospective study may be biased even when the response rate is greater than 80%. The presence of non-response bias or its magnitude can not be estimated with confidence from the associations between response and demographic characteristics, or from the prevalence of risk factors among the non-respondents. If data on the incidence of outcome in the non-respondents are available from indirect sources, the validity of extrapolations from rates among the respondents according to the delay in response should be tested under different conditions.

REFERENCES

1. *Barton J., Bain C., Hennekens CH., ,Rosner B. Belanger C., Roth A.* and *Speizer F.E.* (1980): Characteristics of respondents and non-respondents to a mailed questionnaire. - Am. J. Pub. Hlth., *70:* 823-825.

2. *Clausen J.A.* and *Ford R.N.* (1947): Controlling bias in mail questionnaires. - J. Am. Stat. Assoc., *42:* 497-511.

3. *Cochrane A.L.* (1951): Occupational and domiciliary factors in pulmonary tubercolosis. In: Medical Research Council. The application of scientific methods to industrial and service medicine. - HMSO, London.

4. *Criqui M.H.* (1979): Response bias and risk ratios in epidemiologic studies. - Am. J. Epidemiol., *109:* 394-399.

5. *Criqui M.H., Barrett-Connor E.* and *Austin M.* (1978): Differences between respondents and non-respondents in population-based cardiovascular disease study. - Am. J. Epidemiol., *108:* 367-372.

6. *Ellis R.A., Endo C.M.* and *Armer J.M.* (1970): The use of potential non-respondents for studying non-response bias. - Pacific. Sociol. Rev., *13:* 130-139.

7. *Greenland S.* (1977): Response and follow-up bias in cohort studies. - Am. J. Epidemiol., *106:* 184-187.

8. *Hill A.B.* (1951): General principles of field surveys. In: Medical Research Council. - The application of scientific methods to industrial and service medicine. - HMSO, London.

9. *MacMahon B.* and *Pugh T.F.* (1970): Epidemiology: Principles and methods. - Little, Brown and Co., Boston.

10. *Mausner J.S.* and *Kramer S.* (1985): Epidemiology: An introductory text. - W.B. Saunders Co., Philadelphia.

11. *Menotti A.* (1978): Incidence in non-respondents. In: Holland W.W. and Karhausen L. (eds.). Health care and epidemiology. - H. Kimpton Publishers., London.

12. *Seltzer C.C., Bosse R.* and *Garvey A.J.* (1975): Mail survey response by smoking status. - Am. J. Epidemiol., *100:* 453-457.

13. *Sheikh K.* and *Mattingly S.* (1981): Investigating non-response bias in mail surveys. - J. Epidemiol. Community Hlth., *35:* 293-296.

14. *Sheikh K.* and *Mattingly S.* (1982): Does the type of disabling impairment influence return to work. - Pub. Hlth., *96:* 238-242.

15. *Sheikh K.* and *Mattingly S.* (1984): Employment rehabilitation: outcome and prediction. - Am. J. Ind. Med., *5:* 383-393.

16. *Sims A.C.* (1973): Importance of a high tracing-rate in long-term medical follow-up studies. - Lancet, *2:* 433-435.

17. *Tagliacozzo R.* and *Vaughn S.* (1982): Stress and smoking in hospital nurses. - Am. J. Pub. Hlth., *72:* 441-448.