Systematic review of the occurrence of infantile colic in the community

Lucassen, P. L. B. J.; Assendelft, W. J. J.; van Eijk, J. T. M.; Gubbels, J. W.; Douwes, A. C.; van Geldrop, W. J.

Published in:
Archives of disease in childhood

DOI:
10.1136/adc.84.5.398

Citation for published version (APA):
Systematic review of the occurrence of infantile colic in the community

P L B J Lucassen, W J J Assendelft, J Th M van Eijk, J W Gubbels, A C Douwes and W J van Geldrop

Arch. Dis. Child. 2001;84;398-403
doi:10.1136/adc.84.5.398
Abstract

Aims—To assess the occurrence of infantile colic in the community and the need for professional help; and to study the influences of potential determinants of infantile colic.

Methods—Surveys were identified by a systematic search in Medline (1966–98) and Embase (1988–98). Retrieved publications were checked for references. Studies selected were community based, prospective, and retrospective surveys on the occurrence of infantile colic published in English, German, French, or Dutch. Occurrence rates were calculated as percentages. Methodological quality of the surveys was assessed by two assessors independently with a standardised criteria list containing items on method of data gathering, definition of colic, and drop out rate.

Results—Fifteen community based surveys were identified. The methodological quality varied considerably and was generally low. Even the two most methodologically sound prospective studies yielded widely varying cumulative incidence rates of 5–19%. Referral rates or the need to seek help because of crying were consistently lower than occurrence rates for prolonged crying as such. Gender, socioeconomic class, type of feeding, family history of atopy, and parental smoking were not shown to be associated with colic.

Conclusion—Occurrence rates of infantile colic vary greatly according to methodological quality. A considerable number of parents reporting prolonged crying do not seek or need professional help.

Keywords: infantile colic; community; incidence

The reported occurrence rates of infantile colic vary from 10% to 40%. Many reviews only use data from selected populations. Differences may reflect differences in definitions, methods of data gathering, and study design.

In epidemiological and more clinically oriented studies, infantile colic has been defined in different ways but prolonged crying is an almost constant feature. However, studies define normal and prolonged crying differently. The most commonly accepted is the “rule of 3”: crying during at least three hours per day on at least three days of at least three weeks. The first two features originated from a description by Wessel and colleagues. The third is often not established because of problems in documenting the condition. Other time criteria that have been used are: severe crying for several hours per day, crying for more than two hours per day, and overall duration of more than three hours per day. Studies not using an indication of crying time define infantile colic as unexplained crying, crying seen as a problem, or crying with which parents felt they could no longer cope and for which they sought help. Other sources of variability are the presence of symptoms possibly of gastrointestinal origin (a high pitched pain cry, flatulence, and difficulties with the passage of stools) and the emphasis on consolability. Apart from these differences, studies generally agree that infants with colic are healthy, thriving, and below 6 months of age. The foregoing could be congruent with two underlying concepts. Firstly, infantile colic is conceived as one distinct entity with prolonged crying as the main symptom and several optional additional features. Secondly, infantile colic is conceived as a collection of different entities, each defined separately.

Different methods have been used to establish the diagnosis, including audiotape recordings, parental diaries, the Crying Patterns Questionnaire (CPQ), non-specified questionnaires, and interviews, personal or by telephone. Crying has been classified in various ways, such as amount of crying in hours per day or answers to a broad question on whether the baby cried a lot. Moreover, some studies assessed crying prospectively with an inception early in life, whereas others used a retrospective data gathering method, sometimes up to 14–28 months.

To evaluate the impact of infantile colic on health care and to plan research on infantile colic, exact data on its occurrence are needed. We aimed to perform a systematic review of community based surveys to assess the occurrence of infantile colic and to assess the need for professional help related to study quality (method of data gathering, definition of infantile colic, and drop out rate). We also aimed to assess occurrence rates according to source of recruitment. Finally, we planned to assess the influence of prognostic factors such as gender and socioeconomic class, and aetologic factors such as type of feeding, parental smoking, and family history of atopy on estimates of occurrence of infantile colic.

