

# A confidence interval approach to investigating non-response bias and monitoring response to postal questionnaires

Alan Tennant, Elizabeth M Badley

## Abstract

**Study objective**—The aim was to develop an alternative method of investigating non-response bias in postal surveys, including a method of calculating a final full (100%) coverage confidence interval which avoids the wide intervals of existing approaches.

**Design and Setting**—As part of a two stage survey of disablement in the community, a first phase postal questionnaire was sent to 25 168 households in Calderdale, West Yorkshire, England. Confidence intervals were calculated to investigate the precision of estimates using a “no bias” model, where the prevalence in non-responders is assumed to be the same as in responders.

**Respondents**—A total of 21 889 postal questionnaires were returned (87%), representing households containing 42 826 people aged 16 years and over. This was achieved by the original post (1st wave, 57% response); two further postal follow ups (2nd and 3rd waves, taking the response to 73% and 81% respectively), the latter including a small personal call back; and a final postal follow up (the 4th wave).

**Results**—The cumulative estimated prevalence of those with dependence was plotted as the survey progressed. The final wave full coverage estimated prevalence for those aged 16–64 years was 12.8 per 1000 with 95% confidence intervals of 11.3–14.4 per 1000. The integrity of this estimate holds as long as the true prevalence in non-responders is within the calculated non-response confidence interval under the no bias assumption, 9.7–16.0 per 1000 people. This latter interval represents the tolerance of prevalence in non-responders implied by the no bias assumption.

**Conclusions**—The findings have general implications for monitoring non-response bias in postal screening questionnaires. The confidence interval approach developed in this paper offers an alternative to existing regression based estimates, giving an indication of the range of prevalence amongst non-responders that could be tolerated before the no bias assumption used by the model is breached. It is suggested that this approach can be used to determine both the extent of bias, and to aid decision making about the appropriate juncture to terminate follow up. It highlights the potential, particularly in the context of a computerised survey operation, of methodological investigation occurring simultaneously with survey operation.

An objective of many surveys of the population is to estimate the prevalence of a particular attribute. Incomplete response may introduce uncertainty as to the accuracy of the findings and raise questions as to whether it is desirable to extrapolate over the non-response element of the survey. This might be prompted by a visual inspection of the changing proportion of respondents with the attribute as the survey progresses. The resulting cumulative prevalence rate, that is the plot of estimated prevalence against the response rate, may rise or fall, suggesting a likely under- or overestimate of prevalence. Knowledge of the accuracy of the estimate may be particularly crucial if the survey has been undertaken to inform action, for example the planning of specific services.

This paper considers methods of estimation of final prevalence and suggests a confidence interval approach to monitoring response and the likelihood of bias in surveys.

## ESTIMATION OF FINAL PREVALENCE

A recent survey of disabled people<sup>1</sup> adjusted for non-response in a postal questionnaire using a method suggested by Hochstim.<sup>2</sup> This extrapolates from the end of survey prevalence, where the response is less than 100%, to a full coverage (ie, 100%) prevalence. Hochstim uses a regression approach based on the estimated prevalence of the target group at each stage of the survey, viz the initial response and each follow up enquiry. Thus the values entered into the regression equation reflect the cumulated estimated prevalence of the target group at each successive wave of the survey. The method is deemed appropriate where there is a clear linear trend across these successive waves.

Some criticisms can be levelled at this approach. (1) The assumption that there is a linear response across successive waves leaves the method vulnerable to “hard core” non-responders. If there is any reason to believe that they are quite different to responders, then prevalence in non-responders may differ sharply, and estimates based upon responders may be erroneous. However, Fillion’s<sup>3</sup> findings refute this, suggesting that in a survey with a response rate of about 70%, linear extrapolation would indicate the “presence, direction and extent of non-response bias”. (2) For linear regression there is the assumption that the value of each observation on the independent variable is independent of all values of all the others.<sup>4 5</sup> Cumulative prevalence data breach this assumption as their values are built up on the base of preceding values. (3) If an attempt is made to consider full coverage of a survey it would seem

