WORLD HEALTH ORGANIZATION PUBLICATIONS

Implementation of the Global Strategy for Health for All by the Year 2000

Second Evaluation

Eighth Report on the World Health Situation

This seven-volume work provides a detailed assessment, at global, regional, and country levels, of the extent to which strategies to develop health systems based on primary health care are being successfully implemented. The assessment, which also constitutes the Eighth Report on the World Health Situation, draws upon findings from national evaluations and data submitted by 151 countries covering 5200 million people and representing 96% of the world population, thus providing an unprecedented basis for a frank appraisal of health conditions and the factors that influence them.

Emphasis is placed on changes in the world health situation as measured through data, covering the period 1985–1990, on several well-defined indicators of health status and progress in reaching the social goal of health for all. Noting that the health for all strategy is essentially a quest for social justice and equity, the report gives particular attention to changes in the health status of vulnerable or disadvantaged groups and to the plight of populations in the least developed countries. Throughout, an effort is made to identify the lines of action needed to move forward in a world characterized by continuing population growth, increasing health problems, and shrinking funds for health and development.

The report consists of seven volumes: a global overview (volume 1) followed by individual reports from each of WHO's six regions. All regional reports were prepared according to a common outline. Each begins with an overview of socioeconomic developments, changes in the health system, and changes in health status as indicated by statistics on morbidity and mortality. An evaluation of overall achievements, particularly concerning implementation of the strategy for health for all, is followed by an analysis of problems likely to dominate the future and the actions that might be taken. Each regional volume also features a series of richly detailed profiles on health conditions for each country.

Volume 1: Global Review

This first volume provides a global overview of changes in the world health situation as determined through an analysis of data submitted by 151 countries for 1985–1990. Focused on a number of well-defined indicators of health status and its socioeconomic and environmental determinants, the report aims to discern trends, measure progress, define problems, and thus guide countries in their continuing efforts to strengthen health systems and improve the accessibility and quality of care. Emphasis is placed on factors linked to progress in the achievement of coverage by primary health care, equity in health, and sustainability in the national approaches employed.

The book has eight chapters. The first evaluates global political, economic, demographic and social trends and considers their health implications, including several consequences linked to the continuing global recession. The second chapter, devoted to health systems, evaluates the success of national efforts to achieve universal access to essential health care. Factors that have slowed progress are identified together with other factors consistently linked to the successful development of health systems. Subsequent chapters review progress in health care coverage and assess changes in financial resources, human resources, and health technology. Expanded immunization is identified as the greatest public health success story of the past decade.

A chapter devoted to health status uses data on mortality, morbidity, and disability to discern trends, identify the leading causes of death and ill health, and profile the major determinants of health. Environmental policies, hazards and risks are covered in the next chapter, which concludes that national capabilities for controlling environmental pollution have improved very little since 1984. The remaining chapters analyse the main achievements in terms of health care coverage and changes in health status, discuss major trends in health, and consider how these trends may develop. The final chapter identifies issues that need to be addressed in order to accelerate progress towards health-for-all goals.

Implementation of the Global Strategy for Health for All by the Year 2000:
Second Evaluation
Eighth Report on the World Health Situation
Volume 1: Global Review

1993, vi + 183 pages
ISBN 92 4 160281 3
Sw.fr. 35.-/US $31.50
In developing countries: Sw.fr. 24.50
Order no. 1231008

WHO • Distribution and Sales • 1211 Geneva

Assessing response reliability of health interview surveys using reinterviews

S.J. Fabricant[1] & T. Harpham[2]

Data from interview surveys of households or health facilities are used to assess community parameters such as health status and factors related to the ability and willingness of individuals to pay for health services. Although the effect of sample size on confidence intervals is generally well understood by the survey designers and policy-makers who use the results, the typical survey is also subject to non-sampling errors whose magnitude may exceed that of the sampling errors. The non-sampling errors associated with surveys are only rarely assessed and reported, even though they may have a major effect on the interpretation of findings. The present study reports the non-sampling errors associated with a household survey in Sierra Leone by comparing the results of reinterviews with the responses given during the original interviews. Certain types of questions were subject to greater non-sampling errors than others. The findings should be of use to designers of similar surveys and to those who rely on such surveys for making policy decisions.

Introduction

Health interview surveys

Household surveys have become a valuable tool for health planners and policy-makers in the search for better ways of managing and financing the activities of the health sector. Especially in developing countries, where official health statistics may be facility-based and data from censuses or general population surveys are likely to be incomplete or out of date, the information required to make sound policy decisions is best obtained using specialized surveys. If, for example, a comparison of the users and non-users of health facilities is needed, it is usually necessary to sample the general population using household interviews.

