



UNIVERSITY OF
TECHNOLOGY SYDNEY

UTS:CHERE

**DIARIES OR QUESTIONNAIRES FOR
COLLECTING SELF-REPORTED
HEALTHCARE UTILISATION AND
PATIENT COST DATA?**

A review of the literature

Patricia Kenny, Ajsa Mahmic, Emily Lancsar,
Rob Anderson, Madeleine King, Jane Hall

Project Report 20
April 2003



The Centre for Health Economics Research and Evaluation (CHERE) was established in 1991. CHERE is a centre of excellence in health economics and health services research. It is a joint Centre of the Faculties of Business and Nursing, Midwifery and Health at the University of Technology, Sydney, in collaboration with Central Sydney Area Health Service. It was established as a UTS Centre in February, 2002. The Centre aims to contribute to the development and application of health economics and health services research through research, teaching and policy support.

CHERE's research program encompasses both the theory and application of health economics. The main theoretical research theme pursues valuing benefits, including understanding what individuals value from health and health care, how such values should be measured, and exploring the social values attached to these benefits. The applied research focuses on economic and the appraisal of new programs or new ways of delivering and/or funding services.

CHERE's teaching includes introducing clinicians, health services managers, public health professionals and others to health economic principles. Training programs aim to develop practical skills in health economics and health services research.

Policy support is provided at all levels of the health care system by undertaking commissioned projects, through the provision of formal and informal advice as well as participation in working parties and committees.

ACKNOWLEDGEMENTS

The authors are grateful to Liz Chinchon who completed the literature search. The project was supported by the Cooperative Research Centre for Asthma. The views expressed in this paper and any errors are the responsibility of the authors.



ABSTRACT

The literature comparing diaries and questionnaires was reviewed in order to identify the most appropriate method of collecting patient self-reported data, on health service utilisation and out-of-pocket costs, for a longitudinal study. Nine published studies met the review inclusion criteria; four compared the diary method with a self-completed questionnaire and five with an interviewer administered questionnaire. None of the eligible studies measured patient costs, and only two measured some aspects of health service utilisation. Most of the studies reported higher response rates for questionnaires than for diaries, and there was some evidence of selection bias. There was a tendency to report more symptoms, symptom intensity or health care utilisation by questionnaires compared to diaries, and compared to physician reports (included in only two studies). The review provides some information about the two approaches for collecting self-reported data, but does not provide sufficient evidence to favour either approach.

CONTENTS

Introduction	2
<hr/>	
Methods	3
Inclusion criteria	3
<hr/>	
Results	4
Identification of studies	4
Respondent cooperation	4
Accuracy	5
<hr/>	
Discussion	7
Table 1: Summary of studies comparing diary and self-completed questionnaire data collection methods	9
Table 2: Summary of studies comparing diary and interviewer-administered questionnaire data collection methods	10
<hr/>	
References	11



INTRODUCTION

The measurement of health service utilisation and individual out-of-pocket costs for the economic evaluation of health services often relies on self-reported patient data, because of the limitations of administrative data. The two most common tools used to collect this type of self-reported data are questionnaires and diaries. Both approaches have advantages and disadvantages in terms of accuracy, compliance, response rates and data collection costs.

Questionnaires are highly structured data collection instruments¹ and may be either self-completed or interviewer-administered. When asking about the experiences of individuals, questionnaires refer to a specified time period in the past. By contrast, diaries may be structured, semi-structured or unstructured and are always self-completed. They are intended to be kept over a given period and completed intermittently, either at specified intervals or when particular events occur (rather than at a single sitting). Thus, a structured diary may be seen as a self-completed questionnaire in which the respondent records events as they occur or soon after (usually within 24 hours).

Diaries used for the collection of health care utilisation and patient cost data would usually take a structured approach, comprising questions about specific types of utilisation and costs, which require particular types of responses (such as frequency, date, duration or cost). Some open ended descriptive items may also be included, to identify additional types of services or expenditure. In this context, a questionnaire would ask about the costs of health care used within a specified period in the past, while a diary would be completed as the costs occur throughout the period.

