Validity and accuracy of interview and diary data on children's medical utilisation in The Netherlands

MA Bruijnzeels, JC van der Wouden, M Foets, A Prins and WJ van den Heuvel

J. Epidemiol. Community Health 1998;52;65-69

Validity and accuracy of interview and diary data on children's medical utilisation in the Netherlands

M A Bruijnzeels, J C van der Wouden, M Foets, A Prins, W J A van den Heuvel

Abstract

***Study objective*—To assess the validity and accuracy of children's medical utilisation estimates from a health interview and diary and the possible consequences for morbidity estimates. The influence of recall bias and respondent characteristics on the reporting levels was also investigated.**

***Design*—Validity study, with the medical record of the general practioner (GP) as gold standard. In a health interview and three week diary estimates of medical utilisation of children were asked and compared with a GP's medical record.**

***Setting*—General community and primary care centre in the Netherlands.**

***Participants*—Parents of 1805 children and 161 GPs**

***Main results*—The sensitivity of the interview (0.84) is higher than the diary (0.72), while specificity and κ are higher in the diary (0.96;0.64) than in the interview (0.91;0.5–8). Recall bias, expressed as telescoping and heaping, is present in the interview data. Prevalence estimates of all morbidity are much higher in the interview, except for skin problems. Compared with a parental diary more consultations are reported exclusively by the GP for children from ethnic minorities (OR 1.6), jobless (OR 2.3), and less educated mothers (OR 2.6).**

***Conclusions*—Estimates of medical utilisation rates of children are critically influenced by the method of data collection used. Interviews are prone to introduce recall bias, while diaries should only be used in populations with an adequate level of literacy. It is recommended that medical records are used, as they produce most consistent estimates.**

(*J Epidemiol Community Health* 1998;**52**:65–69)

Department of General Practice, Erasmus University Rotterdam, the Netherlands
M A Bruijnzeels
J C van der Wouden
A Prins

Netherlands Institute of Primary Health Care, Utrecht, the Netherlands
M Foets

Department of Health Sciences, University of Groningen, the Netherlands
W J A van den Heuvel

Correspondence to:
Dr MA Bruijnzeels, Room Ff 318, Department of General Practice, Erasmus University, PO Box 1738, 3000 DR Rotterdam, the Netherlands.

Accepted for publication 12 March 1997

Estimates of medical utilisation (frequency and morbidity) of children are derived from parental interviews or medical records and used to monitor the occurrence of diseases and to assess the need for health services.[1 2] Despite the use of different methods, the validity and accuracy of children's medical utilisation estimates have not been assessed. Since Tennant showed that proxy and saliency effects are larger if a household member reports for other household members (which is often the case for children), special attention on the validity and accuracy of parental responses for their children is justified.[3] Another way to determine children's medical utilisation is using a health diary.[4] The validity and accuracy of a medical utilisation estimate from a diary has also never been assesed.

Comparisons between health interviews and medical records, concerning medical conditions, dietary habits, obstetric histories, hospitalisation, medication use, and chronic diseases, showed that the general population tends to underreport medical events, but that they overreport events as well.[5–7] In an Australian study, Britt *et al* compared the nature of morbidity presented to the general practitioner (GP) as reported in medical records and patient interviews and concluded that no large differences were found between both methods.[8] However, they did not ask whether the consultation took place according to the parent. Hence, we do not know whether medical utilisation estimates (and consequently morbidity estimates) derived from different sources are comparable.[9]

An important factor that influences the validity and accuracy of self reported data in interviews is recall bias expressed by omission of events and telescoping, which means that events are recalled as having occurred either more recently or longer ago than they actually did.[10] As omission leads to underreporting and telescoping may lead to under and overreporting of events, the effect of recall bias on utilisation estimates is ambiguous. For health diaries, omission of events or fatigueness (reduced willingness to complete diaries in the same detail as time passes) are threats to validity.[11] Also, respondent characteristics as age, sex, and socioeconomic status might influence validity and accuracy of self reported data. However, consistent relations for interview data have not been found.[10 12] For diary data, effect of respondent characteristics has to date not been investigated.