Methods

SELECTION OF STUDIES
In March 1998 we performed a Medline (1966–98) and an Embase (1988–98) search with the following search strategy: “colic” and
“crying” as keywords, and “colic”, “cry”, and “fuss” as free text words were combined with “epidemiology” (explode) as key word and “incidence”, “prevalence”, and “morbidity” as free text words. To contain sensitivity of the search, we did not search with the “restrict to focus” option.\textsuperscript{21} The searches were limited to infants younger than 1 year and restricted to English, German, French, and Dutch languages. We checked the references of retrieved publications for missing studies. The first author screened all citations in articles about incidence or prevalence of infantile colic in community based samples. Publications on consultation rates for crying and studies on non-white infants were excluded because our aim was to generate data on infants in Western societies.

We defined community based studies as those including infants recruited from well baby clinics or community populations. Studies recruiting infants born in hospital, but otherwise without problems, were also accepted as representative of the general population. We labelled a study prospective when data collection was performed during the crying period and retrospective when data were collected after resolution of the crying problem.

### METHODOLOGICAL QUALITY

We evaluated methodological quality with a self developed scale, based on Laupacis and colleagues\textsuperscript{22} and Fletcher and colleagues.\textsuperscript{23} We used different versions for prospective and retrospective surveys. Each study was scored on the items “definition of infantile colic”, “method of obtaining data on crying behaviour”, and “drop out”. The adequacy of each item could be scored “yes” or “no”. Each item had a weight of 1 or 2. In prospective studies the quality score ranged from 0 to 5, with a low score indicating high susceptibility to bias. The score of retrospective studies had a range from 0 to 5. Two of us (WJvG, PLBJL) scored all surveys independently. We were not blinded for information on authors and journal, because one of us was well acquainted with the material already. Disagreement was solved by consensus. The degree of agreement before the consensus meeting was expressed as a percentage agreement and as kappa.\textsuperscript{24, 25}

### OCCURRENCE RATES

Occurrence rates of infantile colic in prospective studies are usually given as cumulative incidence rates and occurrence rates from retrospective studies as period prevalence rates.\textsuperscript{26} However, both rates are identical when studying an ailment that only occurs in the first four months of life. Whenever possible, we extracted data from the original publications to calculate occurrence rates over the period in which infantile colic usually presents—that is, the first three months of infancy. For all data,
95% confidence intervals (CI) were calculated. Pooling of the data was a priori considered not suitable because of heterogeneity in design, method of obtaining data, and definition of infantile colic.

**Results**

The Medline search revealed 87 citations (33 on the subject of infantile colic); the Embase search yielded 52 citations (26 on the subject of infantile colic/prolonged crying). Moreover, we located one not yet included in Medline and one submitted study. Reference checking yielded no additional publications. One study was excluded because it was performed in non-white infants. We excluded two other studies: one which gave data on mothers seeking help from their general practitioner, and one which had been performed in infants referred to a private paediatric clinic. Therefore, we report on 15 studies, 13 of which were identified on Medline, and five on Embase. Most surveys had been performed in Scandinavian countries and the UK. Seven studies had a prospective design, and one study had a retrospective part. Tables 1 and 2 present a tabular summary of the surveys.

The quality score ranged from 1 to 4 for the prospective studies and from 0 to 5 for retrospective studies. Tables 3 and 4 gave details of the quality assessment. There was good interrater agreement on the summary quality score of prospective studies (agreement 89%, κ 0.78). Agreement was moderate for retrospective studies (agreement 77%, κ 0.53). For prospective studies interrater agreement (expressed as κ) for drop out rates was 0.75, for adequacy of the diagnosis 0.33, and for assessment of the outcome 0.90. For retrospective studies these figures were 0.25, 0.53, and 0.58 respectively. Most disagreement was caused by differences in interpretation or unclear reporting in the article and could be resolved easily. We reached consensus in all cases.

