Bias resulting from missing information: Some epidemiological findings

ANTONY COX, MICHAEL RUTTER, BRIDGET YULE, AND DAVID QUINTON

From the Department of Child and Adolescent Psychiatry, London

SUMMARY The biases resulting from missing information were examined in three psychiatric epidemiological studies. In each study, cases with missing information could be compared with the main sample because data were available from several sources or at several points in time through a longitudinal study. In almost all instances, cases with missing data differed systematically in terms of variables crucial to the questions being studied. In general, they tended to include a higher proportion with problems of various kinds—such as, behavioural deviance, reading backwardness, child or adult psychiatric disorder, and marital discord. The characteristics or circumstances of those giving information were generally more strongly associated with co-operation in testing or interviewing than the characteristics of those about whom information was sought. In some situations, the nature and degree of distortion resulting from missing information could lead to biased results.

In almost all studies data are lacking from a proportion of the sample and it is sometimes assumed that the missing cases have not distorted the findings. When the matter has been examined, this assumption has often received some support from the finding that the cases missed did not differ from the remainder of the sample in terms of such characteristics as age, sex, and social class. However, just because they have been missed, it has rarely been possible to determine if the missed cases differed in terms of the variables central to the hypotheses being tested or the disorders being studied. Such information is likely to be available only when information is sought from several sources or on several occasions such as during the course of a longitudinal study.

The findings in these situations are much less reassuring. Thus, in the Isle of Wight study (Rutter *et al.*, 1970) children whose parents failed to complete behavioural questionnaires were twice as likely as other children to receive a deviant score on questionnaires completed by teachers. Similarly, West and Farrington (1973) found that boys whose parents refused to be interviewed were more likely to become delinquent than boys whose parents co-operated with the study. In the survey by Shepherd *et al.* (1971) the children whose parents failed to complete questionnaires were more likely to have low attainments, to show behavioural difficulties in class, and to become delinquent later.

In her 30-year follow-up study of psychiatric clinic attenders, Robins (1963) discovered relatively few factors associated with refusal to be interviewed, but sociopathic individuals were more difficult to trace because of their greater mobility both from house to house and also from area to area (Robins, 1966). Sims (1973) noted that patients with neurosis who were difficult to trace in a follow-up study were more likely to have died.

The opportunity to examine more closely the possible biases resulting from missing information arose from three interrelated psychiatric epidemiological studies. Data were available from several sources in all these studies and in two, longitudinal data were also obtained. This paper summarises the main findings.

Samples and methods

INNER LONDON STUDY

In 1970 the total population of 10-year-old children attending local authority schools in an inner London borough was assessed by group tests of reading and non-verbal intelligence and by a questionnaire on behaviour completed by teachers (Berger *et al.*, 1975; Rutter *et al.*, 1975). From this survey two subgroups (a random sample and a sample of children with deviant scores on the teacher's questionnaire) were chosen for more intense studies by means of interviews with

parents and teachers.

Three years, and again four years later, the same total population of children (then at secondary school) was re-assessed by means of group tests of reading and of non-verbal intelligence and by a behavioural questionnaire completed by teachers (Rutter, 1977). Special attention was paid to the children at the 25 secondary schools in, or immediately adjacent to, the borough studied in 1970.

ISLE OF WIGHT STUDY

In 1969 the total population of 10-year-old children living on the Isle of Wight and attending local authority schools was assessed in a somewhat similar way to that in the inner London study (Berger *et al.*, 1975; Rutter *et al.*, 1975). Two similarly defined subgroups were studied intensively by interviewing both the parents and teachers.

FAMILY ILLNESS STUDY

A prospective study of families in which one parent had a psychiatric disorder was carried out in the same inner London borough as that used for the general survey of 10-year-old children. The sample consisted of all patients who spoke colloquial English at home, who had at least one child under the age of 15 years, and who had been *newly* referred to a psychiatric clinic during the course of a 10-month period (Rutter, 1970). At the time of initial referral, and again one and four years later, detailed systematic interviews were held with the patient and spouse (if there was one) to obtain information on patterns of family life, the marriage relationship, and on the psychiatric state of all family members (Rutter and Brown, 1966; Quinton *et al.*, 1976). Shorter interviews were carried out two and three years after the initial referral.