Therefore, in this study we assessed the validity and accuracy of estimates of children's medical utilisation from a health interview and diary by comparing them with medical records. We established the consequences for morbidity estimates. Also, we investigated whether and how respondent characteristics influence these medical utilisation estimates.

Methods

In the Netherlands each inhabitant is listed with their own GP and a consultation with a GP is the point of entry into the Dutch health

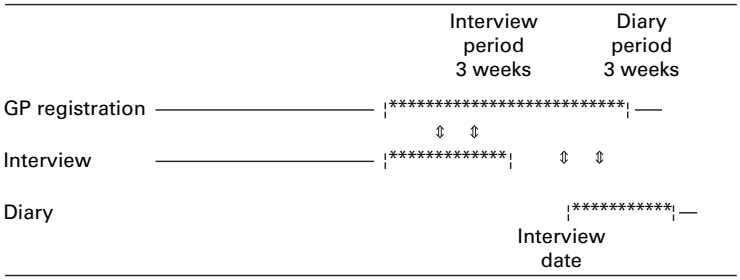

Figure 1 Overlap of the registration periods of the various methods used in the Dutch National Survey carried out in 1987 and 1988.

care system. We used data from the Dutch National Survey of Morbidity and Interventions, carried out in 1987 and 1988. A non-proportional stratified sample of 161 GPs was drawn form the Dutch GP population (n=5826).[13] The practices of the participating GPs contained 63 753 children between 0 and 14 years of age. All participating GPs registered all contacts with their patients during three months (GP registration). A random sample of 100 persons per participating GP was drawn. In this sample 2561 children were represented. Parents of these children were asked to answer a health interview about their child: parents of 2282 children participated (response 89%). Subsequently, these parents were asked to keep a diary during three weeks about their child's health. Parents of 1805 children cooperated (response 79%). One general practice had to be excluded because of non-overlapping registration periods, so 1765 children and 160 GPs participated in the study.

In the GP registration all consultations were registered by either GP or receptionist/nurse on a specially designed form. Among the registered items were date of the consultation and the reason for consultation as expressed by the parent, which was afterwards coded according to the International Classification for Primary Care (ICPC).[14] The health interview took place after two months of GP registration. In the health interview questions regarding a GP consultation were whether, how long ago, and for what reason the child had contacted the GP during the last two months. The diary started the day after the interview took place. It consisted of a one page questionnaire for every day. Precoded questions of interest were among others whether the child had suffered from a health problem that day and if so, what kind of health problem and whether the child had gone to the GP for that problem that day.

Interview and diary referred to different time spans (two months and three weeks); these data were not directly comparable for our purpose. Because the diary was for three weeks, we restricted the comparison between interview and GP registration to three weeks as well (most recent). A consultation was included if it took place within three weeks according to one of the methods. Figure 1 shows the overlapping time periods of the three data collection methods. We considered only consultations during surgery hour and home visits as reported by parent or GP. Thus, telephone consultations were excluded.

KEY POINTS

- Medical utilisation rates estimated by different methods are not comparable.
- Overreporting of salient events in a health interview is partly because of telescoping effects of recall.
- Health diaries should be used with caution in lower educated populations and ethnic minorities.

We matched the consultations from the different methods by the reported consultation date. Next, we classified the matches into a perfect match (same date in both instruments), an almost perfect match (maximum three days of difference between both methods), a problematic match (more than three days and less than 15 days difference), and a mismatch (consultation is reported by one instrument only or the difference is larger than 14 days).

For the validity, we calculated sensitivity and specificity with the GP registration as the gold standard, because the organisation of the GP registration yields probably the most accurate data.[15] We calculated Cohen's κ statistic (as an indicator for the accuracy) for all matches.[16] To assess morbidity differences between the methods, we compared prevalences of specific diagnostic groups classified by the ICPC according to all methods and calculated relative risks with 95% confidence intervals indicating the risk for each diagnostic group of being reported more often by the parent than by the GP.