One of the key differences between the diary and questionnaire approaches relates to the impact of memory recall and the data collection period on data quality. The data period is usually more restricted for the questionnaire method, because it relies on respondents' recollection of events for the entire data collection period. The recommended recall period for questionnaire data is between two weeks and one month for events of low salience (ordinary or routine).² For many studies, particularly in the case of chronic illness, a longer period of data collection is required in order to capture the relevant costs and consequences. With the diary method, respondents are asked to record events for shorter time blocks (eg daily) within the data collection period. The duration of diary keeping can be as long as required, allowing for a longer data collection period, without the recall problems of the questionnaire method. However, the burden on respondents of diary keeping for prolonged periods may have a negative impact on response rates and compliance with study protocol.

Memory recall problems may lead to over or under estimation. Relatively ordinary or routine events are likely to be underreported, because the respondent forgets them. Unusual or high impact events are likely to be remembered, but details such as the date of the event may not. This can lead to over-reporting because prior events are remembered as happening within the reporting period, or underreporting because events from within the reporting period are remembered as happening earlier.²

Another important difference between diary and questionnaire methods is the burden on respondents, which is greater for the diary method because of the frequent recording. Diary methods require investment in procedures to encourage respondents to complete diaries and to maximise data quality³, and can be susceptible to lower response rates from particular population subgroups such as those with low levels of education, low socio-economic status and those in poor health⁴ (selection bias). There is also evidence that paper diaries are often completed after the actual day of reporting⁵, which could introduce the same memory recall issues attributed to questionnaires. In addition, respondents tend to report fewer events as the duration of diary keeping increases, suggesting fatigue effects (declining respondent interest).³

This paper reports a review of the literature comparing diaries and questionnaires, which was conducted in order to identify the most appropriate method of collecting patient reported data for a longitudinal study of asthma costs (the Cost of Asthma Study). The results of the literature review are presented in terms of respondent cooperation and accuracy of the data collection method.

METHODS

Identification of studies

A literature review of studies comparing diaries and questionnaires was conducted using the Econlit, Pre-Medline, Medline, PsychInfo and Sociological Abstracts databases. The search criteria were articles published between 1980 to 2001 in English, with the following keywords (MeSH): "diary or diaries, questionnaire or questionnaires, and comparison".

Inclusion criteria

The inclusion criteria were determined according to the requirements of the Cost of Asthma Study, which did not have the capacity to use electronic diaries or interviewer-administered questionnaires. The data to be collected by patient report for that study included: health service use, medication use, effects of asthma on activities, patient out-of-pocket costs for products or services and patient and carer time lost from work and other activities. The nature of these data items meant that, if it used the diary method, the Cost of Asthma Study would require diary keeping less frequently than daily. Thus, the inclusion criteria for the review were comparative studies that involved reporting of:

1. Health service utilisation,
2. Cost or personal expenditure,
3. Symptom-related events, where the frequency of those events is usually less often than daily,
4. Time use, but only if this includes reporting of time spent at work (paid or unpaid), school or involved in caring activities.

A data collection form was used to record information on the study type and duration, sample size, follow-up procedures, incentives for completion, target group, measured events, recall interval, and results of the study.



RESULTS

Identification of studies

Over 100 studies were identified using the MeSH headings, however only nine studies were included after: a) excluding all studies that did not meet the inclusion criteria; and b) excluding studies that compared electronic diary entries with questionnaires. Of the nine studies, four compared diaries with self-completed questionnaires, and five compared diaries with interviewer-administered questionnaires (included but examined separately). Because of the small number of comparative studies using self-completed questionnaires, those using interviewer-administered questionnaires were included in case they identified issues that were also relevant for comparisons with self-completed questionnaires.

The studies using self-completed questionnaires measured headache frequency and severity (three studies⁶⁻⁸) or menstrual symptoms (one study⁹). We found no studies of health service utilisation or costs which compared diaries with self-completed questionnaires. The studies using interviewer-administered questionnaires measured symptoms (three studies¹⁰⁻¹²), health problems and GP visits (one study¹³), and health care utilisation, pain management and impact on work (one study¹⁴).