Behavioural questionnaires were completed by teachers for all children of school age in the families at the time of initial referral and again at each of the four years of follow-up. Questionnaires were completed in the same way throughout the follow-up study for a control group of children of the same age, sex, ethnic background, and in the same school class as the patients' children at the start of the study.

Results

GROUP TESTS

In the inner London study, reading tests were administered to all 13-year-old children attending 24 secondary schools* serving one borough. Tests

*In later stages of the study data were obtained for 25 schools but at this stage one school failed to carry out testing.

were repeated to include children absent on the day of the first tests. However, even after several testing occasions 8.6% of the children had still not been tested. Because the children had previously been studied in primary school, it was possible to use measures obtained at 10 years of age to compare in the 13-year-old children the absentees with the children tested at that age. This is done in Table 1.

Table 1 *Characteristics of children with missing data on group tests (reading test 1973)*

Characteristics at 10 years	Reading test at 13 years	
	Children absent (n = 136)	Children tested (n = 1448)
Behavioural deviance on teacher's questionnaire	37.5%	20.8%**
Reading at least 2 SD below mean	23.5%	14.0%**

**P < 0.01

Of the children successfully re-tested at the age of 13, 14% when aged 10 years had had a reading score on a group test at least two standard deviations below the standardised sample mean. But, among the absentees at the age of 13, 23.5% had had a similarly low reading score at age 10 years. This difference is significant at 1% level. In the same way, nearly twice as many of the absentees had had deviant scores on the teacher's questionnaire at the age of 10 compared with those tested at 13 years (37.5% v. 20.8%; P < 0.01).

Table 2 *Characteristics of children with missing data on group tests (reading test 1974)*

Characteristics at 10 years	Reading test at 14 years	
	Children absent (n = 116)	Children tested (n = 1559)
Behavioural deviance on teacher's questionnaire	40.5%	14.9%**
Reading at least 2 SD below mean	24.8%	13.6%**

**P < 0.01

The same reading test was given again a year later, in 1974, because one school had not given the reading test and it was hoped that the proportion of children with missing information might be reduced. This time the absenteeism rate was 6.9%. These results are shown in Table 2. Once more, twice as many of the absentees had been behaviourally deviant at 10 years of age compared with those present at 13 (40.5% v. 14.9%; P < 0.01).

Table 3 Characteristics of children with missing data in group tests (non-verbal IQ test 1973/74)

Characteristics at 10 years	IQ test at 13 or 14 years	
	Children absent both years (n = 44)	Children tested (n = 1631)
Behavioural deviance on teacher's questionnaire	40.9%	22.1%**
Reading at least 2 SD below mean	29.3%	14.0%**

* P < 0.01

The non-verbal intelligence test was given in 1974 to children who had repeatedly missed the test in 1973. In spite of giving the test on several different days in 1974, there were still 44 children (2.6%) who were absent on all testing occasions in both years. Table 3 shows the characteristics of these persistent absentees at 10 years of age compared with the rest of the sample. Again the rates of behavioural deviance (40.9% v. 22.1%; $P < 0.01$) and poor reading attainment (29.3% v. 14.0%; $P < 0.01$) were twice as high among the absentees.

PARENTAL INTERVIEWS

In the Isle of Wight study 202 children were selected for study by parental interview. Interviews were not obtained in 16 cases (7.9%); 12 families refused to participate and four could not be traced. In the inner London study there were no parental interviews for 21 (8.4%) out of 249 selected children (20 refusals and one not traced).

