How to identify information bias due to self-reporting in epidemiological research

L Fadnes, A Taube, T Tylleskär

Citation

L Fadnes, A Taube, T Tylleskär. *How to identify information bias due to self-reporting in epidemiological research*. The Internet Journal of Epidemiology. 2008 Volume 7 Number 2.

Abstract

Reality can be distorted in many ways when seen through a questionnaire or an interview. Such distortions may be systematic, introducing bias. Bias can spoil research by indicating false associations or failing to detect true relationships. It is practically impossible to eliminate measurement errors totally, but estimating the extent of disagreement and assessing whether the errors are systematic should be a priority in epidemiological research. The aim of this article is pedagogically oriented. Through Medline searches and cross-references, 1400 articles were identified, of which 53 were chosen. This review gives an overview of information bias, focusing on recall period, selective recall, social desirability, interview situation and interviewing tools, question phrasing, alternative answers and digit preference. We use a problem identification approach and also present some possible solutions, exemplifying the different topics by research conducted in the fields of HIV-AIDS, nutrition and alcohol abuse. Methods for measuring bias are presented.

INTRODUCTION

Bias can be defined as “any trend in the collection, analysis, interpretation, publication, or review of data that can lead to conclusions that are systematically different from the truth” (1). Research results can be undermined by bias leading to false associations or failure to identify true relationships (2). Random errors tend to cancel out when the number of observations increases, whereas systematic error or bias does not (3). It is practically impossible to eliminate measurement errors totally, but estimating the extent of disagreement should be a priority in epidemiological research (4).

Epidemiological research involves studying part of the reality with selection bias as a threat (5). The data collection may influence responses, introducing information bias. The research is put into perspective and analysed with risks as misinterpretations and wrong causations, e.g. due to confounding (5, 6). The terminology of bias is extensive with a range of parallel concepts (5). Information bias will be the focus of this article.

Memory lapses, misinterpretations, simplification, and modification of answers to make them more socially desirable can all complicate interpretation and analysis. We will elucidate these topics with examples from research in the areas of HIV-AIDS, nutrition and alcohol abuse. We selected these topics as they illustrate different aspects of information bias. Some psychological aspects and possible ways of avoiding or minimising information bias will also be discussed.

SEARCH STRATEGY AND SELECTION CRITERIA: ¬

Articles for this review were identified through searches in PubMed with the following search words: “recall bias”, “respond bias”, “bias”, “observer variation”, “digit preference”, “social desirability”, “information bias”, “attribution bias”, “comprehension bias”, “self-report bias” and “mental recall”, combined with the medical subjects headings “alcohol drinking”, “infant nutrition physiology”, “breast feeding”, “disease transmission, vertical”, “HIV infections/transmission”. A limited search was conducted in PsychInfo to complement some psychological aspects. Approximately 1400 articles were considered on the basis of abstract, title and content, from which 53 articles were selected based on their relevance for the review to illustrate the different aspects of information bias.

TOPICS ASSOCIATED WITH POTENTIAL ERRORS

In this presentation of information bias, we have focused on the following areas: recall period, selective recall, social desirability, interview situation and interviewing tools, question phrasing, answer alternatives and digit preference

(table 1).

Figure 1

Table 1: Summary of Different Aspects of Information Bias due to Self-reporting Focusing on Important Concepts, Problem Identification and Possible Solutions.