Respondent cooperation

Diary versus self-completed questionnaire

The four studies reviewed generally reported more respondents with missing data for the diary method than for the questionnaire method (Table 1). However, the questionnaire was self-completed at interview for the study by Woods et al⁹ and in the classroom for the study by van den Brink et al.⁸ These administration conditions would be expected to produce higher response rates than postal questionnaires or diaries. The two studies using postal questionnaire methods^{6,7} used similar follow-up protocols, but produced different response rates. The first of these studies⁶ reported a lower response rate overall (58%) with a small difference between the two methods for incomplete data (2% withdrew from the study, 22% had incomplete diary data and 18% had incomplete questionnaire data). The later study⁷ reported 77% with usable data; 5% withdrew from the study, 16% had incomplete diary data, 0.5% had incomplete questionnaire data and 2 outliers (1%) were excluded. The data collection period for both of these studies was three months, with the diaries completed daily and returned weekly. The studies also used reminder phone calls and reimbursed subjects with \$5.00 (US) for each completed diary.

None of the studies compared diaries and questionnaires for selection bias. Woods et al⁹ found diary respondents to be slightly younger with higher incomes than non-respondents, but only had demographic data from nine of the 19 non-respondents and did not examine the characteristics of questionnaire respondents and non-respondents. Stewart et al⁶ found study respondents who completed both the diary and the questionnaire to be older and more highly educated than the sample of migraine sufferers from which they were drawn. However, there was no information about the extent of this selection bias for diaries and questionnaires separately. Stewart et al⁷ reported similar characteristics between eligible migraine sufferers and study respondents who completed both the diary and the questionnaire, in terms of gender, age and race, but did not report education or income.

Two studies attempted to examine the extent of delayed recording of diary entries, and implemented strategies to minimise this problem.^{6,7} These studies used similar follow-up protocols; if a diary was not received within five days, a supplemental telephone questionnaire was administered to cover the period of the missing diary. The study protocol therefore ensured that the longest recall period was one week. Subjects were asked to record the actual date of diary completion, with 99%⁶ and 86%⁷ of diaries being reported as completed on or within one day of the assigned diary day. The other studies used protocols where all diaries were returned at the end of the diary keeping period which was two months for the study by Woods et al⁹ and one month for the study by van den Brink et al.⁸ Neither assessed delayed recording of diary entries.

Diary versus interviewer-administered questionnaire

All but one of the five studies reviewed, reported a higher response rate for the interviewer-administered questionnaire than for the diary (Table 2). However, two of these studies^{12,13} asked only questionnaire respondents to keep diaries and so the response rates are not strictly comparable between the two methods, as there is no information about the diary response rates of questionnaire non-respondents. A third study¹⁴ reported a small difference between the two methods (78% with complete questionnaire data and 70% with complete diary data), but only a sub-sample was asked to keep diaries. The fourth study reporting different response rates between the two methods¹¹ required diary keeping when symptoms were present over a two year period. Subjects were phoned every two weeks and a questionnaire administered when symptoms were reported. This study reported that, in the first year of the study, 66% had at least six months of data for both methods; 70% had at least six months of diary data and 92% had at least six months of questionnaire data. There was insufficient data for comparisons in the second year and the study concluded that diary keeping beyond one year was not feasible. The study reporting no difference in response rates between the two methods¹⁰, was a small study and the diary keeping burden was low (daily for three days).

Only one paper discussed selection bias. Bruijnzeels et al¹² reported that parents who completed the interview but refused to keep a diary had less education and were more likely to belong to an ethnic minority than diary keepers, suggesting that selection bias was a problem for diaries. However, there is no way of knowing if the same problem applied to the interview, as the study did not collect similar data on parents who refused to be interviewed and these parents were not invited to keep diaries. None of the papers discussed the issue of delayed recording of diary entries, and none attempted to minimise the potential recall period by requiring diaries to be returned frequently. Diaries were collected at the end of the diary keeping period for all studies except one¹¹, which had a two year diary keeping period and collected diaries monthly.