Table 4 Characteristics of children whose parents were not interviewed (Isle of Wight and London studies combined)

Characteristics of children	Parents not interviewed (n = 37)	Parents interviewed (n = 414)
Behavioural deviance on teacher's questionnaire†	75.7%	58.2%*
Mean non-verbal IQ (London only)	86.9	86.9
Mean reading quotient (London only)	85.8	87.9

†The samples interviewed were deliberately stratified so that some three-fifths of the children would be deviant on the teacher's questionnaire. Accordingly, the proportions do not represent population prevalence figures. For prevalence data see Rutter *et al.*, 1975.

*P < 0.05

In both samples a higher proportion of the children whose parents were not interviewed were rated as showing behavioural deviance on the teacher's questionnaire ($P < 0.05$). However, this difference was accounted for by the parents of boys. Of the 178 boys who were behaviourally deviant on the teacher's questionnaire, 17 had parents who were not interviewed compared with only one out of the 94 non-deviant boys. The total groups of children whose parents were not interviewed did

not differ in terms of sex, non-verbal IQ, or reading attainment.

The above data apply to differences between parents interviewed and not interviewed. The inner London study data allow similar comparisons with respect to degree of co-operation. Although very few families had no interviews at all there was a larger proportion whose interviews were incomplete in some respect. Among the 190 families in which the key child was living with both parents there were 43 for whom data were obtained from only one informant*. In an additional eight cases the interview could not be completely finished, usually because the husband would not let it continue. The differences between these 51 families with incomplete information and the 139 with complete data are summarised in Table 5.

Table 5 Characteristics of families with missing data on parental interview (inner London borough study)

Characteristics of family	Incomplete interviews (n = 51)	Complete interviews (n = 139)
Child has conduct/mixed disorder (interview rating)	23.5%	7.2%***
Marked marital discord	38.1%	16.6%**
Non-manual social class	18.0%	20.4%
Psychiatric disorder of father of mother	15.2% 36.2%	12.4% 27.5%
Child has behavioural deviance on teacher's questionnaire†	64.7%	59.7%

†See footnote to Table 4

***P < 0.001

**P < 0.01

Incomplete interviews were not related to the sex of the child, the social class of the family, the psychiatric state of the parents, or behavioural deviance in the child as assessed by teacher's questionnaire. However, children in families with incomplete interviews were more likely to have a conduct or mixed emotional/conduct disorder as rated from parental interview (25.5% v. 7.2%; $P < 0.001$). Marital discord was also twice as common (38.1% v. 16.6%; $P < 0.01$) if the interviews were incomplete. Both the psychiatric rating (Rutter *et al.*, 1975) and the marital rating (Quinton *et al.*, 1976) have been shown to be reliable and valid.

Broadly similar findings were obtained by comparing mothers rated by the interviewer as fully co-operative and those rated as showing some degree of reluctance or non-cooperation. Less than fully co-operative mothers were more likely to have a child with a conduct or mixed emotional/conduct

*In this study the interview with the husband was mainly used as a check on information provided by the wife, rather than as a primary source of data.

Table 6 Family characteristics and degree of informant co-operation (inner London borough study)

Characteristics of family	Informants	
	Any degree of reluctance (n = 85)	Fully co-operative (n = 142)
Child has conduct/mixed disorder (interview rating)	20.0%	7.8%**
Marked marital discord	30.3%	17.2%**
Father has psychiatric disorder (interview rating)	26.5%	12.1%**
Child has behavioural deviance on teacher's questionnaire†	63.5%	65.6%

†See footnote to Table 4

**P < 0.01

disorder (20.0% v. 7.8%; $P < 0.01$), more likely to have a discordant marriage (30.3% v. 17.2%; $P < 0.01$), and more likely to have a husband with some kind of psychiatric disorder (26.5% v. 12.1%; $P < 0.01$).

Various aspects of co-operativeness could also be assessed in the family illness study in which interviews with the husband and wife were of equal importance and were dealt with similarly. At the time of the first family contact one or other was a newly referred psychiatric patient. The only two variables associated with co-operativeness were patient status and sex of informant.