	Concepts	*Identification and possible solutions*
Recall period	– Accuracy decreases over time – Long recall period will often lead to interpretation that researcher is interested in more major events – Retrospective reporting seems to deviate towards normal behaviour and current behaviour	– When possible, skewed reporting can be adjusted for using validating data – Validating studies can sometimes influence behaviour and further complicate interpretation – Try to identify amount of day-to-day variation – Length of recall period must be weighted against representativeness
Selective recall	– Some events are more easily remembered than others – Difficult to predict which groups that will differ in reliability – Information contrasting earlier experiences is often better memorised than congruent information – When people are asked to give reasons, they tend to seek backwards in the causal chain until they discover factors which look familiar to serve as explanation	– Unhealthy behaviour more often underreported – Deaths and major events better recalled – Factors recognised as risk factors better recalled – Inconsistent reporting can be due to re-evaluation of own actions and behaviour over time – Minnesota Multiphasic Personality Inventory can be used as a tool for identifying over-reporters
Social desirability	– Reporting false information – Avoid mentioning information – Modifying information	– Especially important when studying sensitive topics – Can be measured using the Marlowe-Crowne Social Desirability Scale (SDS) – 'Audio computer-assisted self-interviewing' and 'randomized response interview' reduce socially desirable responding
Interview situation and tools	– Interviewers identify similar participants more precisely than dissimilar participants	– Interview situation can be framed to promote openness
Question phrasing and answer alternatives	– Wording can alter responses – Interpretation of simple questions can be dissimilar – Respondents have preferences for rounded and approximated answers – Respondents try to put the questions into a context, giving answers they perceive as sufficiently informative for the research	– Avoid a question wording that influences answers – Attempt to be as clear as possible without choosing bothersome phrasing – Few answer alternatives increase tendency to omit question and can bias reporting – Avoid expression loaded with prejudice – Using a question vignette to underline an open minded view – Scales valid for all participants are preferable
Digit preference	– Respondents have preferences for rounded answers – Last digit preference is often 0 and 5	– Tendency to report a digit preference number is culturally different
Bias	– Bias can be proportional (relative to its size) or fixed (similar in absolute values).	– Can be detected using a modified McNemar test in categorical variables and ordinary linear product regression for continuous variables – Performing piloting studies aiming to compare different measurements may be a key step for validating – When comparing measurement methods it is important that the methods are independent

RECALL PERIOD

The period of time over which respondents are required to remember past events in order to answer a question can be denoted the "recall period"; e.g. one month when respondents report behaviour during the last month or behaviour that took place one month ago. The recording period is the length of time covered by a question and will overlap with the recall period when the time period covered includes the present. These periods need careful consideration in the design of a study.

Recall usually deteriorates with time. A Swiss study that compared recall of drinking events reported one week previously with those during the past day showed that one-third of drinking events were not reported after a week (7). This tendency was rather constant; recall decreased progressively with increasing time since the drinking event (8). Sporadic drinkers underreported more than regular drinkers (7).

In cross-sectional studies, increasing the recall period decreases the accuracy of recall but may also make the period more representative. When day-to-day variation is marked, a prolonged recording period may be required (9, 10), but this may be counteracted by less accurate reporting owing to fatigue among the participants at the end of a prolonged recording period (11). Representativeness and recall accuracy need to be well balanced.

When mothers are asked retrospectively in breastfeeding studies about breastfeeding duration, the reports become increasingly inaccurate with increasing time since cessation (12, 13). Surprisingly, one American study reported the same mean breastfeeding duration in 40-year recall and prospective data (14). Mothers who had breastfed their infants for longer systematically underreported their breastfeeding duration; while in contrast, those who had breastfed for a shorter time over-reported the duration. Mothers with many children reported duration more accurately than those with fewer children, possibly indicating that memory was enhanced by multiple similar experiences. South African studies focusing on exclusive breastfeeding have indicated that it is difficult to obtain reliable measures from retrospective recall, and that prospective one-week frequency recall gives more valid and representative information (15).

Behavioural parts of our daily life such as energy intake from food consumption can sometimes be difficult to estimate. One approach to address this challenge has been to compare reported energy intake through dietary recall with assessed energy expenditure (16). Different methods using e.g. doubly-labelled water, urine nitrogen analysis and assessment of resting metabolic rate and physical activity can all help in estimating total energy expenditure. With such strategies, calibration may be possible, adjusting the

scales (17). Most such studies have revealed roughly 25% underreporting of energy intake, while obese persons underreport even more (16). To complicate this still further, food items that are considered "bad for health"(e.g. fat containing) were selectively more strongly underreported than other macronutrients. Underreporting of food intake varies among individuals (18). A person who underreports food intake on one occasion is significantly more likely to underreport it subsequently than a person who does not (19). Food intake also seems to be more underreported among adults and adolescents than among children (20).