Sources of error and the need for quality control

There are two basic types of errors associated with sample surveys: *sampling errors*, which arise because a selected sample is used to obtain information about an entire population; and *non-sampling errors*, which arise because of variations between the response given by interviewees and the true answer. Sampling errors can be estimated *a priori* and can be reduced by increasing the size of the sample and by designing the survey to ensure that the sample bias is minimized. Non-sampling errors can be defined as all errors other than sampling errors, and are caused by biases introduced at any stage of the survey process—design, questionnaire development, data collection and analysis, and reporting of the findings—that affect the validity or reliability of the results. The causes of some non-sampling errors, such as mistakes in recording respondent's answers or in data processing, can be minimized by exercising care and using appropriate quality control procedures, while others, such as the respondent's inaccurate recall or wish to conceal the truth, may be extremely difficult to avoid.

The magnitude of both sampling and non-sampling errors is of great importance when reporting the results of the survey. For example, the level of significance of an observed difference between two groups is related to the errors associated with both groups. Usually, computers are used to analyse data from surveys and the results obtained are those determined by the software package to be statistically significant. The total error of a sample variable is the geometric sum of the sampling and non-sampling errors, but the latter are not usually taken into account by the software. Studies based on widely used guidelines for survey design that claim, for example, a 95% confidence interval (CI) of ±10% if a particular calculated sample size is used, are therefore misleading; such "rules" ignore the effect of non-sampling errors. In health economics, some of

[1] Health Policy Unit, London School of Hygiene and Tropical Medicine, London, England. Requests for reprints should be sent to Dr Fabricant at the following address: 2026 Ashland Mine Road, Ashland, OR 97520, USA.

[2] Senior Lecturer, Health Policy Unit, London School of Hygiene and Tropical Medicine, London, England.

Reprint No. 5392

the seminal reports on user fees and the price elasticity of demand for treatment are based on household survey data of unspecified validity. The reporting of results based on sophisticated econometric methods may have masked deficiencies in data quality.

The importance of non-sampling error is mentioned in a classic monograph on survey methods (*1*), where the lack of attention paid to non-sampling errors by researchers using household surveys is highlighted. Another standard text on surveys in developing countries (*2*) suggests that as a minimum the quality of survey results should be reported using the following rough grading:

— A) The basic data are reliable (accurate, objective methods were used): errors are mainly due to sampling.

— B) Some of these estimates rely on the respondent's recall: non-sampling errors may be as important as sampling errors.

— C) The respondents were reluctant to answer the question on which this item depends: non-sampling error is substantially greater than sampling error.

Costly efforts to reduce sampling errors by increasing the sample size may reduce the total error only slightly if the non-sampling errors are large, and indeed can increase it if use of a larger sample results in a relaxation of other standards of quality.

Although non-sampling errors are routinely reported for national censuses, smaller-scale surveys often ignore such errors. For example, one review of health interview studies in developing countries cited only a few studies that made any validity check at all, although important variations in reporting were measured in the two morbidity surveys that used reinterviews (*3*). Recall bias in case–control studies has been reported in only one study of reproductive hazards (*4*).

In contrast to the sampling errors, which can be calculated, non-sampling errors can only be estimated empirically. Several methods have been used to check the validity and repeatability of health interview surveys. These include the following: internal consistency checks, comparison with "expected" patterns, reinterviews, medical examinations, cross-checks against medical records, repetition of the study using a different sample, and observational studies. Moser & Kalton suggest that post-enumeration surveys (PES) consisting of reinterviews should become standard practice. The PES of the 1981 United Kingdom national census showed that true non-sampling error rates are hard to predict and are frequently underestimated in the absence of a PES (*5, 6*). One problem with large-scale PESs is the inevitable delay between the original survey and the repeat survey; this leads to a confounding of true error with recall error arising from informants' inability to remember the situation at the time of the original survey. Also, reinterviews have been criticized as a validation technique (*7*) because the result of the second interview will be unpredictably influenced by that of the first, with the degree of influence likely to increase the shorter the interval between the two interviews.

Despite these limitations, reinterviews are a practical and useful method of assessing non-sampling error in household surveys. In a recent study consistency checks were made on a questionnaire used first in a clinical setting and then repeated in households about a week later (*8*). This case–control study of the health impact of improved water supply and hygienic practices examined responses to the maternal education level and use of water source. For both these variables, there was disagreement between the data collected in the clinics and the households. For example, for maternal education the mean duration of maternal schooling reported in the clinic interviews and the household interviews differed by 0.4 years, which was statistically significant for comparisons of cases and controls. There was also a high rate of misclassification (23%) for the variable "type of water source used by the household", but it was concluded that the differences found were most probably due to a real change in water sources that had occurred between the two interviews; this highlights the importance of minimizing the interval between interview and reinterview.