Accuracy

Diary versus self-completed questionnaire

The purpose of all four studies was to examine recall error in self-reported symptoms, by comparing a retrospective questionnaire to a diary. Because of the nature of the information (symptoms experienced), none of the studies was able to use a third data source (not self-reported) for verification. Two of the studies explicitly used the diary as the gold standard or reference against which to validate the questionnaire.^{6,7} The remaining studies appeared to do this implicitly because, while the stated purpose was to compare the estimates derived from the two methods, both studies set the issue and discussed the results in terms of recall error of the retrospective questionnaire.^{8,9}

Woods et al⁹ reported poor agreement between the Menstrual Distress Questionnaire and a diary of menstrual symptoms for the same menstrual cycle (κ less than 0.3 for all items). The prevalence of all symptoms and days away from work or school was higher by questionnaire report for the menstrual and premenstrual phases and was higher by questionnaire report for most symptoms during the remainder of the cycle. However, in addition to issues of recall and compliance, the discrepancy may also be related to instrument design. The questionnaire listed menstrual symptoms and asked respondents to rate their presence/severity, while the diary asked respondents to report health problems and if they were menstruating without reference to a symptom list.

The remaining three studies⁶⁻⁸ found that items related to the experience of the headache, such as intensity and duration, were scored higher by questionnaire report than by diary. However, headache frequency and work days missed due to headache (reported by two studies^{6,7}), were similar between the two methods. For the study by van den Brink et al⁸, the differences might be explained by the different time period covered, as the questionnaire was completed prior to commencement of diary keeping. Two studies reported moderate to high correlation between questionnaire and diary reports but, the use of Pearson⁶ or Spearman⁷ correlation coefficients does not take account of chance correlation and is therefore not a good measure of agreement.

None of the four studies attempted to examine fatigue effects over the period of diary keeping.



Diary versus interviewer-administered questionnaire

The purpose was to compare the two methods of data collection for two studies^{11,12} and to assess the validity of the data collection instruments for the remaining three studies.^{10,13,14} Two studies also compared diary and questionnaire reported health service utilisation with physician reports.^{13,14}

Four of the five studies reported higher rates of symptoms or health service utilisation by questionnaire than by diary. Nandha et al¹⁰ found no difference between diaries and questionnaires for reports of medication effectiveness, but this study had a short diary keeping period (3 days) and a small sample size (n=48). Thus, this study would have been less susceptible to recall error and less likely to detect a difference than the other studies.

Gold et al¹¹ found the incidence of all respiratory symptoms was higher by questionnaire than by diary but the magnitude of the difference varied by the type of symptom. They also found that, while total rates of respiratory illness were similar between questionnaire and diary reports, parents with three or more children reported more illness by questionnaire than by diary. Bruijnzeels et al¹² found that overall symptoms were reported more frequently by questionnaire than by diary, but that this varied according to the nature of the symptom and some types of symptoms were reported more frequently by diary. This study used different question formats between questionnaire and diary method. The questionnaire used a list of symptoms from which to choose, whereas the diary used open ended questions, which may have influenced the higher rate of symptoms reported in the questionnaire. This study also found the discrepancy between questionnaire and diary reporting was greater for less educated respondents.

The two studies which compared questionnaire and diary reported health service use with physician reports found more service use reported by questionnaire than by diary. Bruijnzeels et al¹³ found that more GP consultations were reported by parents (for their children) than by general practitioners, regardless of the method of parent report, and that the discrepancy was higher for the questionnaire method than for the diary method. Diaries showed slightly better agreement with GP reports than did questionnaires (kappa 0.64 and 0.58 respectively), but this difference was not statistically significant. Assuming that GP reports reflect the true number of consultations, questionnaires showed better sensitivity than diaries (0.82 and 0.70 respectively) and diaries showed better specificity than questionnaires (0.96 and 0.90 respectively). The difference was statistically significant at the 5% level for the specificity but not for the sensitivity. Guzman et al¹⁴ compared diary and questionnaire measures of health service utilisation and questionnaire and physician measures (not diary and physician). This study measured a large number of items and reported a large range of agreement depending on the item. In general, it found more services and medications reported by questionnaires than by diaries. It also found moderate to good agreement (kappa 0.38-0.78) between questionnaires and diaries and fair to good agreement (kappa 0.29-0.85) between questionnaires and physician reports for most items. However, the sample size was small for these analyses (n=32 had sufficient diary and questionnaire data; n=48 had sufficient physician and questionnaire data) and confidence intervals for kappa were not reported.

None of the five studies attempted to examine fatigue effects over the period of diary keeping.