Sub-optimal recall is particularly problematic if different groups have different recall periods, which may be the situation in a case-control study. If the control period proceeded the case period, the accuracy of recall of the former will often be lower than that of the latter. In some circumstances an ambi-directional design can improve the situation; in which the control period can be both before and after an event defining the case period. Prospective measurements must also be analysed with caution because many study participants modify their behaviour during observation periods, e.g. reduces food intake. A recall period of one year will often lead to an interpretation among the study participants that the researcher is more interested in major events than a recall period of one week, where minor events are also likely to be reported (21).

SELECTIVE RECALL

Some events are better recalled than others. You may memorise the dog that scared you to death two years ago, but maybe not the one you saw walking in the park yesterday. An East-African study investigated recall of reasons for child deaths (22). Easily recognisable clinical pictures such as kwashiorkor, measles and tetanus were well recalled, while gastrointestinal symptoms, coughing and nasal flaring were poorly recalled. A study investigating the reasons for stopping breastfeeding found that some reasons such as the mother having to work, inflammation of the breast or death of child were reliably reported, but others such as breastfeeding being inconvenient or the child being ill were not (13, 23).

Retrospective recall inaccuracies are often reported in a way that increases agreement with current behaviour (24). Research into sexual behaviour indicates that frequently-occurring behaviour seems to be less accurately recalled than less frequent behaviour (25), whereas in the area of nutrition, extremes such as "consume daily" and "almost never consume" seem to be best preserved (20). Qualitative information such as the type of food item consumed is in general more accurate than quantitative information such as the amount of the item consumed.

Bias and false associations may emerge when different groups have different recall reliabilities (26). Educated mothers in Brazil tended to misclassify their breastfeeding duration upwards (to a longer duration) twice as often as downwards (12). It has been hypothesised that such misclassification tends towards the behaviour that was attempted. In a Malaysian study, mothers were interviewed twice with a 12-year time gap, and a similar distribution of reasons for stopping breastfeeding emerged, though the consistency was fairly low because a high proportion of the subjects changed their reports. Though, it is important to keep in mind that inconsistent reporting does not necessarily mean unreliable reporting. Reporting other factors than initially can be due to re-evaluation of one's own actions and behaviour over time (23). When people are asked to give reasons, they seem to search backwards in the causal chain until they discover factors that look familiar to serve as explanations (26).

Information that contrasts with earlier information is often better memorised than information that is congruent with preceding experience (27). Recall bias can also be suspected when the respondents report that they suspect that a certain factor is a risk factor (26). Persons who have experienced a specific outcome are more likely to recall suspected factors. A tool that can be used to validate reporting is Minnesota Multiphasic Personality Inventory (MMPI). This scale investigates the respondents' prior exposure to "fake" risk factors. Respondents who over-report these factors are more likely to over-report true risk factors too, which makes it possible to adjust for the bias.

SOCIAL DESIRABILITY

Social desirability has been stated to consist of self deception, "the conscious tendency to see oneself in a favourable light" and impression management, "the conscious presentation of a false front, such as deliberatively falsifying test responses to create a false front" (26). Socially desirable responding can have three different manifestations: reporting incorrect information, omitting information or altering the magnitude of the reported information. This is often linked to the fear that information will be revealed publicly. An American study focusing on drug abuse showed that under-reporters experienced more socio-desirability

pressure than those who did not under-report (28). Some study participants over-reported their drug abuse, but this was associated with memory difficulties. A review evaluating doctors' clinical adherence to guidelines indicated that, in nearly all studies, clinicians overestimated their adherence (29). Older participants are reported to give more socially desirable responses than young ones, while income and socio-economic status are inversely correlated with socially desirable responding (30). Some research topics such as sex work are particularly sensitive (31), but even in food consumption research, social desirability bias is clearly evident (32, 33). The tendency to seek praise reported in nutritional research can be classified as social approval bias and is related to social desirability (33, 34).

Social desirability can be measured using the Marlowe-Crowne Social Desirability Scale (SDS), in which high values indicate that respondents are more reluctant to disclose unpopular beliefs or behaviour (30). This scale contains 33 strong statements that tempt respondents to deny statements such as "I have never intensely disliked anyone", because few can honestly respond to these questions in a desirable way. A high score reflects an individual who self-represents as acting socially desirably, but we must keep in mind that the person might also behave accordingly. Thus, adjusting for SDS can undermine real variance in the data in addition to removing some of the information bias. Similarly, a Martin-Larsen Approval Motivation-score can be used to measure social approval (33).