We checked whether telescoping in the interview or fatigueness, or both, in the diary occurred by comparing the number of consultations that parents reported during the registration period. The influence of respondent characteristics is assessed separately for under and overreporting parents and is expressed by odds ratios resulting from bivariate logistic regression analyses with the matched group as reference category. The respondent characteristics studied were child, maternal, and family characteristics. For the child age, sex, birth order (firstborn compared with laterborn), and ethnicity were considered. Ethnicity was divided into children belonging to an ethnic minority or not.[17] For the mother age (over and under 35 years of age), educational level (elementary compared with continued/-university) and working status (has a paid job or not) were considered. For the family the socioeconomic status level according to the profession of the wage earner (high/middle compared with low),[18] and the composition (one or two parent family) were considered. Characteristics with $p < 0.25$ in the bivariate analysis are added to a multivariate logistic regression model.[19]

Results

Parents reported 355 consultations of 281 children (mean (SD) 1.3(0.6)) in the interview and 213 consultations of 160 children (mean (SD) 1.3(0.6)) in the diary. GPs registered 216 consultations of 183 children (mean (SD) 1.2(0.5)) during the interview period and 191 consultations of 165 children (mean (SD)

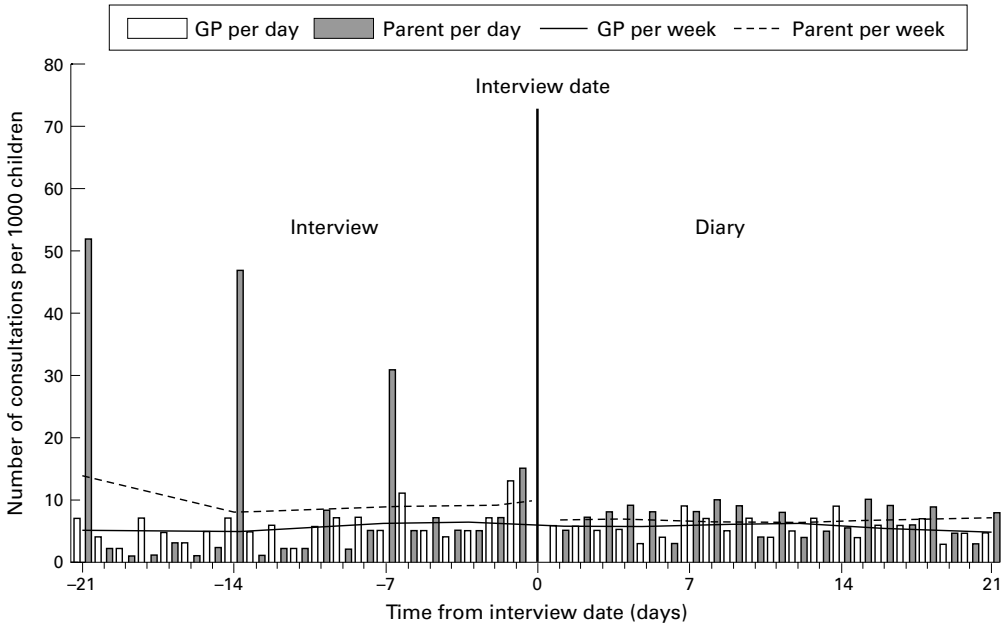

Figure 2 Number of consultations per day (bars) and mean number of consultations per week (lines) per 1000 children as reported by the parent in interview and diary (each during three weeks) and by GP (during six weeks).

1.2(0.4)) during the diary period. The number of matched consultations was 177 in the interview period and 133 in the diary period. So, parents reported more consultations exclusively than the GPs, more pronounced in the interview (178 versus 39) than in the diary (80 versus 58). If we breakdown the matches by perfectness, in the diary period 81% matched perfectly, 17% almost perfectly, and only 2% matched problematically whereas these percentages are for the interview period 36%, 46%, and 18% respectively. Figure 2 shows a possible explanation for these differences in perfectness. The reported consultations in the interview clustered around seven, 14, and 21 days ago, whereas in the GP registration and diary the number of consultations were constant over all days. Also in the figure, an indication of telescoping is present, as the number of consultations in the interview increased between 15 and 21 days ago.