DISCUSSION

The purpose of the review was to identify the most appropriate method of collecting patient self-reported data (or parent reported data) about health service utilisation and out-of-pocket costs for the Cost of Asthma Study. Financial and logistical factors restricted the study to using a self-completed method for the collection of this data. Consequently, the review initially set out to examine studies which compared patient diary methods with self-completed questionnaires for measuring health service use and patient costs. Because of the limited number of comparative studies, the criteria were broadened to include other data types (such as symptoms) and interviewer-administered questionnaires. The search found no studies comparing diary and questionnaire methods to measure patient costs and only two studies comparing diary and interviewer-administered questionnaire methods to measure some aspects of health service utilisation.

Most of the studies reported higher response rates for questionnaires than for diaries, including the two studies using postal questionnaires (although the difference was small for one of these studies). These two studies used rigorous follow-up protocols for the diaries, including telephone calls and reimbursement. The study with the lowest diary keeping burden (fewest days of diary keeping) was the only study with the same response rates for diary and questionnaire. This would suggest that the differential response was related to the burden of diary keeping rather than study protocols. This contradicts an earlier review by Verbrugge³ which found very high response rates for diaries provided there were good follow-up and collection protocols in place, and that response rates were not related to the duration of diary keeping required. Verbrugge's review did not assess selection bias and none of the studies in the current review compared diaries and questionnaires for selection bias. While there was some evidence of selection bias associated with diaries, where respondents were more highly educated than non-respondents, none of the studies in this review examined selection bias for questionnaires separately.

This review found a tendency to report more symptoms and service use by questionnaires than by diaries. This contradicts Verbrugge's³ findings that diaries produce higher levels of reporting. In the current review, some of this difference in reporting may be explained by the different formats used for the questionnaires and diaries. Several studies used a symptom list in the questionnaire but not the diary, which would be likely to increase the number of symptoms reported in the questionnaire (particularly the less important symptoms). However this does not explain all of the differences, including those of health service use where higher rates of use were reported by questionnaires relative to diaries and relative to physician reports.

One explanation of the higher levels of reporting in questionnaires relative to diaries is memory recall error due to telescoping (remembering events from an earlier period as occurring in the study period). Further, there is likely to be a degree of underreporting in diaries due to memory recall error when diary entries are completed after the due date or simply not completed at all where entries are only required when events occur. Two studies used rigorous follow-up procedures to minimise delayed recording in diaries and found the majority of daily diary entries reported as entered within one day of the relevant diary day. However, reported date and time of entry may not be an accurate measure of diary compliance. Stone et al⁵ found that while 90% of diary entries were reported as complying with the diary protocol, only 11% actually complied. Underreporting may also explain the greater discrepancies between diaries and questionnaires for some sub-groups such as those with low education levels (who may not understand what is required and complete fewer diaries or omit some events) or those with large families (who may complete diaries after the due date because of time pressures). Another potential source of underreporting in diaries is fatigue effects where the number of events recorded decreases over the diary keeping period as the respondent's interest wanes. This is more likely to be an issue for studies with a long duration of diary keeping. None of the studies in the current review examined this issue.



The discrepancies identified between diaries and questionnaires give little indication as to which method is the more accurate. Studies which use a third data source (not self-reported) are likely to be most informative here. However, these studies are limited by the lack of alternative data sources (which is the reason for collecting self-reported data). Two studies in the current review included data from physician reports. One of these studies found that, while both diary and questionnaire reports of GP attendances were higher than the general practitioners' reports, sensitivity was better for interviewer-administered questionnaires and specificity was better for diaries. This suggests that diaries and questionnaires may be more suited to particular situations depending on the implications of different error types for the study. For example, an interviewer-administered questionnaire might be better if it is more important to detect all events that actually occurred rather than to ensure that the events that are reported truly occurred (when a diary would be preferable).