INTERVIEW SITUATION AND INTERVIEWING TOOLS

Framing the interview situation to promote openness is important. A British study investigating drinking habits found that when the wife of a participant was present at the interview, the participant reported lower alcohol consumption than when the wife was absent (35). It has also been shown that interviewers who are similar to the participants are more readily able to detect behaviour patterns than interviewers who are dissimilar. Regular drinkers identify regular drinkers more precisely, while abstainers identify abstainers accurately (35). A Japanese study indicated that age difference between interviewer and interviewee strongly affected the bias (36). A small age difference resulted in a strong increase in bias, in contrast to a difference of many years between the interviewer and the interviewee. On the other hand, a study of HIV-stigma in Ghana and Nigeria indicated that a smaller age difference increased the accuracy of the reporting (31). Health workers were reported to be well respected, but some study participants reported less of the stigmatised behaviour to their health workers, probably to 'please' their relationship with the health worker.

Tools have been developed to improve the validity of data collected on socially sensitive topics. Audio computer-assisted self-interviewing (ACASI) is a software system that conducts an automated interview with a study participant by giving instructions on how to report responses. This technique has been tried in some studies related to sensitive topics such as HIV-stigma in Kenya (37). ACASI detected sensitive items substantially better than an ordinary interview. In such situations, participants have preferred ACASI over standard interviews. Stigmatised behaviour was also reported more frequently with an ACASI-situation than in a face-to-face interview in a clinic for sexually transmitted diseases (38). Persons who opted to skip questions in the computer-assisted interview tended to deny stigmatised behaviour in the face-to-face interview. Older persons were more sceptical of computer-assisted interview techniques than younger people (31).

A less technological approach is the randomized response interview technique (39). Using e.g. a die throw to introduce a random factor can make responses less vulnerable to socially desirable responding while making it possible to control statistically for the random factor. Participants rolling a specific number on a die not seen by the interviewer are instructed to give the socially desirable response regardless of the true answer. Consequently, true positive responses to less socially desirable questions can be given without the interviewers being able to know whether the individual gave them because of the randomisation or because they were true. The randomized response technique was also successfully tried in a low-educated setting in Ethiopia to estimate the incidence of induced abortions, demonstrating that advanced techniques are not necessarily culturally or educationally dependent (40).

A Danish study compared the use of questionnaires and interviews to investigate binge drinking during pregnancy (41). They concluded that interviews gave a higher detection rate of binge drinking, a higher proportion providing information and better consistency. They hypothesised that many participants made more effort in an interview situation than in a questionnaire. In contrast, a Japanese study comparing the same methods of data collection reported substantial underreporting of habits such as alcohol drinking and psychological stress in interviews compared to self-

administered questionnaires (36). Behaviour that is recognized as risky seemed to be underreported most. In a Danish study about musculoskeletal and gastrointestinal symptoms, information obtained through a questionnaire was compared with a telephone interview (42). Infrequent symptoms were often not mentioned in the telephone interview, while the questionnaire was better able to identify them (table 2).

Figure 2

Table 2: Comparison of Different Data Collecting Methods in Terms Recall Efforts, Accuracy, Question Understanding and Socially Desirable Responding

	Recall efforts and accuracy	*Question understanding*	*Socially desirable responding*
Questionnaire	– Less effort is made to choose accurate answer	– Make less room for individual differences	– Less vulnerable
Interview with open questions	– More effort is made to increase accuracy – More missing answers. Fewer false positives than prompting interview	– Makes room for clarification	– Vulnerable
Prompting interview	– More effort is made to increase accuracy – Increased reporting of recalled information – More false positive answers than open questions	– Makes room for clarification	– Vulnerable
Telephone interview	– Less effort is made to choose accurate answer alternative	– Some room for clarification	– Vulnerable
Audio computer-assisted self-interviewing	– Slightly more missing answers than interview	– Depending on structure – Can give pre-programmed instructions	– Less vulnerable

QUESTION PHRASING

We should keep in mind that respondents try to put the questions into a context, giving answers they perceive as sufficiently informative for the research (21). Therefore, even the affiliation of the researchers can influence the respondents' reports. Information that is considered obvious or already known is less likely to be reported. Respondents also tend to report consistent information in line with what they have reported to similar questions (43). Phrasing a question so that it is loaded with prejudice can also bias the answers. Using a question vignette to underline absence of prejudices can improve the situation (10).

