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


Updated information and services can be found at:
http://jech.bmj.com/cgi/content/abstract/52/1/65

These include:

References
3 online articles that cite this article can be accessed at:
http://jech.bmj.com/cgi/content/abstract/52/1/65#otherarticles

Rapid responses
You can respond to this article at:
http://jech.bmj.com/cgi/eletter-submit/52/1/65

Email alerting service
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article

Notes

To order reprints of this article go to:
http://www.bmjjournals.com/cgi/reprintform

To subscribe to Journal of Epidemiology and Community Health go to:
http://www.bmjjournals.com/subscriptions/
Validity and accuracy of interview and diary data on children’s medical utilisation in the Netherlands

M A Bruijnizeels, 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 practitioner (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.

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 assessed.

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 fatiguess (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
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. We calculated Cohen’s $k$ statistic (as an indicator for the accuracy) for all matches. 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 fatigue, 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 later-born), and ethnicity were considered. Ethnicity was divided into children belonging to an ethnic minority or not. 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), 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.

**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)
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 $\kappa$ statistic is higher for the interview method (0.64) than for the interview method (0.58), although the confidence intervals for both $\kappa$ 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

<table>
<thead>
<tr>
<th></th>
<th>Interview versus GP registration</th>
<th>Diary versus GP registration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GP Consult</td>
<td>No consult</td>
</tr>
<tr>
<td>Mother</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consultation</td>
<td>140</td>
<td>137</td>
</tr>
<tr>
<td>No consultation</td>
<td>30</td>
<td>1458</td>
</tr>
<tr>
<td>Total</td>
<td>170</td>
<td>1595</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>GP Consult</th>
<th>No consult</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>0.82</td>
<td>0.76, 0.88</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.91</td>
<td>0.90, 0.92</td>
</tr>
<tr>
<td>$\kappa$ Statistic</td>
<td>0.58</td>
<td>0.52, 0.64</td>
</tr>
</tbody>
</table>

*95% Confidence intervals.
of the Dutch health care system enabled us to determine these false negative consultations. The higher sensitivity but lower specificity for the interview-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 statistic is invariant to asymmetry between the disagreements, the sensitivity and specificity are more informative for validity aspects. The lower sensitivity of the diary contradicts general statements that in a diary more valid data are collected than in an interview. 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

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

<table>
<thead>
<tr>
<th>Health interview-GP registration</th>
<th>Diary-GP registration</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Underreporting</strong></td>
<td><strong>Overreporting</strong></td>
</tr>
<tr>
<td>(n=170)</td>
<td>(n=277)</td>
</tr>
<tr>
<td>OR 95% CI</td>
<td>OR 95% CI</td>
</tr>
<tr>
<td>Sex (boys)</td>
<td>1.1</td>
</tr>
<tr>
<td>Birth order (firstborn)</td>
<td>0.7</td>
</tr>
<tr>
<td>Ethnicity (minority)</td>
<td>1.3</td>
</tr>
<tr>
<td>Age (&lt;35 years)</td>
<td>0.9</td>
</tr>
<tr>
<td>Educational level (low)</td>
<td>1.1</td>
</tr>
<tr>
<td>Work status (does not work)</td>
<td>1.4</td>
</tr>
<tr>
<td>Family characteristics</td>
<td>1.4</td>
</tr>
<tr>
<td>Socioeconomic status (low)</td>
<td>1.0</td>
</tr>
<tr>
<td>Composition (one parent family)</td>
<td>0.9</td>
</tr>
</tbody>
</table>

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

**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 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 statistic is invariant to asymmetry between the disagreements, the sensitivity and specificity are more informative for validity aspects. The lower sensitivity of the diary contradicts general statements that in a diary more valid data are collected than in an interview. 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. 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 to these statements for self-administered questionnaires, while these limitations can be avoided with face to face interviews. 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.

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 reproducible 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.