Occurrence rates in prospective studies varied from 3% to 28% and in retrospective studies from 8% to 40% (tables 1 and 2); refusal state (percentage of invited/recruited persons not participating) varied from 6% to 49%; and loss to follow up in prospective studies (percentage of recruited persons not completing the study) varied from 1% to 27%. Five studies did not use a time criterion in defining colic. Six studies assessed occurrence of colic by measuring mothers’ subjective interpretation of their infants’ crying as “prolonged” or “colicky”, without using a time criterion. “Gas problems” and “problems with comforting the infant” were each included once in the definition. All except one survey (Canivet and colleagues, retrospective part), assessing infant crying behaviour and care seeking behaviour separately, found consistently lower occurrence rates for colic seen as a problem needing professional advice than for colic seen as prolonged crying for a certain period.

One of the two best prospective studies—the one using Wessel’s “rule of 3” stringently—
reported a relatively low incidence of infantile colic (5%). The other high quality prospective study reported a much higher figure (19%).

None of the surveys separately reported the presence of gastrointestinal symptoms. Three surveys measured consolability.

Surveys recruiting cases from well baby clinics reported lower occurrence rates compared to recruitment from birth registers and hospitals. One survey did not state the source of recruitment.

Eight studies reported on occurrence in boys and girls separately: seven found no difference in occurrence of colic, and one reported a significantly higher proportion of boys crying more than three hours per day.

Five studies reported on influence of socioeconomic class: three found no differences, while two reported slightly higher rates in higher socioeconomic classes.

Seven surveys compared breast fed and formula fed infants: four found no difference, in two studies the occurrence rates among breast fed infants were slightly higher, and in one it was slightly lower. Only two studies reported separately on a positive family history of atopy: both found no association between the presence of a positive family history of atopy and the presence of infantile colic. In two studies influence of parental smoking was detected.

### Discussion

Occurrence rates of infantile colic in community based samples vary greatly because of differences in study design, site of recruitment, definition, and method of data collection. The two best prospective studies yielded occurrence rates of 5% and 19% respectively. Our review stresses the importance of a uniform definition and good documentation methods.

Gender, socioeconomic status, type of feeding, parental smoking, and family history of atopy have not been consistently measured and in studies that presented data, no consistent influence on occurrence estimates were detected. For some surveys small study size with related low power, might have caused the inability to detect differences in subgroups.

We are interested to find that in a prospective study of 160 Korean infants (which we excluded) no case of infantile colic was found. This survey was adequate according to the quality criteria in this study as the researchers used a 24 hour diary and a definition of infantile colic that included a time criterion. Confirmation of this finding in other non-Western societies is needed as differences in diet or care taking activities may be responsible for differences in occurrence of infantile colic, possibly providing clues for prevention or treatment.

In our systematic review, we used an unvalidated quality assessment method. We were not aware of any existing methods for assessing quality of occurrence studies. Although the low agreement for some items may be a result of inadequate reporting in the original publication, inconsistencies in our method could also have played a role. Further development of quality assessment methods for occurrence studies is therefore needed.

The various surveys in the review actually assessed three different concepts of infantile colic: crying of a certain duration; crying as a problem for the mother; and crying leading to a need to seek professional help. Comparing the need for professional help with crying of certain duration (for example, more than three hours per day) is therefore needed.

### Table 3: Quality score of prospective studies, individual items and total score

<table>
<thead>
<tr>
<th>Item (weight) *</th>
<th>van der Wal**</th>
<th>Canivet**</th>
<th>Lothe**</th>
<th>St James-Roberts**</th>
<th>Altavara**</th>
<th>Crosse**</th>
<th>Thomas**</th>
<th>Ståhlberg**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selection: % not included &lt;20% (1)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>?</td>
<td>+</td>
<td>?</td>
<td>−</td>
<td>?</td>
</tr>
<tr>
<td>Follow up: loss to follow up &lt;20% (1)†</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>?</td>
<td>+</td>
<td>−</td>
<td>?</td>
<td>−</td>
</tr>
<tr>
<td>Diagnosis—definition sufficient (1)</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Outcome (method)</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Diary at least 7 days (2)</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Diary less than 7 days (1)</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Other method (0)</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Outcome (timing)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>Multiple measurement periods (2)</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Single measurement period (0)</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Total quality score (0–6)</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

*Weight: score to compliance with item.
+ complied to item; − not complied to item; ? not stated.
† Loss to follow up includes non-participation and not completing the study after recruitment.