The studies in the current review provide some information about two approaches to collecting self-reported data, diaries and self-completed questionnaires, but do not provide sufficient evidence to favour either approach in the current context. The published studies reviewed here suffered from some or all of the following limitations:

1. Study design problems including different data collection periods, use of both instruments for the same period on the same sample (which may mean that questionnaire reporting is enhanced by diary keeping as a memory recall aid), the lack of a third data source to verify accuracy rather than just identifying differences, insufficient demographic data to examine bias (particularly education and socio-economic status), and no examination of fatigue effects on diary results
2. The follow-up protocol lacked procedures to detect or minimise delayed diary completion
3. Sample selection problems including small sample size and inadequate or no examination of selection bias (ie different response rates between sub-groups)
4. The data collection instruments were not comparable, particularly in relation to the use of check lists in questionnaires but not in diaries.

A large randomised study is likely to be the most appropriate way of further informing the problem. Such a study should incorporate the capacity to test specific issues such as the effects of the duration of data collection and the effects of different follow-up procedures on response rates and data quality. However, given the heterogeneity of the two approaches, such a study is unlikely to provide general recommendations about the preferred approach. This will depend on the nature of the variables of interest and the relevance of the issues discussed above for the specific study.

In view of the duration of the data collection required and the costs of good diary follow-up procedures, the Cost of Asthma Study will use repeated questionnaires to collect self-reported data. Different recall periods will be used for different types of events, and administrative data collections will be accessed where possible.

Table 1: Summary of studies comparing diary and self-completed questionnaire data collection methods

DIARY PERIOD	QUESTIONNAIRE PERIOD	RESPONSE RATE	MEASURES	ANALYSIS	FINDINGS
Study: Woods et al 1982 ⁹					
Study method: Comparison of prospective diary and retrospective questionnaire by same sample for same time period					
Sample: 100 women aged 18-35 years & not pregnant, selected randomly from census					
Daily for 2 months, reminder phone calls, 50% offered a \$20 (US) financial incentive	End of 2 mth diary period, self-completed at interview	73 completed diaries and questionnaire (used in analysis) 6 refused, 19 did not complete diary, 10 did not complete questionnaire, 2 returned diary too late	Presence and severity of symptoms associated with menstrual cycle, 8 summated scales constructed from 47 symptoms rated on scale 1-6	Percent difference between quest and diary for symptom prevalence at each phase of cycle, kappa to estimate concordance between quest and diary for presence/absence of symptoms at each phase of cycle	>20% discrepancy in symptom prevalence between quest and diary on many symptoms, most overestimated by quest and during menstrual phase, kappa poor agreement between quest and diary at all phases
Study: Stewart et al 1999 ⁶					
Study method: Comparison of prospective diary and retrospective questionnaire by same sample for same time period					
Sample: 226 migraine sufferers aged 18-65 years. Random population sample of 438 invited, 239 agreed of whom 226 met migraine criteria					
Daily for 3 months, returned weekly, reminder phone calls, financial incentive \$5 (US) per completed diary	Beginning and end of 3 mth diary period. Beginning self-completed at clinic, end not stated	132 (58%) had data for both methods and used in analysis. 5 withdrew from study, 49 had incomplete diary data and 40 did not complete questionnaire	Headache frequency, and experience. Frequency items = days with headache, missed work etc. Experience items = mean pain intensity, reduced effectiveness at work etc. Total score also calculated	Pearson's Correlation Coefficient to measure correlation between quest and diary for each variable and total score, systematic bias assessed by comparing quest and diary mean and median of items	Moderate correlation for total score, item correlations ranging from low to high, experience items systematically higher on quest than diary, no systematic bias on frequency items
Study: Stewart et al 2000 ⁷					
Study method: Comparison of prospective diary and retrospective questionnaire by same sample for same time period					
Sample: 188 migraine sufferers aged 18-55 years. Random population sample of 426 eligible, 188 agreed to participate					
Daily for 3 months, returned weekly, reminder phone calls, financial incentive \$5 (US) per completed diary	End of 3 mth diary period, postal with reminder phone call	144 (77%) used in analysis. 10 withdrew from study, 31 had incomplete diary data, 1 had incomplete questionnaire data and a further 2 outliers excluded	Headache frequency, pain level and lost productivity due to headache (number of days lost and number days reduced productivity). Total score also calculated	Spearman's correlation coefficient to measure correlation between quest and diary for each variable and total score, systematic bias assessed by comparing quest and diary mean and median of items, paired t-test to compare means	Moderate to high correlation between quest and diary for items and total scores. Quest significantly underestimated number of days with headache, overestimated pain intensity and missed days from work and household work. Reduced productivity days were similar between quest and diary.
Study: van den Brink et al 2001 ⁸					
Study method: Comparison of prospective diary and retrospective questionnaire by same sample for different time period					
Sample: 212 school children aged 9-16 years with headache complaints					
4 entries daily for 4 weeks	3 months prior to diary, completed in classroom (no specific period referred to)	181 (85%) used in analysis. 26 did not complete all diaries, 5 had incomplete questionnaire data	Headache frequency, duration intensity and severity. Severity = 6pt likert, intensity = 100mm VAS for Quest & Diary (median diary score used). Frequency = 6pt likert (eg 1/year) converted to monthly for quest and median frequency in month for diary. Duration = 7pt likert (eg 1/2hour) for quest and median duration in month for diary	Recall error assessed with Wilcoxon test and Spearman rank correlation (between quest and diary for each variable). Recall bias assessed with differential and proportional frequency and intensity scores and dichotomous duration agreement variable as dependent variables in regression analysis to explain size of bias	No difference between quest and diary for frequency. Intensity and duration significantly greater on quest. Explanatory variables did not predict duration error. Older children underestimated frequency and overestimated intensity in quest. Higher severity associated with over-estimation of frequency and less over-estimation of intensity. Depressive children underestimated frequency in quest