ARC Epidemiology  
Research Unit,  
University of  
Manchester, Oxford  
Road, Manchester  
A Tennant  
E M Badley

Correspondence to:  
Mr Tennant, at The  
University of Leeds, School  
of Medicine, Rheumatology  
and Rehabilitation Research  
Unit, 36 Clarendon Road,  
Leeds LS2 9NZ, United  
Kingdom

Accepted for publication  
March 1990

highly desirable, particularly from a service planning point of view, to have confidence intervals for the "best estimate". Confidence intervals, however, are not so useful when used with extrapolated data as they widen rapidly after the known data; the lower the response rate and the greater the range of non-response over which to extrapolate the wider the confidence intervals for the full coverage estimate. This is a strong argument for maximising response rates, particularly as there is the risk of falling into the extrapolation trap,<sup>6,7</sup> where the greater the extrapolation, the greater the chance of ending up with an estimate which has no valid place in the current regression structure.

An alternative method of trend estimation which might be considered is the "exponentially weighted moving average".<sup>6,8</sup> This both smooths the data and gives greater weight to the most recent results. This method makes no assumptions about linearity, or independence of measures from one point to the next, although like the Hochstim method it has the disadvantage that confidence intervals widen after the end of the known data.

#### CONFIDENCE INTERVAL APPROACH TO NON-RESPONSE BIAS

Following recent emphasis upon the use of confidence intervals<sup>9-11</sup> we looked at an alternative approach to investigating non-response bias. Following the binomial distribution, in large samples when  $i$  members are found with a particular attribute, the proportion  $p'$  of the sample population possessing the attribute (sample size  $n$ ) is  $i/n$ , with a standard deviation  $\sqrt{(p'q'/n)}$  where  $q'$  is  $1-p'$ . The approximate 95% confidence interval for  $p$  (population proportion) is  $p' \pm 1.96[\sqrt{(p'q'/n)}]$ .<sup>7,12</sup> If  $np'$  is less than 15, the Poisson distribution, rather than the binomial distribution should be used.<sup>12</sup>

A problem with using this method is that  $n$  is greater than the actual number of respondents. Cochran<sup>12</sup> divides  $n$  into  $n(r)$  (responders) and  $n(n)$  (non-responders), so that at any wave of the survey  $n = n(r) + n(n)$ . The survey proportion  $p(r)$  and confidence intervals  $CI(r)$  that we are able to calculate relates only to  $n(r)$ . Another unknown proportion  $p(n)$  with attendant confidence intervals  $CI(n)$  represents non-responders. Modifications to the formulas for confidence intervals have been developed to take account of the effect of non-response<sup>7,13</sup> which in practice requires the addition of the non-response proportion to the upper confidence level. However in circumstances where the non-response proportion is large in relation to the magnitude of the prevalence this would lead to a very large upper confidence interval.

An alternative method would be to assume no bias so that the prevalence in non-responders is deemed to be the same as in responders. In this case  $p'(r) = p'(n) = p'$ . To test whether this no-bias assumption applies, confidence intervals can be calculated for the outstanding non-response group at a particular wave of the survey, assuming  $p'(r) = p'(n)$ , and comparing these with the actual prevalence obtained in the succeeding wave of the survey when these subjects are sent a reminder

questionnaire. If the prevalence, or more specifically its confidence intervals, overlaps the confidence intervals calculated for the former non-response group, then this indicates that they both belong to the same population. Succeeding follow ups giving similar no bias results would suggest that bias in the ultimate non-responders would be unlikely. Alternatively it would show the magnitude of prevalence that would have to be found in remaining non-responders, if the no bias assumption were to be breached.