Table 1 shows the distribution of the children with and without consultations according to the reporter. As more consultation reports per child occurred, the numbers do not correspond to the distribution of consultations. The validity of both parental methods is satisfactory, but slightly better for the interview than for the diary. The sensitivity is higher for the interview method (0.82) than for the diary method (0.70), whereas the specificity is higher for the diary. The κ statistic is higher for the diary method (0.64) than for the interview method (0.58), although the confidence intervals for both κ statistics overlap. The number of false positives (GP no, mother yes) is much higher in the interview than in the diary, in contrast with the false negatives (GP yes, mother no). Table 2 shows the consequences of these larger number of consultations in the interview for the morbidity estimates. All categories had higher prevalences according to the interview than according to both other methods. Three diagnostic categories show a deviant pattern. Firstly, musculoskeletal problems had an even higher prevalence according to the interview, than was expected. Secondly, skin problems had according to both parental methods lower prevalences than according to the GPs. Thirdly, "other problems" had higher prevalences in both parental methods.

Table 3 shows what respondent characteristics caused over or underreporting. In the case of the interview, parents overreported if they had a high socioeconomic status. The educational level of the mother had no effect. The multivariate analysis of the interview data yielded similar odds ratios. In the multivariate

Table 1 Validity and accuracy of interview and diary regarding a GP consultation in 1765 children

	Interview versus GP registration			*Diary versus GP registration*		
	GP Consult	*No consult*		*GP Consult*	*No consult*	
Mother						
Consultation	140	137	277	108	60	168
No consultation	30	1458	1488	46	1551	1597
Total	170	1595	1765	154	1611	1765
			*95% CI**			*95% CI**
Sensitivity		0.82	0.76, 0.88		0.70	0.62, 0.77
Specificity		0.91	0.90, 0.92		0.96	0.95, 0.97
κ Statistic		0.58	0.52, 0.64		0.64	0.57, 0.71

*95% Confidence intervals.

Table 2 Prevalence of reasons for encounter per 1000 children by diagnostic category and relative risks of a health problem being reported more often by parent than by GP for both comparisons (1765 children)

	Interview - GP registration				*Diary - GP registration*			
	Prevalence Interview (n=355)	*GP (n=216)*	*RR**	*CI**	*Prevalence Diary (n=213)*	*GP (n=191)*	*RR**	*CI**
General and unspecified problems	32	19	1.7	1.1, 2.6	13	11	1.2	0.6, 2.1
Digestive problems	22	12	1.8	1.1, 3.0	12	11	1.2	0.6, 2.1
Eye problems	9	5	1.7	0.7, 3.8	2	3	.7	0.2, 2.4
Ear problems	15	10	1.4	0.8, 2.6	14	12	1.1	0.6, 2.0
Musculoskeletal problems	20	8	2.6	1.4, 4.8	12	11	1.2	0.6, 2.1
Respiratory problems	51	29	1.8	1.3, 2.5	40	29	1.4	0.9, 1.9
Skin problems	26	29	.9	0.6, 1.3	11	24	.5	0.3, 0.8
Other problems	26	10	2.7	1.6, 4.7	16	6	2.6	1.3, 5.3

*RR=relative risk; CI=95% confidence intervals.

Table 3 Bivariate odds ratios (ORs) with confidence intervals (95% CIs) of parental under and overreporting for respondent characteristics

	Health interview GP registration				*Diary GP registration*			
	Overreporting (n=277)		*Underreporting (n=170)*		*Overreporting (n=168)*		*Underreporting (n=154)*	
	OR	*95% CI*	*OR*	*95% CI*	*OR*	*95% CI*	*OR*	*95% CI*
Child characteristics								
Age (> 4 years)	0.9	0.6, 1.5	1.3	0.6, 2.8	0.8	0.4, 1.4	0.7	0.31, 1.4
Sex (boys)	0.9	0.5, 1.4	0.8	0.4, 1.8	1.8*	0.9, 3.4	1.2	0.6, 2.5
Birth order (firstborn)	0.7*	0.5, 1.2	0.6*	0.3, 1.4	1.8*	0.9, 3.4	1.3	0.7, 2.7
Ethnicity (minority)	1.2	0.4, 3.1	†		0.8	0.2, 3.1	1.8*	0.5, 5.9
Maternal characteristics								
Age (< 35 years)	1.0	0.6, 1.6	1.2	0.5, 2.9	1.2	0.6, 2.2	1.2	0.6, 2.4
Educational level (low)	1.1	0.5, 2.4	1.4	0.4, 4.6	0.5*	0.1, 2.5	3.9*	1.4, 11.0
Work status (does not work)	1.1	0.6, 2.1	0.9	0.3, 2.6	0.5	0.2, 1.4	2.8*	1.3, 6.2
Family characteristics								
Socioeconomic status (low)	0.6*	0.3, 1.2	1.4	0.5, 3.4	0.4*	0.1, 1.2	2.3*	1.0, 5.3
Composition (one parent family)	0.9	0.3, 2.5	1.8	0.5, 7.4	2.8	0.5, 17.2	2.4	0.3, 17.7