In one study of the variables used to examine the association between other sexually transmitted diseases and infection with human immunodeficiency virus (HIV), the effect of misclassification of the variable "number of sexual partners in the last year" was estimated (*9*). If, for example, 50% of the subjects who had had five or more sexual partners in the last year claimed that they had had fewer than this number, and 5% of those who had had less than five partners claimed they had had five or more, the authors showed that, because sexual activity was measured imprecisely, there was an apparent association between syphilis and HIV infection when in reality there was none.

Designers of surveys that attempt to obtain new types of information, especially in unfamiliar cultural settings, would benefit from knowing the reliability of certain types of questions and the magnitude of the non-sampling errors that should be expected. This article examines these needs in the context of a recent health utilization and expenditure survey carried out in Sierra Leone.

Materials and methods

In Sierra Leone the Ministry of Health has been introducing user fees for primary health care since 1982, and in 1989 carried out an operational research study (*10, 11*) to obtain information on the effects of payment on the equity and utilization of services. The study included a survey of 1156 households in two rural districts, using a questionnaire consisting of modules on demographics, illnesses over a 2-week recall period and actions taken in response to them, household socioeconomic indicators, and questions on equity and the seasonality of household finances.

Questionnaire design and survey technique were refined and tested in order to minimize potential sources of error. No hypothetical questions were asked, e.g., where a person would go for care if they were sick or how much they would be willing to spend. A list of lay terms for common symptoms was read to respondents to assist them in remembering any illnesses during the 2-week recall period (*12*). The questions were pretested and modified to minimize confusion, and experienced interviewers were used who had received specific training on the questionnaires and spoke the local languages. Nevertherless, it was considered possible that the answers given by respondents (mainly illiterate subsistence farmers who had relatively little contact with health services or "modern" concepts of health and illness) might not be completely reliable. Responses to questions such as illness of household members, treatments used, and amounts spent depended on the subjects' recall of past events; others, such as household membership, might be subjective, while still others, such as the quantity of crops sown or harvested, and the possession of livestock, might elicit deliberately false answers.

Quality control was addressed by checking questionnaires for completeness and internal consistency while the interviewers were still in the village, and by rectifying any discrepancies by sending the original interviewer back to the household. After corrections had been made, reinterviews were conducted in 15% of the households originally surveyed. The original questionnaires were selected at random and kept aside, and different interviewers returned to the selected houses. The interval between the original interview and reinterview varied from less than 1 hour to overnight. The respondents were not told whether their answer differed from the one they had given in the original interview and no explanation was sought by the interviewer.

In terms of attributes relevant to the study, the characteristics of the reinterviewed households were similar to those of the population interviewed originally. In about 25% of the revisited households the original respondent (usually the head of the household) was unavailable and hence a different household member was questioned at the reinterview. In such instances analysis of the differences between the replies given in the original interview and the reinterview also provided information about the reliability of using respondents other than the heads of households and mothers.

Reporting non-sampling errors

The estimation of response reliability from reinterview results has received relatively little attention. For this purpose, a reliable response is taken to be one that is repeated when the respondent is asked the same question during the reinterview. A reliable response may not necessarily be a valid one, since validity implies that the question was interpreted by the respondent to mean what the interviewer intended and was answered truthfully; however, if a response is repeatable, a major type of non-sampling error can be assumed to be absent. In some studies, including large-scale PESs, non-sampling errors have been reported in a way that provides detailed information about the errors associated with each response category for a question, defined as described below.

• *Gross error rate* is the proportion of responses in a given category for which a response different from the original one was given at the reinterview. This is a measure of the reliability of individual responses to a particular question and also of the level of random error that would be undetected by a gross comparison of the original interview and reinterview data. Errors that occur randomly tend to weaken or mask true relationships by introducing anomalies into the observed results.

• *Net bias* is the difference between the proportion of answers to a given question in a particular category in the original interview and in the reinterview. This estimate of error is generally smaller than that indicated by the gross error rate because erroneous responses, to some extent, cancel each other out because they fall at random into different categories. The net bias reveals systematic errors that tend to distort true relationships, and can therefore be considered to be a measure of the reliability of the question for the entire sample, i.e., the degree to which a response could be expected to vary between interviews. The sum of the net biases for all the response categories of a given question equals zero.