Questions with alternative answers can give a distribution different from questions without fixed alternatives (21). Respondents are less likely to give answers not included in a fixed list of alternatives and will often revise their interpretation of the question if they have considered other alternatives than those suggested. It has been reported that prompted questions give a higher detection rate with more true positives than open questions, but also more false positive answers (22).

The sequence of the questions may matter (21). Responses to later questions can be influenced by earlier questions, which may be seen as a contrasting reference. As an example, questions about life happiness may be seen in relation to happiness in marriage or as independent of marriage, depending on the question structure. Wording can also alter responses substantially. A group of students were asked questions about velocity estimation and whether broken glass was seen in a short film (44). A third of the students "observed" broken glass, which was not evident, when the question was phrased "About how fast were the cars going when they smashed into each other?" In contrast, fewer than one in six made this "observation" when "contacted" was chosen instead of "smashed into." Velocity was also estimated to be higher when "smashed into" was utilized in place of "contacted." The reason is probably that memory information is a complex mixture of perceptual and external information.

Different respondents may interpret a question differently (45). Interestingly, seemingly simple questions elicited very different interpretations and different responses. Questions such as "Have you smoked at least 100 cigarettes in your entire life?" were sometimes interpreted as including only finished cigarettes; sometimes as any cigarettes from which a puff had been taken, irrespective of inhalation, or only cigarettes inhaled; only cigarettes, or cigarettes and cigars. Only half the respondents assumed the same definition. Asking the question with a definition resulted in a change in reporting among a proportion of the respondents. Respondents who were instructed to ask for clarification when needed, rarely asked for it. Training the interviewers to recognise uncertainty among the respondents is another strategy. "Um", "uh", long pauses, restarts and repairs can be signs of uncertainty. It has been reported that this is difficult, with the interviewers seldom offering to clarify a question. Making uniform definitions about all questions is difficult and will often make the interview time-consuming and bothersome. Putting the research into its context and recognising that question structure influences how the respondents report, can reduce some possible causes of interpretation error (21).

ANSWER ALTERNATIVES

From alcohol consumption studies, the numbers of alternatives seem to be important (10). Responders who in reality would fit into the highest–consumption alternative often tend to decrease their reporting. Having a wider range of alternative answers than expected can minimise this

potential bias (10). Questions with few alternatives are more often skipped, possibly because it is more difficult to find an alternative that reflects the situation of the participant (42). Furthermore, participants are in general more reluctant to report drinking 52 times each year than once a week (46). A rating scale with low reference alternatives leads to lower reports than one with higher reference alternatives, probably because it is expected that the median alternative answer reflects "average" behaviour (21). Alternatives such as "sometimes" or "frequently" will be relative to the respondents' subjective standards and are the worst option. Using open questions specifying the measurement unit without giving fixed alternatives can reduce the likelihood of influencing the answers.

A rating scale from -5 to 5 can give a respondent the impression of a bipolar dimension (failure to success) while a scale from 0 to 10 is interpreted as unipolar (absence of success to success) (21). This is a challenge for the researcher, who needs to consider this when discussing the findings.

When interviewing children, it is important to take age into consideration. Children around two years old have a tendency to answer "yes" to most questions regardless of the true answer (47). With increasing age, this effect decreases. When children of around 4 to 5 years of age are asked a question they do not understand, they have a tendency to answer "no" instead of "I don't know". Consequently, avoiding closed yes/no questions when interviewing young children can prevent scientific headaches.

DIGIT PREFERENCE

Respondents often have preferences for rounded and approximated answers, e.g. 0 and 5 being the last digit, to which we refer as 'digit preference'. This trend may be seen in most areas but has been described in greater detail for reported measurements of blood pressure (48). The tendency to report a digit preference number is culturally variable. Malay mothers were more likely to report rounded answers than Indian mothers (49).