#### Methods

A survey was carried out to help plan the provision of services for those in the population of Calderdale in West Yorkshire, England, who were very severely physically disabled and likely to be on the margins of institutional care.<sup>14</sup> A primary focus of the survey was the so called younger physically disabled population, those aged 16-64 years, although the survey was extended to look at those aged 65 years and older who were disabled by rheumatic conditions. A two stage methodology was employed, similar to that used in other surveys of disablement in the population.<sup>15</sup> The first stage consisted of a postal questionnaire to every third household in the area, using the domestic rating list as the sampling frame. The sample size was determined by the need to provide meaningful analysis for a rare group in the population, namely those very severely physically disabled aged 16-64 years, based on an estimate by Harris of 12 per 10 000 people.<sup>16</sup> The second stage was an in depth personal interview with a stratified random sample of those ascertained to be disabled in phase 1 of the survey. This paper is concerned with response in the phase 1 screening survey.

The four page screening questionnaire used in phase 1 elicited demographic characteristics of the household, and information about individuals within the household who experienced difficulties or dependence in various activities of daily living. As the main target of the survey was those who were very severely disabled and who might be expected to have high dependency needs, the principal entry criteria for the second wave of the survey was set at the level of being dependent on others. This meant requiring help to get in or out of bed, with dressing, to get to and use the toilet, or to get out of the house. This latter condition sets a relatively low threshold for entry and is consistent with an approach to screening for rare groups in populations which minimises false negatives and produces a manageable group for further investigation.<sup>17,18</sup> This paper concerns the ascertainment of this dependent group in the population, and considers non-response in terms of non-return of the first phase questionnaires.

Questionnaires were sent to 25 168 occupied dwellings in Calderdale and 21 889 were returned, a response rate of 87%. This was achieved by the original post (1st wave, 57% response); two further postal follow ups (2nd and 3rd waves, taking the response to 73% and 81% respectively), the latter including a small personal call back; and a final postal follow up (the 4th wave). The final postal follow up was encouraged by findings from the personal call back which

indicated that one quarter of respondents still outstanding would return another questionnaire which was left with them or put through their door.

Table I shows the number of people responding within each major age group at each stage of the study. The demographic data from households responding at phase 1 was used as the denominator to calculate the rate per thousand dependent people in each age group. The cumulative rate was plotted in order to monitor trends in ascertained prevalence as the survey progressed, together with an exponentially weighted moving average.<sup>6,8</sup> This method of trend estimation smooths the data and gives greater weight to the most recent results. The same method was used to extrapolate to a final (100% coverage) prevalence estimate.

**Results**

Figure 1 shows the cumulative prevalence as the survey progressed of those with dependence aged 16–64 years. In the latter stages of the survey there is a generally falling prevalence, and this raises the question of extrapolation to get a full (100%)

Table I Numbers of adults, by age group, in responding households at each wave of the phase 1 survey

Wave of response	Age (years)					Total
	16–34	35–54	55–64	65–74	75+	
1	8941	8564	4175	4091	2720	28 491
2	2407	2097	924	749	562	6739
3	2195	1701	606	389	267	5158
4	1037	813	262	212	127	2451
Total	14 580	13 175	5967	5441	3676	42 839