### Table 4: Quality score of retrospective studies, individual items, and total score

<table>
<thead>
<tr>
<th>Items (weight) *</th>
<th>van der Wal**</th>
<th>Canivet**</th>
<th>Lothe**</th>
<th>St James-Roberts**</th>
<th>Altavara**</th>
<th>Crosse**</th>
<th>Thomas**</th>
<th>Ståhlberg**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drop-out</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

*Weight: score to compliance with item.
+ complied to item; − not complied to item; ? not stated.
hours per day), in those studies which assessed both,
both most mothers seem to cope without professional help. It is not clear whether this discrepancy is caused by characteristics of the
mother, characteristics of the infant (crying pattern and related features), or both. It is pos-
sible that consolability, the presence of gastro-
testinal symptoms, and acoustic cry features (a more “painful” sound) are determinants of the decision to seek professional help. Moreo-
ver, distinguishing between crying of normal
and excessive duration remains a problem. An
older study had shown that healthy infants in
Western societies cry for about 150 minutes
per day at 2 weeks of age, for almost three
hours per day (median) at 6 weeks and for
about 60 minutes per day at 12 weeks. Recent
research shows similar figures: St James
Roberts and Hall87 measured a mean cry dura-
tion in normal infants aged 1–3 months of
about two hours per day; and Lehtonen and
Korvenranta88 assessed maximum crying levels
of about three hours per day in 4 week old
healthy infants. So, although arbitrary, the
three hour criterion seems to make sense as a
distinction between normal and excessive.

In our opinion, surveys with a prospective
design yield more reliable estimates of occur-
cence rates than retrospective studies, as the
latter are prone to recall bias. The influence of
this bias is greater in conditions measured sub-
jectively, such as infantile colic. In spite of this,
even the occurrence rates reported from
prospective studies vary widely. Differences in
methods of data collection may contribute to
this. We assume the validated 24 hour diary89
to be the best diagnostic method in occurrence
studies, but it seems impossible for parents to
use this method daily for 12–16 weeks. There-
fore, as a compromise, Canivet et al used this
method on one predetermined day each week.
To further improve the diagnostic qual-
y of estimating crying duration, this method
could be supplied with the Crying Patterns
Questionnaire,90 a validated91 retrospective
method, which asks for an estimate of crying
duration during the previous week.

As the studies in this review used a wide
range of definitions and measurement meth-
ods, we can conclude little about infantile colic
as a collection of different entities.92 Future
research should therefore aim to discern the
importance of distinguishing the following
entities: firstly, mothers complaining about
their infants’ crying, even though it is within
normal limits; secondly, mothers complaining
about an excessively crying infant who is
consolable and has no additional gastro-
testinal features; thirdly, mothers complaining
about an infant with prolonged crying, who is
unconsolable and has gastrointestinal fea-
tures; and fourthly, mothers not complaining
about, but reporting on request an infant who
cries excessively. Important items to be as-
essed in a valid way are: the time spent crying,
consolability, and gastrointestinal symptoms,
including a high-pitched pain cry; and mater-
nal characteristics, for example, general wellbe-
ing and the presence of anxiety or depression.
Data should preferably be gathered prospec-
tively with diaries, combined with instruments
aimed at measuring symptoms, such as the
“Colic Symptom Checklist”.93

We thank Anja van Gulick for her help in searching
Medline/Embase and obtaining the articles, and Rosemarie
Tomes for checking the English language.


43 Wal MF van der, Boom DC van den, Pauw-Plomp H, Jonge GA de. Mothers’ reports of infant crying and soothing in the multi-cultural city of Amsterdam, the Netherlands. Submitted.