BIAS AND ESTIMATION OF BIAS

Bias is a systematic error and can be either proportional or fixed (50). A fixed bias indicates that the difference is similar in absolute value for high and low measurements, while proportional bias means that the difference is relative to the size of the measurement. Proportional bias will be higher in absolute terms for higher values of the variable studied. There are different strategies to estimate the extent of disagreement in epidemiological research. It is essential that bias is quantified, but this should not be performed by comparing means measured with different methods, or by correlation analysis (4, 51). An important point when comparing measurement methods is to ensure that the methods are independent and the data preferably gathered by different observers (4).

Bias in categorical data can be detected using a modified McNemar test to compare matched observations (50, 52). This test compares discordant observations with concordant ones. The following example illustrates the use of the modified McNemar test (Table 3). Observations are coded into e.g. 4 different categories and are compared with validation data (observation 2). Numbers with concordant observations are listed on the diagonal of the matrix in bold font. Discordant observations are in either the upper-right or the lower-left corner. This will be evaluated using chi-square (2) with one degree of freedom.

Figure 3

Table 3: Example Illustrating the use of the Modified McNemar Test

		Observation 2				
	Category	I	II	III	IV	*Total*
Observation 1	I	**8**	3	2	2	*15*
	II	2	**7**	2	1	*12*
	III	1	2	**3**	1	*7*
	IV	1	1	2	**7**	*11*
	Total	*12*	*13*	*9*	*11*	**45**

$$\chi^2 = \frac{((\text{sum of upper right observations}) - (\text{sum of lower left observations}))^2}{(\text{sum of upper right observations}) + (\text{sum of lower left observations})}$$

$$\chi^2 = \frac{((3+2+2+2+1+1)-(2+1+2+1+1+2))^2}{((3+2+2+2+1+1)+(2+1+2+1+1+2)} = 0.2 \qquad p = 0.65$$

In the example, the p value of 0.65 indicates that no significant bias could be found. It is important to highlight that this test does not guarantee detection of bias even when present. When the modified McNemar test is statistically significant, it does not necessarily mean that the bias is clinically relevant. However, a positive test is a good reason to read the material through very critical glasses.

As for categorical variables, different forms of factor analysis can also be used as statistical tools to diagnose bias

for continuous variables (43). The validity of these methods has been debated. Regression analysis is one method for comparing different measurements to look for bias. Ordinary linear squares regression is not the best choice because the method assumes that the error is random, which is not the case when we deal with bias. Least product regression is a better choice for this purpose, or weighted least products regression when the spread of values is not constant (50, 53). This method has been well described (53).

CONCLUSION

Identifying bias is often possible and should be a priority in epidemiological research. Being prepared from the beginning to avoid many of the pitfalls is often the best solution. This preparation includes choosing an appropriate way to collect the data, looking for strategies to validate the collected data, selecting a recall period to balance representativeness with recall accuracy, defining the questions to be asked and phrasing questions with matching answer alternatives carefully. For sensitive topics, there may also be a need to compensate for socially desirable responses using a scale identifying this pattern. Performing pilot studies with the aim of comparing different measurements may be a key step in validation. Searching for disagreement and bias in comparing with validation data is an important next step to increase the reliability of epidemiological research.

FUNDING

Funding was provided by the involved institutions and by a research grant from the Research Council of Norway, project no 172226 “Focus on Nutrition and Child Health: Intervention Studies in Low-income Countries.”

ACKNOWLEDGEMENTS

We would like to thank Randi Bolstad and Therese Skarås Skagen for assisting in the search for articles.