• *Relative net bias*, which is derived from net bias, is the net bias for a given category expressed as a percentage of the proportion of responses in the reinterview in that category. The usefulness of the relative net bias is doubtful, since it is always biased upwards for categories with fewer responses.

Table 1 shows examples of these three measures of errors from the 1981 PES of the United Kingdom census.

Where appropriate, we used this scheme to estimate the reliability of interview questions and response categories from the Sierra Leone survey. While the error parameters described above can measure the reliability of discrete response categories within questions, they are inappropriate for continuous variables such as age and expenditures. Additional parameters were therefore used to analyse errors in the continuous variables, including the percentage of individual responses that did not match exactly and the mean difference between pairs of values and the absolute value of the mean. These provide information about both the magnitude and frequency of the error.

Table 1: **Selected results of data errors and biases from a post-enumeration survey of the 1981 United Kingdom census**[a]

Response category	Gross error (%)	Net bias (%)	Relative net bias (%)
Shared use of a bath or shower	7	-0.5	-29.1
Marital status: divorced	3	-0.1	-1.3
Travels to work by car pool	68	2.4	142.7
Travels to work by bus	9	0.9	5.9

[a] Adapted from ref. *6*.

Reliability of survey variables

As an example, the reliability of the variable "age of ill household member" was analysed by comparing the original interview and the reinterview data. The results of the analysis, both by age category and as a continuous variable, are shown in Table 2 and 3, respectively, which reveal some of the weaknesses of the conventional reporting method for continuous variables. Gross errors and net biases, as calculated by age ranges (Table 2), are less stringent than the criteria of exact match and mean absolute error as a percentage of age by single year interval (Table 3).

A summary of similar analyses for other variables used in the survey is shown in Table 4. In general, the means, medians, and distribution of reinterview responses closely matched the original responses; nevertheless, the individual case error was high for some important questions, such as the reason for self-treatment. This reflects non-sampling error and should be taken into account when results are reported.

Table 4 also compares the repeatability of responses in reinterviews with the original respondents and the "reinterview" results obtained from respondents who were not interviewed originally. Although "reinterviewing" a different respondent may not be methodologically rigorous, the fact that such individuals supplied the same answers to certain questions as the original respondents indicates that some, but not all, types of information were well known to most adult household members.

Table 2: **Response errors calculated from the age ranges used in the Sierra Leone survey**

	All respondents			Same respondent			Different respondent		
Age range of respondent (years)	Gross error (%)	Net bias (%)	Relative net bias (%)	Gross error (%)	Net bias (%)	Relative net bias (%)	Gross error (%)	Net bias (%)	Relative net bias (%)
0	16.0	0.5	+8.4	15.8	0.8	15.7	16.7	-1.1	-12.7
1-5	9.8	-1.7	-11.7	6.2	-1.3	-10.9	15.8	-4.9	-17.0
6-15	18.2	1.3	8.5	15.4	1.2	7.8	28.6	1.8	11.2
Adult	6.1	-0.1	-0.2	3.4	0.8	-1.1	20.0	4.3	9.2

Table 3: **Response errors calculated from 1-year intervals used in the Sierra Leone survey**

	All respondents				Same respondent				Different respondent			
Age range of respondent (years)	Not exact match (%)	Mean error (%)	Mean error (% age)	ABS mean error[a] (% age)	Not exact match (%)	Mean error (%)	Mean error (% age)	ABS mean error[a] (% age)	Not exact match (%)	Mean error (%)	Mean error (% age)	ABS mean error[a] (% age)
0	16.0	0.05	—	—	15.8	0.06	—	—	16.7	0	—	—
1-5	19.6	-0.02	0.2	6.1	15.6	-0.07	2.7	4.3	30.0	-0.06	5.4	9.3
6-15	56.1	-0.39	4.0	15.6	51.1	-0.06	0.6	13.2	45.4	-1.50	-19.6	23.6
Adult	6.1	—	—	—	3.4	—	—	—	20.0	—	—	—
All ages	13.0	—	—	—	12.0	—	—	—	17.4	—	—	—

[a] ABS = absolute.