Table 2: Summary of studies comparing diary and interviewer-administered questionnaire data collection methods

DIARY PERIOD	QUESTIONNAIRE PERIOD	RESPONSE RATE	MEASURES	ANALYSIS	FINDINGS
<p>Study: Gold et al 1989¹¹ Study method: Comparison of prospective diary and retrospective questionnaire on same sample for same time period Sample: 422 parents of children aged 5-11 years selected at random from schools in a defined geographical area</p>					
Entry on days child had symptoms over a 2 year period	2 weekly phone calls over 2 year period. Phone call followed by home interview when child had symptoms in previous 2 weeks	66% (277/422) had 6 months of data by 2 methods. 70% had at least 6 months diary data and 92% had at least 6 months questionnaire data. Larger families provided less diary data per person	Respiratory illness symptoms from symptom list	Rates per 100 person months calculated for each symptom class and illness and for each method. Paired t-test to compare rates obtained by each method	Incidence of all resp symptoms was significantly higher by quest than by diary. The magnitude of the difference was greater for lower resp symptoms than upper. Total respiratory illness rates similar between quest and diary but parents with 3 or more children reported more illness on quest
<p>Study: Bruijnzeels et al 1998¹² Study method: Comparison of prospective diary and retrospective questionnaire on same sample for different (consecutive) 2 week period Sample: 2227 parents of children aged 0-14 years selected at random from patients listed with participating GPs</p>					
Daily for 3 weeks, first 2 weeks used in analysis	Completed before diary commenced, referred to previous 2 weeks	1630 used in analysis (restricted study pop to respondent = mother for both methods), diary response rate was 81% (1805/2227), only questionnaire respondents (87%) asked to keep diary	Any symptoms — from checklist for quest and open ended for diary	Pearson Chi-square to test if diary and quest produced same estimates for the presence of symptoms. Specific symptoms tested by binomial test of proportions	Occurrence of any symptom reported more frequently in quest than diary. More ear problems, colds/flu, weakness reported in quest and more diarrhoea reported in diary. Discrepancy between quest and diary significantly greater for less educated mothers
<p>Study: Bruijnzeels et al 1998¹³ Study method: Comparison of prospective diary and retrospective questionnaire on same sample for different (consecutive) 3 week periods and comparing both to GP reports Sample: 2282 parents of children aged 0-14 years selected at random from patients listed with participating GPs</p>					
Daily for 3 weeks	Completed before diary commenced, referred to the last 2 months but only used last 3 weeks in analysis	1765 used in analysis (only used if diary/quest overlapped with GP reporting period), diary response rate was 79% (1805/2282), only quest respondents (89%) asked to keep diary	Any health problems, if and when child saw GP	Sensitivity and specificity of diary and quest (GP as gold standard), kappa to assess agreement between diary/GP and quest/GP, prevalence of problems and RR of report by parent relative to GP	Higher parental report than GP report, discrepancy greater for quest than diary. Quest better sensitivity and diary better specificity, Agreement better for diary (CI overlap). Prevalence of most problems significantly higher than GP in quest but not in diary. Low education/SE status more under-reporting of problems
<p>Study: Guzman et al 1999¹⁴ Study method: Comparison of retrospective interview questionnaire with prospective diary and provider records for the same 6 week period Sample: Convenience sample of 102 workers aged >18 years, with occupational lower back pain</p>					
Daily for 6 weeks	Completed at end of 6 wk diary keeping period	32 used in diary v quest analysis, 48 used in physician record v quest. Quest response 78% (80/102), diary response 70% (32/46), physician response 69%	Health care utilisation, medication, management strategies for lower back pain and impact on work	Kappa used to assess agreement between different methods for reported utilisation	Moderate to good agreement between quest and diary for most items, more health services and medications reported in quest than diary. Fair to good agreement between quest and physician reports, prescription medication only fair, more referrals reported by physicians
<p>Study: Nandha et al 2001¹⁰ Study method: Comparison of retrospective interview and prospective diary on same sample for same 3 day period Sample: 48 patients aged >18 years who purchased a study medication for cough symptoms at a participating pharmacy</p>					
Daily for 3 days	Beginning and end of diary keeping period	44 used in analysis, 92% response for diary and quest	Cough symptoms and medication efficacy	Spearman's rank correlation coefficient and Wilcoxon signed rank test	No difference between diary and quest for relief from medication. First day symptoms rated better on quest than diary, no difference for days 2 and 3