References

1. Last JM, Abramson JH, International Epidemiological Association. A Dictionary of epidemiology. New York: Oxford University Press, 1995.
2. Giovannucci E, Stampfer MJ, Colditz GA, et al. Recall and selection bias in reporting past alcohol consumption among breast cancer cases. Cancer Causes Control 1993;4:441-8.
3. Pearce N, Checkoway H, Kriebel D. Bias in occupational epidemiology studies. Occup Environ Med 2007;64:562-8.
4. Bland JJM, Altman DDG. Measuring agreement in method comparison studies. Statistical methods in medical research 1999;8:135-60.
5. Delgado-Rodriguez M, Llorca J. Bias. J Epidemiol Community Health 2004;58:635-41.
6. Maclure M, Schneeweiss S. Causation of bias: the episcope. Epidemiology 2001;12:114-22.
7. Gmel G, Daeppen JB. Recall bias for seven-day recall measurement of alcohol consumption among emergency department patients: implications for case-crossover designs*. J Stud Alcohol Drugs 2007;68:303-10.
8. Ekholm O. Influence of the recall period on self-reported alcohol intake. Eur J Clin Nutr 2004;58:60-3.
9. Willett WC. Nutritional epidemiology issues in chronic disease at the turn of the century. Epidemiol Rev 2000;22:82-6.
10. Embree BG, Whitehead PC. Validity and reliability of self-reported drinking behavior: dealing with the problem of response bias. J Stud Alcohol 1993;54:334-44.
11. Berg C, Jonsson I, Conner MT, et al. Sources of bias in a dietary survey of children. Eur J Clin Nutr 1998;52:663-7.
12. Huttly SR, Barros FC, Victora CG, et al. Do mothers overestimate breast feeding duration? An example of recall bias from a study in southern Brazil. Am J Epidemiol 1990;132:572-5.
13. Gillespie B, d'Arcy H, Schwartz K, et al. Recall of age of weaning and other breastfeeding variables. Int Breastfeed J 2006;1:4.
14. Promislow JH, Gladen BC, Sandler DP. Maternal recall of breastfeeding duration by elderly women. Am J Epidemiol 2005;161:289-96.
15. Bland RM, Rollins NC, Solarsh G, et al. Maternal recall of exclusive breast feeding duration. Arch Dis Child 2003;88:778-83.
16. Westerterp KR, Goris AH. Validity of the assessment of dietary intake: problems of misreporting. Curr Opin Clin Nutr Metab Care 2002;5:489-93.
17. Kaaks R. Uses and limitations of statistical accounting for random error correlations, in the validation of dietary questionnaire assessments. Public health nutrition 2002;5.
18. Pryer JA, Vrijheid M, Nichols R, et al. Who are the 'low energy reporters' in the dietary and nutritional survey of British adults? Int J Epidemiol 1997;26:146-54.
19. Black AE, Cole TJ. Biased over- or under-reporting is characteristic of individuals whether over time or by different assessment methods. J Am Diet Assoc 2001;101:70-80.
20. Livingstone MBER. Issues in dietary intake assessment of children and adolescents. The British journal of nutrition 2004;92:S213.
21. Schwarz N. HOW THE QUESTIONS SHAPE THE ANSWERS. American psychologist 1999;54:93.
22. Snow R, I Basto De Azevedo, D Forster, S Mwankuyse, G Bomu, G Kassiga,C Nyamawi, T Teuscher and K Marsh. Maternal Recall of Symptoms Associated with Childhood Deaths in Rural East Africa. International Journal of Epidemiology 1993;22:677-83.
23. Kuate-Defo B, DaVanzo J. Reliability of reasons for early termination of breastfeeding: application of a bivariate probability model with sample selection to data from surveys in Malaysia in 1976-77 and 1988-89. Popul Stud (Camb) 2006;60:83-98.
24. Lee MM, Whittemore AS, Lung DL. Reliability of recalled physical activity, cigarette smoking, and alcohol consumption. Ann Epidemiol 1992;2:705-14.
25. Graham CA, Catania JA, Brand R, et al. Recalling sexual behavior: a methodological analysis of memory recall bias via interview using the diary as the gold standard. J Sex Res 2003;40:325-32.
26. Raphael K. Recall bias: a proposal for assessment and control. Int J Epidemiol 1987;16:167-70.
27. Hamilton DL, Grubb PD, Acorn DA, et al. Attribution difficulty and memory for attribution-relevant information. J