Table 4: **Summary of non-sampling errors in the Sierra Leone survey**

Variable	Type of variable	Case error[a] (%)	Sample error[b] (%)	Effect of "different" respondents
Illness incidence	Binary	11	2 [c]	15% fewer cases reported
Age of ill household members	Continuous	13	2	Somewhat higher error
Sex of ill household members	Binary	0.5	0.5	Same error
Seriousness of illness	Binary	26	3	Nearly double the error
Action taken	*Multiple*	19	1	Slightly greater error
Result of action taken	Multiple	28	1.5	Double error rates
Amount paid for treatment	Continuous	1-41 [d]	0-22	60% greater error
Reason for nonmedical action	Multiple	27	2	Not analysed
Receiving an injection	Binary	10	3	Similar error
Availability of money	Binary	16	1	Much greater error
Source of money	Multiple	10	2.5	Similar error
Amount thought medical care would have cost	*Continuous*	17	15	*Not analysed*

[a] For binary or multiple-choice variables, the case error is the mean of gross errors for all categories; the error for individual categories may be much higher or lower. For continuous variables the case error is the mean of absolute errors for the entire range.
[b] For binary or multiple-choice variables, the sample error is the mean net bias for all categories. For continuous variables the sample error is the mean percent error for the entire range.
[c] Percent difference in mean incidence rates.
[d] Depends on price range—see text for details.

Discussion

Our results indicate that a typical small-scale household health interview survey can be subject to various degrees of non-sampling error. However, the meaning of repeatability as an appropriate measure of validity is open to question. If during reinterview, a respondent gives the same answer to a question asked during the original interview by a *different interviewer some time after the initial* interview, the respondent was probably sure of the answer in the first place; when given the opportunity to change the answer, the respondent chose not to do so. While such an answer is reliable, it implies little about its validity. Consistency of response does not exclude the possibility that the original reply was biased or a deliberate fabrication. When the interval between interviews is short, as in the present survey, a deliberate lie in the first interview would tend to be repeated in the second.

Similarly, a different answer to the same question could mean that the respondent originally had no firm answer in mind but guessed. The response in the reinterview might be nearer to the truth if the respondent deliberately reconsidered the question, or it could be another guess, or an attempt to satisfy the interviewer's perceived desire for an answer. Since respondents were always given the opportunity to say that they did not know an answer, the only reason for a deliberate lie should have been when it would benefit the respondent; such questions on questionnaires can usually be identified and minimized by explaining that truthful and accurate answers will be most valued.

For questions for which there is no evident motive for a lie, repetition of the initial response might indicate that it was valid, while a different response in the reinterview could mean that the respondent was uncertain and that the answer was probably not valid. In the latter case, leaving aside any interviewer recording errors, the original answer (normally the only one available for analysis) was probably wrong and it cannot be certain that the second answer is any more accurate. This is suggested by the lower overall rate of consistent responses obtained when a different respondent is "reinterviewed" about an event within the household. The following types of error can occur in such instances: the two respondents may have different opinions or interpretations of the question or one may simply not be as well informed about the situation.

Such difficulties may be insoluble in the absence of additional data about the validity of both the consistent and the inconsistent responses. An analysis of the quality of the U.S. Current Population Survey using reinterviews found that proxy data produced *lower* response variance than self-reported data (*13*). This arose partly because response variability is an inadequate measure of data quality and reliability is a necessary, but not a sufficient, condition for accurate data: the proxy reports may have been consistent without being valid.

Relevance to health interview surveys

The findings in Table 4 permit some generalizations to be made about the relative reliability of different types of questions often asked in health information surveys, as outlined below.

- Questions for which high response reliability (case error <10%) was observed included the following: the sex of household members: actions taken in response to illness: whether an injection was received during treatment; identification of household members who were ill during the recall period; and the source of money used to pay for treatment. If zero or medium-to-high amounts were reported to have been paid for treatment, the response reliability was also high.
- Case errors of 10–20% were associated with the following variables: incidence of illness; whether enough money was available to pay for treatment; age of the ill household members: the amount the respondent thought medical treatment would have cost in instances where self-treatment was used; and when low-to-medium amounts were paid for treatment.
- The highest case error (20–30%) related to the following: the result of the action taken; the seriousness of illnesses; the reason for choosing nonmedical treatment; and high levels of expenditure on treatment.
- Replies to questions of a subjective nature, such as the seriousness of an illness or the outcome of treatment, are subject to considerable uncertainty both for sick people themselves and proxy respondents. Answers to other questions, such as the occurrence of an illness and the sex of household members, were clearly recalled and known to other household members, as were actions taken in response to illness or important curative events such as injections.
- Response errors on the expenditures for treatment had a wide range. For expenditures greater than zero but less than 50 *leones* (US$ 1.0 = Le 65),[a] discrepancies were frequent, reflecting the small expenditures that occurred up to 2 weeks earlier and which were probably of no great importance to the respondent. Infrequent, large errors (recording errors or deliberate falsifications) also had a wide range, resulting in a high average error. Repeatability improved markedly with higher cost (50–300 *leones*, which included nearly all curative primary health care charges), perhaps reflecting the increased importance of these greater amounts to the household. The high error level for amounts >300 *leones* may have been due to deliberate misrepresentation in the original interviews, and is amplified by the small number of cases involved.