REFERENCES

1. de Vaus DA. *Surveys in social research*. 3rd ed. Sydney: Allen & Unwin, 1991.
2. Foddy W. *Constructing questions for interviews and questionnaires. Theory and practice in social research*. Melbourne: Cambridge University Press, 1993.
3. Verbrugge LM. Health diaries. *Medical Care* 1980;18(1):73-95.
4. Carp FM, Carp A. The validity, reliability and generalizability of diary data. *Experimental Aging Research* 1981;7(3):281-96.
5. Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient non-compliance with paper diaries. *British Medical Journal* 2002;324:1193-1194.
6. Stewart WF, Lipton RB, Simon D, Liberman J, Von Korff M. Validity of an illness severity measure for headache in a population sample of migraine sufferers. *Pain* 1999;79(2-3):291-301.
7. Stewart WF, Lipton RB, Kolodner KB, Sawyer J, Lee C, Liberman JN. Validity of the Migraine Disability Assessment (MIDAS) score in comparison to a diary-based measure in a population sample of migraine sufferers. *Pain* 2000;88(1):41-52.
8. van den Brink M, Bandell-Hoekstra ENG, Abu-Saad HH. The occurrence of recall bias in pediatric headache: A comparison of questionnaire and diary data. *Headache* 2001;41(1):11-20.
9. Woods NF, Most A, Dery GK. Estimating perimenstrual distress: a comparison of two methods. *Research in Nursing & Health* 1982;5(2):81-91.
10. Nandha D, Goodyer L, Woodruffe-Peacock C. Diary cards and the assessment of cough symptoms in community pharmacies. *Pharmacy World & Science* 2000;22(1):17-20.
11. Gold DR, Weiss ST, Tager IB, Segal MR, Speizer FE. Comparison of questionnaire and diary methods in acute childhood respiratory illness surveillance. *American Review of Respiratory Disease* 1989;139(3):847-9.
12. Bruijnzeels MA, Foets M, van der Wouden JC, Prins A, van den Heuvel WJ. Measuring morbidity of children in the community: a comparison of interview and diary data. *International Journal of Epidemiology* 1998;27(1):96-100.
13. Bruijnzeels MA, van der Wouden JC, Foets M, Prins A, van den Heuvel WJ. Validity and accuracy of interview and diary data on children's medical utilisation in The Netherlands. *Journal of Epidemiology & Community Health* 1998;52(1):65-9.
14. Guzman J, Peloso P, Bombardier C. Capturing health care utilization after occupational low-back pain: development of an interviewer-administered questionnaire. *Journal of Clinical Epidemiology* 1999;52(5):419-27.