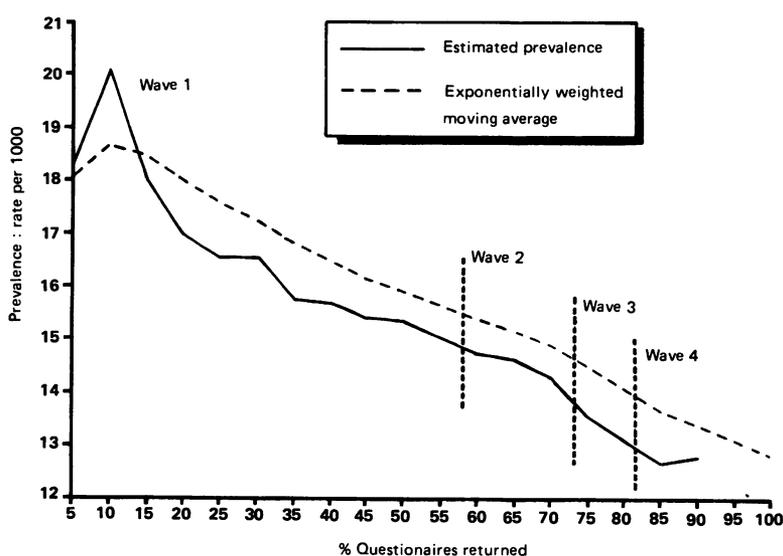


Figure 1 Cumulative rate of dependence. Respondents aged 16–64 years

Table II Survey prevalence estimates with 95% confidence intervals for dependent 16–64 year olds: rates per 1000

Wave	Wave prevalence rate	LCI	UCI	Cumulative prevalence rate	Non-response confidence interval		Full coverage confidence interval	
					Lower	Upper	Lower	Upper
1	14.9	13.2	16.5	14.9	13.0	16.7	12.9	16.9
2	10.7	8.0	13.4	14.0	11.8	16.2	12.2	15.9
3	6.2	3.9	8.5	12.9	10.3	15.6	11.3	14.5
4	11.8	7.2	16.4	12.8	9.7	16.0	11.3	14.4

LCI, UCI = lower, upper confidence interval

coverage prevalence. Using Hochstim's<sup>2</sup> method the prevalence of dependent 16–64 year olds is reduced from 12.8 per 1000 persons (at 87%) (table II) to 11.7 per 1000 persons (at 100%). For planning services, when such estimates are likely to be applied to over 100 000 people, this would make a difference of 110 per 100 000 to the estimate of the number of dependent people.

In contrast, extrapolation with an exponentially weighted moving average, shown by the broken line in the figure, gives an end of survey prevalence of 13.3, marginally above the end of survey prevalence.

Table II gives survey prevalence estimates for the four waves of the survey with confidence intervals for dependent 16–64 year olds, calculated using the no bias assumption. At wave 1 a prevalence of 14.9 per 1000 was found with confidence intervals 13.2–16.5. Assuming no bias, prevalence in non-responders gives confidence intervals of 13.0 to 16.7 per 1000. In order to calculate confidence intervals for full coverage the range of "outcomes" possible among non-responders under the no bias condition can be added to the existing survey findings. For example, the lower confidence interval is adjusted for non-response by allowing for 43% (amount of non-response) of the survey at a prevalence of 13.0 per 1000, which gives a lower confidence for the full coverage prevalence. Repeating the exercise, 43% is added at a prevalence of 16.7 per 1000 to give the upper full coverage prevalence confidence interval. This method to determine lower and upper bounds is similar to the principle expounded by Cochran,<sup>12</sup> and has been used in other contexts, for example, in actuarial life table methods<sup>19</sup> and more specifically, to adjust boundaries of confidence intervals when estimating the morbidity risk of illness from survey data.<sup>20</sup> Confidence intervals determined in this way are deemed to be conservative<sup>20</sup> but avoid the obvious disadvantage of extreme width.

The full coverage prevalence estimate at wave 1 is thus 14.9 per 1000, with confidence intervals of 12.9–16.9 per 1000. The integrity of these data will hold as long as the true prevalence within the non-response group lies within the stated non-response confidence intervals. This can be tested by looking at the results from the second wave, as the potential responders for wave 2 of the survey are the non-responders to wave 1. The second wave prevalence, 10.7 per 1000, falls outside the non-response interval at wave 1, but its confidence intervals overlap, indicating that second wave responders came from the "same" population as first wave non-responders, assuming no bias at wave 1, despite a sharply falling wave prevalence. The third wave response continued the decline in prevalence, so much so that its confidence intervals were no longer

consistent with a no bias assumption for non-responders at the second wave. This situation was rectified at wave 4 such that confidence intervals overlapped those for the previous wave's non-response group under the no bias assumption. The final wave full coverage prevalence (assuming no bias) is 12.8 per 1000 with 95% confidence intervals of 11.3–14.4 per 1000. The integrity of these data will not be breached unless the true prevalence in non-responders is outside the range 9.7–16.0 per 1000 people, the calculated non-response confidence interval. This is the tolerance of prevalence in non-responders implied by the no bias assumption.