How these errors should be interpreted depends on the use to be made of the data. The nature of the variances is critical: for expenditures on treatment, a large number of small deviations from zero produced a low correlation between the data sets, but whether a certain group spent an average of zero or 4 *leones* for treatment is probably unimportant against expenditures of hundreds or thousands of *leones*. For the range of expenditures as a whole, only a few instances of respondents reporting grossly higher amounts at the reinterview or a similar magnitude of interviewer recording error can degrade the overall repeatability of the data. A consistent system of treating "outliers" is therefore essential.

In studies where the misclassification of categorical responses is high, the minimum requirement is that the researchers should draw attention to this, and preferably indicate how it affects the interpretation of the results. In the absence of an accepted technique for taking such non-sampling errors into account, a common-sense approach can be used in reporting results. For example, a χ^2 test of the data from the Sierra Leone survey shown in Table 5 indicated that the difference between the rate of response for "not enough money" in urban (URB) and other areas was significant at the P >0.01 level. This high level of significance arose because the 95% CI for the percentages of individuals who had insufficient money was ($P \pm 1.2$)%[b] if only sampling errors were taken into account; the difference between, e.g., 53.8% and 41.8%, was therefore highly significant. As shown in Table 6, the net bias for the category "not enough money" was about –8% and the gross error was 11%. If this non-sampling error is taken into consideration, the confidence intervals are wider and the probability that there is a true difference between the two proportions is lower. The difficulty is then to determine which of the error statistics most closely corresponds to a confidence interval widened by non-sampling error. In this instance, the net bias of –8% was interpreted to mean that the true proportion would be ($P \pm 8$)%; combining this with the confidence interval due to sampling error resulted in a new 95% CI of ($P \pm 8.1$)%. Since the confidence intervals for the proportions being considered now overlap, the differences cannot be significant at the P <0.05 level.

[a] Agricultural workers earned Le 30–60 per day, depending on the district.

[b] Confidence interval : $P \pm z' \times$ s.e., where P is the proportion observed in the sample, s.e. is the standard error ($\sqrt{P(1-P)/n}$) and z' is the standard normal deviation for the desired confidence level.

Table 5: **Survey data: reason given by respondents for not using medical treatment in the Sierra Leone survey**

Reason[b]	% of individuals by geographical area[a]					
	RUR 1	RUR 2	RUR 3	RUR 4	URB	All
Thought not serious	11.0	12.5	6.1	15.4	31.3	11.9
Got better soon	0.7	4.5	1.1	1.2	3.0	2.1
Too far to travel	5.7	8.0	24.4	10.3	1.5	12.1
Not enough money	54.6	51.6	55.2	53.8	41.8	53.1
Don't like PHU staff	1.4	1.5	0.3	1.2	—	1.0
Thought drugs unavailable	2.1	0.5	1.3	2.8	—	1.5
Not effective for problem	8.2	7.3	2.1	2.0	—	4.7
Knew a self-treatment	13.1	7.3	4.2	8.3	14.9	8.2
Other	0.7	2.8	3.7	2.8	7.5	2.8
PHU was closed	0.7	2.3	—	—	—	0.8
Don't know	1.8	1.8	1.6	2.4	—	1.7
Total	100.0	100.0	100.0	100.0	100.0	100.0
No. of cases, by area	282 (20.5)[c]	399 (29.0)	377 (27.4)	253 (18.4)	67 (4.9)	1378 (100.0)

[a] RUR 1–4 = rural areas. URB = urban area.
[b] PHU = primary health care unit.
[c] Figures in parentheses are percentages.

Table 6: **Error analysis: reason given by respondents for nonmedical treatment in the Sierra Leone survey**

Reason[a]	Original survey (%)	Gross error (%)	Reinterview survey (%)	Net bias (%)	Relative net bias (%)
Thought not serious	11.0	48.0	13.5	–2.52	–18.6
Got better soon after	1.8	75.0	1.3	0.45	34.5
No transport available	16.3	54.0	8.3	8.00	96.4
Not enough money	51.5	11.1	59.4	–7.85	–13.2
Did not like provider	1.8	75.0	0.9	0.89	101.7
No drugs at PHU	0.9	0.0	0.9	0.01	0.9
Provider not effective	6.2	35.7	4.8	1.36	28.4
Knew self-treatment	6.2	42.9	5.2	0.93	17.7
Other	3.5	62.5	3.5	0.03	0.9
PHU was closed	0.9	0	0.9	0.01	0.9
No response	0		1.3	–1.31	100.0

[a] PHU = primary health care unit.