Thus, in the initial contact with the family (in cases when the husband and wife were living together), fewer interviews were obtained with spouses than with patients (85% v. 98%; $P < 0.001$), and there was a tendency for men to co-operate less well than women (88% v. 95%; $P < 0.1$; patients and spouse combined). The poorest co-operation was by male spouses (79.2%) and the best by female patients (99.0%). During the course of the four-year longitudinal study the pattern remained much the same.

Thus in the first and fourth years of follow-up, spouses were twice as likely as patients to refuse interview or to be repeatedly out (first year, 18% v. 9%; $P < 0.05$; fourth year, 23% v. 9%, $P < 0.01$). Similarly, twice as many men could not be interviewed (first year, 19% v. 7%, $P < 0.01$; fourth year, 21% v. 8%, $P < 0.01$). Among the male spouses it was those who were completely well or those with personality disorders (rather than those with acute transient psychiatric problems) who were most likely to be uncooperative ($P < 0.025$).

This longitudinal study also made it possible to observe the extent to which non-cooperation was a persistent family characteristic. There were many examples of people who refused interview on one occasion but yet agreed on another. Furthermore, there were very few families in which both husband and wife persistently refused interview.

Nevertheless, when at least one member of a cohabiting couple refused to be interviewed on the first occasion, there was an increased chance that one or both would refuse to be interviewed subsequently. When both husband and wife were interviewed at the outset, both were interviewed in the fourth year of follow-up in 72% of cases in which they continued to live together. In contrast, if only one had been interviewed at the outset, in only 33% of cases were both co-operative in the final year ($P < 0.01$).

MOVE OF HOUSE

In longitudinal studies a further source of missing information is migration of the family to other parts of the country (or abroad). The family illness study afforded the opportunity to determine whether the migrating families differed from those remaining in London. During the course of the four-year follow-up, 13.5% of the patients' families and 10.6% of the controls moved out of inner London. The very small difference between these two groups indicates that parental psychiatric state is unrelated to the likelihood of internal migration. Within the patient sample migration was also unassociated with the presence of marital discord (37.9% of the migrants showed discord at the initial assessment compared with 44.8% of the non-migrants). Furthermore, within the control group behavioural deviance in the children at the start of the study was unrelated to migration out of London during the next four years (13.6% of the migrant children were behaviourally deviant compared with 18.6% of the non-migrant children).

Discussion

In all the epidemiological studies we have considered there have been substantial and important differences between individuals who co-operated with interviews or tests and those who did not. The findings need to be discussed from two rather different perspectives. Firstly, what light is thrown on the process of co-operation or non-cooperation in research projects? Secondly, what implications for epidemiological research stem from the nature and degree of bias resulting from missing information?

In keeping with previous studies, cases with missing data were not usually distinguishable in terms of demographic characteristics such as social class. On the other hand, in almost all instances the children or families with missing information *did* differ in terms of the variables crucial to the questions being studied. In general, the missing cases included a higher proportion who were

deviant or disordered in some way. However, the pattern was not the same in all situations.

It appears that co-operation or willingness to participate in a study is best understood in relationship to the characteristics or circumstances of the person being interviewed or tested. Only indirectly is co-operation related to the characteristics of the individuals about whom information is being sought.

Thirteen or 14-year-old children who were repeatedly absent from school at the time tests were administered were twice as likely as other children to have been behaviourally deviant and/or backward in reading when they were younger. Persistent truancy is more common in boys who are delinquent or who have low scholastic attainments, and boys who have poor educational progress and bad conduct at 10 years of age are quite likely to have the same difficulties at 14 (Graham and Rutter, 1973; Rutter, 1977). Labouvie *et al.* (1974) also found that adolescents who refused to participate or who were unavailable for testing had lower scores on cognitive tests than those who had co-operated.