The second major age group considered in the survey was those aged 65 years and older. Extrapolation using the exponentially weighted moving average (fig 2) suggests little difference between the end of survey prevalence (131.3/1000), and the estimated full coverage prevalence (132.1/1000). Hochstim's method is inappropriate here as there is no linear trend of cumulative prevalence across successive waves.

Table III shows the confidence interval data for those in this age group. Here the overall prevalence estimates are higher, consistent with an increased frequency of disablement in the elderly.<sup>21</sup> Follow up wave prevalences are consistent with no bias from the outset. The wave 2 prevalence of 114.4 is lower than the lower non-response confidence interval at wave 1, but its confidence intervals overlap, signifying that they come from the same population. The wave 3 prevalence confidence intervals also overlap the non-response confidence intervals for wave 2 non-responders, and likewise the final wave overlaps wave 3 intervals. The final wave full

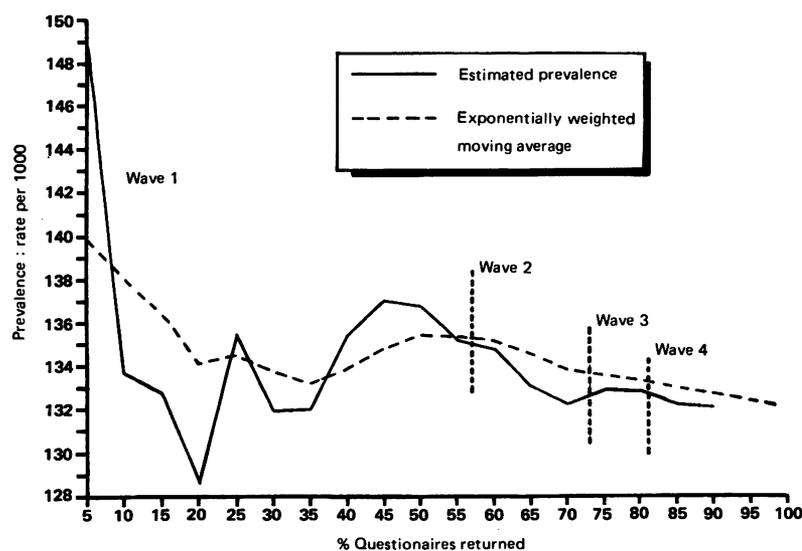


Figure 2 Cumulative rate of dependence. Respondents aged 65+ years

Table III Survey prevalence estimates with 95% confidence intervals for dependent 65+ year olds: rates per 1000

Wave	Wave prevalence rate	LCI	UCI	Cumulative prevalence rate	Non-response confidence interval		Full coverage confidence interval	
					Lower	Upper	Lower	Upper
1	135.1	127.0	143.2	135.1	125.6	144.6	125.0	145.3
2	114.4	97.2	131.6	131.8	120.1	143.4	122.3	141.3
3	137.8	111.2	164.3	132.2	117.1	147.3	123.1	141.4
4	123.9	88.8	159.0	131.9	113.9	149.8	123.1	140.8

LCI, UCI = lower, upper confidence interval

coverage prevalence (assuming no bias) is 131.9 per 1000 with 95% confidence intervals of 123.1–140.8 per 1000. The tolerance of non-response is a prevalence ranging from 113.9 to 149.8 per 1000.