Pers Soc Psychol 1990;59:891-8.
28. Johnson T, Fendrich M. Modeling sources of self-report bias in a survey of drug use epidemiology. Ann Epidemiol 2005;15:381-9.
29. Adams AS, Soumerai SB, Lomas J, et al. Evidence of self-report bias in assessing adherence to guidelines. Int J Qual Health Care 1999;11:187-92.
30. Welte JW, Russell M. Influence of socially desirable responding in a study of stress and substance abuse. Alcohol Clin Exp Res 1993;17:758-61.
31. Guest G, Bunce A, Johnson L, et al. Fear, hope and social desirability bias among women at high risk for HIV in West Africa. J Fam Plann Reprod Health Care 2005;31:285-7.
32. Hebert JR. Social Desirability Bias in Dietary Self-Report May Compromise the Validity of Dietary Intake Measures. International Journal of Epidemiology 1995;24:389.
33. Hebert JR, Ma Y, Clemow L, et al. Gender differences in social desirability and social approval bias in dietary self-report. Am J Epidemiol 1997;146:1046-55.
34. Miller TM, Abdel-Maksoud MF, Crane LA, et al. Effects of social approval bias on self-reported fruit and vegetable consumption: a randomized controlled trial. Nutr J 2008;7:18.
35. Crawford A. Bias in a survey of drinking habits. Alcohol Alcohol 1987;22:167-79.
36. Okamoto K, Ohsuka K, Shiraishi T, et al. Comparability of epidemiological information between self- and interviewer-administered questionnaires. J Clin Epidemiol 2002;55:505-11.
37. Waruru AK, Nduati R, Tylleskar T. Audio computer-assisted self-interviewing (ACASI) may avert socially desirable responses about infant feeding in the context of HIV. BMC Med Inform Decis Mak 2005;5:24.
38. Ghanem KG, Hutton HE, Zenilman JM, et al. Audio computer assisted self interview and face to face interview modes in assessing response bias among STD clinic patients. Sex Transm Infect 2005;81:421-5.
39. Rittenhouse BE. Respondent-specific information from the randomized response interview: compliance assessment. J Clin Epidemiol 1996;49:545-9.
40. Chow LP, Gruhn W, Chang WP. Feasibility of the randomized response technique in rural Ethiopia. Am J Public Health 1979;69:273-6.
41. Kesmodel U, Frydenberg M. Binge drinking during pregnancy--is it possible to obtain valid information on a weekly basis? Am J Epidemiol 2004;159:803-8.
42. van Ooijen M, Ivens UI, Johansen C, et al. Comparison of a self-administered questionnaire and a telephone interview of 146 Danish waste collectors. Am J Ind Med 1997;31:653-8.
43. Podsakoff PM, MacKenzie SB, Lee JY, et al. Common method biases in behavioral research: A critical review of the literature and recommended remedies. Journal of Applied Psychology 2003;88:879-903.
44. Loftus EF, Palmer JC. Reconstruction of automobile destruction: An example of the interaction between language and memory. Journal of Verbal Learning and Verbal Behavior 1974;13:585-9.
45. Suessbrick AL. Different respondents interpret ordinary questions quite differently. Journal of business & economic statistics 2000.
46. Dawson DA. Methodological issues in measuring alcohol use. Alcohol Res Health 2003;27:18-29.
47. Fritzley VHVH, Lee KK. Do young children always say yes to yes-no questions? A metadevelopmental study of the affirmation bias. Child development 2003;74:1297-313.
48. Hessel PA. Terminal digit preference in blood pressure measurements: effects on epidemiological associations. Int J Epidemiol 1986;15:122-5.
49. Haaga JG. Reliability of retrospective survey data on infant feeding. Demography 1988;25:307-14.
50. Ludbrook J. Statistical techniques for comparing measurers and methods of measurement: a critical review. Clin Exp Pharmacol Physiol 2002;29:527-36.
51. Bland JM, Altman DG. Applying the right statistics: analyses of measurement studies. Ultrasound Obstet Gynecol 2003;22:85-93.
52. Ludbrook J. Detecting systematic bias between two raters. Clin Exp Pharmacol Physiol 2004;31:113-5.
53. Mullineaux DR, Barnes CA, Batterham AM. Assessment of Bias in Comparing Measurements: A Reliability Example. Measurement in Physical Education and Exercise Science 1999;3:195-205.

Author Information

Lars T. Fadnes
Centre for International Health, University of Bergen, Norway

Adam Taube
Department of Information Science, Statistics, Uppsala University, Sweden

Thorkild Tylleskär
Centre for International Health, University of Bergen, Norway