Conduct disorders in childhood and adolescence are linked with school truancy and with personality disorders in adult life (Robins, 1966), so it is to be expected that childhood antisocial problems would tend to predict co-operation in follow-up studies. Thus Bebbington (1970), in a 10-year follow-up study of London adolescents, found that those who stalled or who refused to be interviewed at follow-up tended to have been of lower intelligence when they were aged 13. These people were also more likely to have been delinquent.

In the general population studies, husbands were less likely to be interviewed if their wives gave evidence that the marriage was unhappy. In some instances the women dissuaded the interviewers from approaching their partners. The lack of co-operation by the husbands in discordant marriages could have been either because of their reaction when they found out that their wives had disclosed family difficulties or because wives blocked the contact with the interviewer for fear of that reaction. The reduced co-operation in women who had husbands with a psychiatric problem may have been based on similar fears. On the other hand the children whose parents could not be interviewed did not differ from other children in IQ or reading attainment. The difference in the proportion of those with conduct difficulties was presumably because of family discord and parental problems (Rutter, 1971).

Co-operation was better and not so clearly related to family or personal problems in the

longitudinal investigation of patients' families. This is probably a consequence of several factors. Firstly, strenuous efforts were made by the project team to maintain contact and to obtain interviews. Secondly, there was such a high rate of parental discord and psychiatric disorder in the group as a whole that it was less likely that these factors would distinguish individuals who co-operated less well. Thirdly, they constituted a group who had recently attended a psychiatric service and might therefore be more inclined to co-operate with workers who were linked with the services which had helped them at their time of need. There is support for this interpretation in the finding that patients remained more co-operative than their spouses throughout the four years of the study.

In all the current studies men proved more difficult to interview than women. In part this may have been because the interviews were primarily focused on children and men may not have seen children as their province. In part, too, it may have been that practical difficulties stemming from long working hours may have prevented interviews with men who would have been willing to be seen had time permitted. However, it is unlikely that this entirely explains the difficulty as other studies have noted that fathers were crucial in determining co-operation in conjoint family treatment (Shapiro and Budman, 1973; Slipp *et al.*, 1974).

Our findings, as well as those of others (Bebbington, 1970; West and Farrington, 1973) indicate that uncooperative informants are not a homogeneous group. West and Farrington suggest that uncooperative parents tend to be of one of two types: either those with serious social or psychological difficulties or alternatively those without gross disturbance but who are overprotective, self-contained, and independent. Our findings suggest that men with a personality disorder were most likely to be uncooperative. Those with acute disorders and those who had sought psychiatric help, on the other hand, were usually very willing to be interviewed. Perhaps some mental disorders are associated with a great willingness to discuss personal matters whereas with others the opposite is true (Cozby, 1973). Lack of co-operation may reflect antisocial attitudes and disturbed personal relationships or it may stem from a dislike of self-disclosure or intimacy which has nothing to do with psychiatric disorder.

In the current study, internal migration was not associated with any measure of disorder in parents or children. This was also so in the Bebbington (1970) study, but not in Robins's investigation (Robins, 1966). The matter requires further study and it would not be justified on these findings to

assume that migration is never a source of 'bias'.

It is necessary to consider the implications for epidemiological research which stem from the findings of bias resulting from missing information. Our evidence, as well as that from previous studies, clearly shows that individuals or families who are not interviewed or tested, or for whom data are partially lacking, tend to differ in important ways from the main samples with complete data. Frequently cases with missing data included a higher proportion with problems of various kinds.

Obviously, the nature of this bias could cause serious distortions in results, but the extent to which results will in fact be distorted depends on a variety of other considerations. In all the studies we undertook, persistent efforts by the investigators led to a very high level of co-operation so that refusal rates were kept below and usually well below 10%. In these circumstances the distortion of total population rates which results from missing data is trivial. Moreover, because data were always available from multiple sources it was possible to allow for the missing information if this was required.