## Discussion

For the 16–64 year old dependent population, use of Hochstim's method and extrapolation using an exponentially weighted moving average gives different results. Without any indication of the magnitude of likely errors it is difficult to choose between the two, and in any case both methods of extrapolation have the disadvantage of widening confidence intervals. The use of the confidence interval approach gives end of survey estimates based on an assumption of no bias, together with useful confidence intervals. These intervals provide an indication of the tolerance of prevalence in non-responders under the no bias assumption. Using this method, the decision that the end of survey prevalence for those aged 16–64 years was substantially free of bias is not entirely clear cut, as not all waves indicated a no bias situation. The third wave prevalence and its confidence intervals suggested the presence of bias at that stage (table II). In the event we feel that the evidence from the confidence interval calculations, taken together with smoothed trend which gives greater weight to the final wave prevalence, supports the no bias decision. Even with 13% of the sample outstanding, a broad range of non-response prevalence (9.7–16.0/1000) could be tolerated before the estimated range, expressed by the full coverage confidence interval, would be breached. Population estimates would only be affected by the order of 40 per 100 000 people, should prevalence in non-responders be at the lower or upper limits of their confidence interval.

In the 65 year and older age group, all follow up wave confidence intervals overlapped the previous wave's non-response no bias confidence intervals (table III). Therefore the evidence suggests that bias is unlikely to be found in remaining non-responders, particularly when consideration is given to the increasing tolerance of non-response prevalence explicit in the model. This, taken with the plot of the cumulative prevalence rate (fig 2) and its associated smoothed trend which suggests little difference between the end of survey prevalence and the estimated full coverage prevalence, led us to decide that the data for those over 65 years were free from non-response bias.

The principal weakness of using the confidence interval approach lies in its vulnerability to measurement error. It can be shown that in practice with large samples and modest rates of measurement error the a priori probability that the conventional confidence interval will include

the true proportion of positive cases could be small.<sup>22</sup> This indicates the crucial nature of ensuring that measurement error is small, or, if there is doubt about this, to use the confidence interval approach as one of several inputs into making decisions about non-response bias.

Wherever possible failure to conform to the no bias assumption should be met with further attempts to increase the response rate. Should bias still be evident, and it is not possible to improve the response rate further, some form of extrapolation might be in order. Decisions about the nature of any adjustment to be made where bias is shown to exist require a close examination of the data. Tests for linearity amongst proportions (not cumulative) are available,<sup>23</sup> and in addition to the type of smoothing we have used, other packages such as GLIM<sup>24</sup> offer a variety of ways to investigate the data for the most appropriate trend. The different estimates we have shown suggest that choice would be best made from some a priori theoretical position, as we did with the choice of the exponentially weighted moving average. In this way we allied ourselves to evidence that late responders are more similar to non-responders.<sup>25 26</sup>

Often methodological investigations, if they are done at all, are undertaken retrospectively. However, we have become aware of the potential, particularly in the context of a computerised survey operation, of methodological investigation occurring simultaneously with survey operation. A confidence interval approach to evaluating non-response offers potential both for continual monitoring, and for helping to determine if bias exists at the completion of field work. A computerised survey management system, dealing with returns and follow ups, could easily monitor the relevant response patterns, and is particularly well suited to microcomputer spread sheet operation.

In practice where samples are very large this might focus upon weekly returns,<sup>27</sup> or successive proportions of the target population. From a practical perspective it is at the end of each postal wave or personal call back wave that major decisions are taken. However, monitoring returns within waves can function as an early warning system to indicate, for example, that resources should be mobilised for an additional follow up.

When this approach is being used for survey monitoring, early detection of a no bias situation also raises the question as to whether the survey can be terminated sooner. For instance, in the case of the 65+ year age group (table III) a no bias situation was indicated from the early waves. If this had been the primary group of interest in the survey, and using this approach as a decision making tool, the survey could have been terminated after wave 2 (with a final response rate of 73%) or after wave 3 (with an 81% final response rate). Such decisions are a matter of fine

judgement and will depend both upon the objective of the survey, and on a combination of favourable indicators such as we have illustrated above. Early conclusion to the survey may be appropriate where the objectives of the survey are to obtain a prevalence estimate or where the attribute of interest is not uncommon. However, if the prime objective is case ascertainment, particularly of a rare group, a high response rate will be an end in itself.