*Wald p < 0.25; CI=95% confidence intervals.

analysis of the diary data, no characteristic was statistically significant ($p < 0.05$) related to overreporting, although in the bivariate analysis parents reported more consultations for their girls and firstborns. Resulting from the bivariate analysis, parents underreported if their child belonged to an ethnic minority, if the educational level of the mother was low, if they were jobless, and if they had a low socioeconomic status. Odds ratios for these characteristics resulting from the multivariate analysis (not shown in table 3) remained high, though not statistically significant: ethnic minority 1.6 (95% CI 0.4, 5.9), low maternal educational level 2.6 (95% CI 0.7, 9.4), jobless mothers 2.3 (95% CI 1.0, 5.3), and for low socioeconomic status 1.4 (95% CI 0.5, 3.9).

Discussion

This study is the first to check on the accuracy and validity of medical utilisation data for children. The utilisation rates determined by different methods vary, especially between GP and interview. The results indicate that consultation rates of the past three weeks reported by the parent are overestimated by about 60% (355/216). Consultations reported in the diary yielded a better result. Obviously, utilisation rates of various methods are not comparable.

Considering the accuracy of the measurements, the κ statistic indicated a substantial agreement between the diary and medical record and only a moderate agreement between the interview and the medical record, which is strengthened by the much larger number of perfect matches for the diary-GP registration comparison than for the interview-GP registration comparison. This difference in accuracy between both parental methods can be explained by heaping or clustering of the reported consultations in the interview to exactly one, two, and three weeks before. As the κ statistic is invariant to asymmetry between the disagreements, the sensitivity and specificity are more informative for validity aspects.[20] The lower sensitivity of the diary contradicts general statements that in a diary more valid data are collected than in an interview.[4 11] A typical error of the diary data was simply missing the whole illness episode or forgetting to tick the GP item. This deviant finding compared with other studies is probably caused by the fact that false negative consultations (recorded by the GP, but not by the parent) in diary and interview could not be detected in previous studies. The organisation of the Dutch health care system enabled us to determine these false negative consultations. The higher sensitivity but lower specificity for the interview than for the diary indicate that there are relevant differences between both methods, most pronounced by the large overreporting in the interview. This overreporting can partially be explained by telescoping. Another explanation is that parents tend to forget less salient reasons like skin problems and overreport more salient reasons like musculoskeletal and "other" problems compared with the GP. Probably, a combination of telescoping, the proxy effect, and the saliency principle strengthens the reporting of consultations for more salient illnesses, even more for

problems that also affect the parent's activities.[3] If the period of reference is longer the saliency principle will probably influence this recall bias more heavily. Hence, consequences of these incomparable results for prevalence estimates are large.

Another important finding is that some respondent characteristics were related to reporting behaviour. The underreporting of the parent in the diary was higher for parents with children belonging to an ethnic minority, which is probably because of language problems. That mothers with a low education and who are jobless report less consultations may be caused by literacy limitations. Other studies subscribe these statements for self administered questionnaires, while these limitations can be avoided with face to face interviews.[21] Our results cast doubt on the usefulness of a self administered diary for low educated respondents and ethnic minority groups. Also we found that the overreporting of the parent was positively associated with socioeconomic status. Perhaps, these parents report a lot of consultations to demonstrate that they are good parents.[10]

In evaluating these results some remarks should be made. Firstly, we excluded telephone and preventive consultations. However, as these consultations mostly involve less salient problems, inclusion of these consultations would probably have worsened the validity of the self reported data. Secondly, given the organisation of the GP registration (during or directly after the consultation with a weekly check by a research associate), we assume that the GP registration is the most accurate. However, the GP may have under or overreported consultations as well. Overreporting by the GP is unlikely, because then he would have to make up consultations. Underreporting (forgetting to fill out the registration form) may have taken place. In case of substantial underreporting by the GP, the number of false negatives would increase and the specificity of both diary and interview would be higher. Nevertheless, the large discrepancy in false negatives between interview and diary indicates that parental overreporting in the interview remains substantial. Finally, not all parents were willing to cooperate in the diary and interview study. Non-response is often because of lack of motivation and time constraints. In case of participation of the non-responders, these constraints would probably have caused an even worse outcome.