Conclusions

Although sampling errors receive considerable attention in the design and implementation of health interview surveys, non-sampling errors are largely ignored. In this article we have stressed the importance of measuring one aspect of non-sampling errors—response reliability. We have described a method of analysing the results of reinterviews, and using the example of reinterviews in a morbidity and health expenditure survey in Sierra Leone we have highlighted the type of questions that are prone to this kind of error. Because reinterviews are rarely conducted and their results are even less often reported, there are few findings with which to compare our results. However, in view of the burgeoning number of health interview surveys being carried out in developing countries to provide data on which to base policies, it is crucial that the quality of such surveys be more routinely measured and reported.

Acknowledgements

We gratefully acknowledge the advice of Dr T. Mertens, Mr S. Cousens and Mr T. Marshall during the preparation of this article. The field research was supported by UNICEF.

Résumé

Evaluation de la fiabilité des réponses aux enquêtes de santé à l'aide d'entretiens complémentaires

Les enquêtes auprès des ménages sont souvent utilisées dans les pays en développement pour mesurer l'état de santé et les facteurs liés à la capacité et à la bonne volonté des individus à payer les services de santé reçus. Ces enquêtes sont sujettes à deux types d'erreurs fondamentaux: les erreurs d'échantillonnage et les erreurs indépendantes de l'échantillonnage. Si les personnes qui conçoivent les enquêtes et les décideurs politiques ont souvent conscience des erreurs d'échantillonnage, nombreux sont les chercheurs qui ne tiennent pas compte des erreurs indépendantes de l'échantillonnage, souvent plus graves. La question de l'importance des erreurs non liées à l'échantillonnage est traitée dans les manuels classiques sur les méthodes d'enquête mais très peu de recherches ont été faites pour déterminer les types de variables qui donnent le plus souvent lieu à ces erreurs. La présente étude examine les erreurs autres que les erreurs d'échantillonnage dans une enquête auprès des ménages faite en Sierra Leone en comparant les résultats des entretiens complémentaires aux réponses données lors de l'entretien initial. Elle propose diverses façons de signaler les erreurs non liées à l'échantillonnage dans les enquêtes et dresse la liste des questions qui semblent s'accompagner d'un faible pourcentage d'erreurs indépendantes de l'échantillonnage (mesures prises en cas de maladie, source de l'argent utilisé pour payer le traitement, etc.) ou qui s'accompagnent d'un fort pourcentage d'erreurs (gravité de la maladie, raison pour laquelle un traitement non médical a été choisi, résultat du traitement, etc.). Des stratégies de contrôle de la qualité doivent être utilisées dans les études qui incluent des variables s'accompagnant d'un taux élevé d'erreurs non liées à l'échantillonnage (entretiens complémentaires pour mesurer la fiabilité des réponses, par exemple). Des entretiens complémentaires devraient être faits plus souvent et leurs résultats publiés.

References

1. **Moser, C.A. & Kalton, G.** *Survey methods in social investigations.* London, Heinemann, 1971.
2. **Casley, D.J. & Lury, D.A.** *Data collection in developing countries.* Oxford, Oxford University Press, 1981.
3. **Ross, D.A. & Vaughan, J.P.** *Health interview surveys in developing countries.* London, London School of Hygiene and Tropical Medicine, 1984 (EPC Publication 4).
4. **Mackenzie, S.G. & Lippmann, A.** An investigation of report bias in a case–control study of pregnancy outcome. *American journal of epidemiology,* **129**: 67–73 (1989).
5. **Britton, M. & Birch, F.** *1981 Census post-enumeration survey.* London, HMSO, 1985.
6. **Thomas, R.** Data quality in the National Census of Population. *Survey methods newsletter (Spring)*: 6–7 (1987).
7. **Cochrane, A.L. et al.** Observer's error in taking medical histories. *Lancet,* **1**: 1007–1009 (1951).
8. **Mertens, T.E. et al.** Health impact evaluation of improved water supplies and hygiene practices in Sri Lanka: background and methodology. *Tropical medical parasitology,* **41**: 79–88 (1990).
9. **Mertens, T.E. et al.** Epidemiological methods to study the interactions between HIV infection and other sexually transmitted diseases. *AIDS,* **4**: 57–65 (1990).
10. **Fabricant, S.J. & Kamara, C.W.** *The financing of community health services in Sierra Leone: initial results of an operations research study.* Bamako, August 1990 (Bamako Initiative Technical Report Series No. 1).
11. **Fabricant, S.J. & Kamara, C.W.** *The financing of community health services in Sierra Leone: seasonality, socioeconomic status, gender, and location factors affecting access to services.* Bamako, November 1991 (Bamako Initiative Technical Report Series No. 8).
12. **Kroeger, A.** Health interview surveys in developing countries: a review of the methods and results. *International journal of epidemiology,* **12**: 465–481 (1983).
13. **O'Muircheartaigh, C.** The quality of reporting on the U.S. Current Population Survey. *Survey methods newsletter (Spring)*: 7–10 (1987).