We believe that applying this method in routine survey practice will enhance decision making, both with regard to the likelihood of bias where it offers a useful adjunct to visual inspection of trends, and possibly in determining an appropriate end point for the survey.

- Locker D, Wiggins R, Sittampalam Y, Patrick DL. Estimating the prevalence of disability in the community: the influence of sample design and response bias. *J Epidemiol Community Health* 1981; 35: 208-12.
- Hochstim JR. A critical comparison of three strategies of collecting data from households. *J Am Stat Assoc* 1967; 62: 976-89.
- Fillion FL. Estimating bias due to nonresponse in mail surveys. *Public Opinion Q* 1976; 39: 482-92.
- Poole J, O'Farrel M. The assumptions of the linear regression model. *Trans Inst Br Geographers* 1971; 52: 145-58.
- Durbin J. Estimation of parameters in time-series regression models. *J R Stat Soc: B* 1960; 22: 139-53.
- Yeomans KA. *Statistics for the social scientist: 2 Applied statistics*. London: Penguin, 1968.
- Armitage P. *Statistical methods in medical research*. Oxford: Blackwell, 1971.
- Gregg JV, Hossell CH, Richardson JT. *Mathematical trend curves: an aid to forecasting*. (ICI Monograph No. 1.) Edinburgh: Oliver and Boyd, 1964.
- Bulpitt CJ. Confidence intervals. *Lancet* 1987; 8531: 494-497.
- Gardner MJ, Altman DG. Confidence intervals rather than p values: estimation rather than hypothesis testing. *Br Med J* 1986; 292: 746-50.
- Gordon I. Sample size estimation in occupational mortality studies with use of confidence interval theory. *Am J Epidemiol* 1987; 125: 158-62.
- Cochran WG. *Sampling techniques*, 2nd ed. New York: Wiley, 1963.
- Konijn HS. *Statistical theory of sample survey design and analysis*. Amsterdam: North-Holland, 1973.
- Badley EM, Tennant A. *Calderdale health and disablement survey*. Manchester: Arthritis and Rheumatism Council for Research, 1988.
- Knight R, Warren M. *Physically disabled people living at home: a study of numbers and needs*. London: HMSO, 1978.
- Harris AI, Head E. *Sample surveys in local authority areas*. London: OPCS, 1971.
- Kalton G, Anderson DW. Sampling rare populations. *J R Stat Soc: A* 1986; 149: 65-82.
- Deming WE. An essay on screening, or on two-phase sampling, applied to surveys of a community. *Int Stat Rev* 1977; 45: 29-37.
- Lawless JF. *Statistical models and methods for lifetime data*. New York: John Wiley, 1982.
- Newman SC, Bland RC. Estimating the morbidity risk of illness from survey data. *Am J Epidemiol* 1989; 129: 430-8.
- Martin J, White A. *The prevalence of disability among adults*. (OPCS surveys of disability in Great Britain Report 1.) London: HMSO, 1988.
- Schwartz JE. The neglected problem of measurement error in categorical data. *Sociol Methods Res* 1985; 13: 435-66.
- Fleiss JL. *Statistical methods for rates and proportions*. New York: John Wiley, 1973.
- Baker RJ, Nelder JA. *The GLIM system, Release 3*. Oxford: Numerical Algorithms Group, 1978.
- Edgerton HA, Britt, SH, Normal RD. Objective differences among various types of respondents to a mail questionnaire. *Am Sociol Rev* 1947; 12: 435-44.
- Korelitz J, Visscher B. A method for examining the direction and amount of bias due to survey nonresponse (abstract). *Am J Epidemiol* 1988; 128: 945.
- Kemsley WFF. Some technical aspects of a postal survey into professional earnings. *Appl Statistics* 1962; 11: 93-105.