However, there are two circumstances in which missing information may cause more serious distortions. The first is when conclusions are being drawn from deviant subsamples rather than from the population as a whole. This may be illustrated by reference to Table 2. Only 7% of children missed the reading test at 14 years. Although this 7% was nearly twice as likely to be backward in reading, the missing data will make very little difference to the population mean scores on reading, simply because 7% is such a small proportion. On the other hand, if the study concerned the reading attainments of behaviourally deviant children the distortion would be much greater because 17% of behaviourally deviant children missed the tests.

The second circumstance is when refusal rates are much higher than they were in our investigations. It is not at all uncommon to publish epidemiological or longitudinal studies in which data are missing for more than one-quarter of the population, and sometimes even for half. In these cases distortions could be considerably greater. What we can conclude is that it behoves all investigators to keep non-cooperation rates as low as they can and to check as far as possible on the characteristics of cases with missing data. Multiple sources of data and repeated tests or interviews are invaluable for this purpose.

Reprints from M. Rutter, Department of Child and Adolescent Psychiatry, Institute of Psychiatry, De Crespigny Park, Denmark Hill, London SE5 8AF.

References

- Bebbington, A. C. (1970). The effect of non-responses in the sample survey with an example. *Human Relations*, 23, 169-180.
- Berger, M., Yule, W., and Rutter, M. (1975). Attainment and adjustment in two geographical areas. II. The prevalence of specific reading retardation. *British Journal of Psychiatry*, 126, 510-519.
- Cozby, P. C. (1973). Self-disclosure: a literature review. *Psychological Bulletin*, 79, 73-91.
- Graham, P., and Rutter, M. (1973). Psychiatric disorder in the young adolescent: a follow-up study. *Proceedings of the Royal Society of Medicine*, 66, 1226-1229.
- Labouvie, E. Q., Bartsch, T. W., Nesselroade, J. R., and Baltes, P. B. (1974). On the internal and external validity of sample longitudinal designs. *Child Development*, 45, 282-290.
- Quinton, D., Rutter, M., and Rowlands, O. (1976). An evaluation of an interview assessment of marriage. *Psychological Medicine*, 6, 577-586.
- Robins, L. N. (1963). The reluctant responder. *Public Opinion Quarterly*, 27, 276-286.
- Robins, L. N. (1966). *Deviant Children Grown Up*. Williams and Wilkins: Baltimore.
- Rutter, M. (1970). Sex differences in children's responses to family stress. In *The Child in His Family*. Edited by C. Koupennik. Wiley: New York.
- Rutter, M. (1971). Parent-child separation: psychological effects on the children. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 12, 233-260.
- Rutter, M. (1977). Prospective studies to investigate behavioural change. In *Methods of Longitudinal Research in Psychopathology*. Edited by J. S. Strauss, H. M. Babigian, and M. Roff. Plenum: New York.
- Rutter, M., and Brown, G. (1966). The reliability and validity of measures of family life and relationships in families containing a psychiatric patient. *Social Psychiatry*, 1, 38-53.
- Rutter, M., Cox, A., Tupling, C., Berger, M., and Yule, W. (1975). Attainment and adjustment in two geographical areas. I. The prevalence of psychiatric disorder. *British Journal of Psychiatry*, 126, 493-509.
- Rutter, M., Tizard, J., and Whitmore, K., Editors (1970). *Education, Health and Behaviour*. Longman: London.
- Shapiro, R. J., and Budman, S. H. (1973). Defection, termination and continuation in family and individual therapy. *Family Process*, 12, 55-67.
- Shepherd, M., Oppenheim, B., and Mitchell, S. (1971). *Childhood Behaviour and Mental Health*. University of London Press: London.
- Sims, A. C. P. (1973). Importance of a high tracing rate in long-term medical follow-up studies. *Lancet*, 2, 433-435.
- Slipp, S., Ellis, S., and Kressel, K. (1974). Factors associated with engagement in family therapy. *Family Process*, 13, 413-428.
- West, D. J., and Farrington, D. P. (1973). *Who Becomes Delinquent?* Heinemann: London.