This study shows that medical utilisation rates and derived prevalence estimates of health problems for children are critically influenced by the method of data collection used. GP registrations, if well organised, can provide reproducable estimates of utilisation rates, although some systematic underreporting may be present. Interviews produce too high rates and should be handled with caution. Utilisation rates of especially more salient problems are prone to be overestimated. Diaries should only be used in populations with an adequate level of literacy. Hence, given their consistency we recommend to use medical records if possible.

Funding: the study received funding from the Dutch Ministry of Education and Science.
Conflicts of interest: none.

1 Starfield B, Katz H, Gabriel A. Morbidity in childhood. A longitudinal view. *N Engl J Med* 1984;**310**:824–9.
2 Bruijnzeels MA, Wouden JC van der, Foets M. General practice consultation in childhood in The Netherlands: sociodemographic variation. *J Epidemiol Community Health* 1995;**49**:532–3.
3 Tennant A, Badley EM, Sullivan M. Investigating the proxy effect and the saliency principle in household based postal questionnaires. *J Epidemiol Community Health* 1991;**45**: 312–16.
4 Roghmann KJ, Haggerty RJ. The diary as a research instrument in the study of health and illness behavior. *Med Care* 1972;**10**:143–63.
5 Linet MS, Harlow SD, McLaughlin JK, *et al*. A comparison of interview data and medical records for previous medical conditions and surgery. *J Clin Epidemiol* 1989;**42**:1207–13.
6 Harlow SD, Linet MS. Agreement between questionnaire data and medical records. *Am J Epidemiol* 1989;**129**:233–48.
7 Heliövaara M, Aromaa A, Klaukka T, *et al*. Reliability and validity of interview data on chronic diseases. *J Clin Epidemiol* 1993;**46**:181–91.
8 Britt H, Harris M, Driver B, Bridges-Webb C, O'Toole B, Neary S. Reasons for encounter and diagniosed health problems: convergence between doctors and patients. *Fam Pract* 1992;**9**:191–4.
9 O'Toole B, Driver B, Britt H, *et al*. Using general practitioners to measure community morbidity. *Int J Epidemiol* 1991; **29**:1125–32.
10 Coughlin SS. Recall bias in epidemiologic studies. *J Clin Epidemiol* 1990;**43**:87–91.
11 Verbrugge L. Health diaries. *Med Care* 1980;**18**:73–95.
12 Mechanic D, Newton M. Some problems in the analysis of morbidity data. *J Chron Dis* 1965;**18**:569–80.
13 Foets M, Velden J van der. *Dutch national survey of morbidity and interventions in general practice. A summary of the survey design*. Utrecht: NIVEL, 1992.
14 Lamberts H, Wood M. *International classification of primary care*. Oxford: Oxford University Press, 1987.
15 Rose G, Barker DJP. Reliability and validity. *BMJ* 1978;2:1070–1.
16 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;**33**:159–74.
17 ISEO. *Beter meten 1*. Rotterdam: ISEO, 1987.
18 Erikson R, Goldthorpe JH, Portocarero L. Intergenerational class mobility and the convergence thesis: England, France and Sweden. *Br J Sociol* 1983;**34**:303–43.
19 Hosmer DW, Lemeshow S. *Applied logistic regression*. New York: John Wiley, 1989.
20 Maclure M, Willet WC. Misinterpretation and misuse of the kappa statistic. *Am J Epidemiol* 1987;**126**:161–9.
21 Sullivan LM, Dukes KA, Harris L, *et al*. A comparison of various methods of collecting self-reported health outcomes data among low-income and minority patients. *Med Care* 1995;**33**:AS183–94.