Dengue in China: a clinical review

Fu-Xi Qiu,[1] D.J. Gubler,[2] Ji-Chang Liu,[1] & Qiu-Qian Chen[3]

Three etiologically proven outbreaks of dengue fever and one etiologically confirmed epidemic of dengue haemorrhagic fever have occurred in south China since 1978. The first of these, an epidemic of dengue due to virus type 4 took place in Shiwan town, Foshan city, Guangdong Province, in 1978; the epidemic began in May and ended in November. The clinical manifestations of 583 hospitalized patients were observed from August to October. The majority (81.3%) of patients were aged 21–50 years (male : female = 1.2:1). The course of illness was about 1 week in most cases; three patients (0.5%) died.

A local outbreak of dengue due to virus type 1 occurred in Shiqi town, Zhongshan County, Guangdong Province, from September to November 1979. The majority of patients were older children and adolescents. There was no marked difference between males and females in terms of the course of the illness, and there were no complications or deaths.

A large epidemic of dengue due to virus type 3 occurred on Hainan Island in 1980. The clinical manifestations of 510 hospitalized patients (mostly adolescents and adults) were observed from April to September. Some patients developed rare complications, such as loss of hair, acute intravascular haemolysis, and multiple peripheral paralysis; there were four deaths (0.78%).

The first known epidemic of dengue haemorrhagic fever in China occurred among 10–29-year-olds on Hainan Island in 1985 and 1986. There were no essential differences between males and females. Some cases had rare complications such as acute intravascular haemolysis, while others had diffuse intravascular coagulation and altered mental status; 10 patients (6.5%) died.

History of dengue in China

Epidemics of dengue-like illness in China were not uncommon in the 1940s. For example, in 1947, an epidemic of febrile illness, suspected to be dengue fever (DF), occurred along the south-east coast of China, including Shanghai city and the coastal areas of Jiangsu, Zhejiang and Fujian Provinces. Etiological and epidemiological investigations were not carried out on this or other epidemics. From 1950 to 1978, outbreaks of dengue-like illness were not reported in China.

Since 1978 three etiologically documented outbreaks of DF and an epidemic of dengue haemorrhagic fever (DHF) caused by all four virus serotypes have occurred in south China. An epidemic of DF due to dengue virus type 4 (DEN-4 virus) occurred in Foshan city, Guangdong Province in 1978 (*1–3*). A more localized outbreak of mild DF due to dengue virus type 1 (DEN-1 virus) occurred in Zhongshan County, Foshan Prefecture, Guangdong Province, in 1979, followed by a large epidemic caused by dengue virus type 3 (DEN-3 virus) on Hainan Island in 1980 (*4–6* and Qiu, F.-X., unpublished data, 1980). In 1985–86 an epidemic of DHF caused by dengue virus type 2 (DEN-2 virus) occurred in the same areas of Hainan Island (*7*).

Circumstantial evidence suggested that the dengue viruses causing these four outbreaks may have been imported. In the late 1970s there was considerable commercial activity between Guangdong Province, Hainan Island, and south-east Asia. Beginning in 1977, many south-east Asian tourists visited southern China, thus providing the ideal means of importing new strains or serotypes of dengue viruses into the country (*8*). Moreover, the population densities of *Aedes* mosquitos were high in these areas of south China, thus putting them at high risk for epidemic transmission of dengue fever when new strains/serotypes of viruses were introduced. Serological data from the 1978 epidemic, however, suggested that most cases had a secondary-type serological response, indicating that, despite the absence of previous epidemics, dengue viruses were endemic in the area.

[1] Professor, Institute of Epidemiology and Microbiology, Chinese Academy of Preventive Medicine, Beijing, China. Requests for reprints should be sent to Professor Fu-Xi Qiu at the following address: 6 Cang Nan Hu Tong, Chao Yang Men Nei, Beijing 100 010, China.

[2] Director, Division of Vector-borne Infectious Diseases, National Center for Infectious Diseases, Centers for Disease Control, Fort Collins, CO, USA.

[3] Chief Physician, Hainan Provincial Ocean Fishery Company Hospital, Zhan County, Hainan Island, China.

Reprint No. **5393**