Prevalence of family violence in adults and children: estimates using the capture–recapture method

Arie Oosterlee¹, Remy M. Vink², Filip Smit³

Background: Reliable prevalence estimates of family violence in adults and children are difficult to obtain. Most are based on surveys or registration counts, whose research designs and methods are often questionable. The results of questionnaire-based surveys can be unreliable, due to the risk of selection bias caused by people’s unwillingness to disclose such sensitive information, or due to recall bias, especially in the case of the abuse of young children. Selection bias may also work in the opposite direction, when a survey investigates one specific subject. This is because people who are interested in a certain subject tend to be more willing to participate than average.

Disadvantages are also inherent to registration counts, which may represent only the tip of the proverbial iceberg, even when registrations from several different health services and other organizations are combined. Combining the cases registered by different organizations can be complicated, as the organizations often use different definitions and indicators for the concept of family violence.

The capture–recapture method (CRC) is a research method for estimating otherwise unknown numbers in a target population. Originally developed in biology (in wildlife and fisheries studies), it can be used to estimate the prevalence of child abuse, this suggestion apparently attracted little attention. To our knowledge, the application of CRC to family violence has been only twice to date: once for babies under 1 year of age, and once for older children.

CRC appears to be a valid and feasible research method for estimating the prevalence of family violence and child abuse. It can be used to complement other methods, especially in young children, in whom valid results are otherwise difficult to obtain.

Methods

Study population

The study was conducted in Haarlem, a city with a population of 149,000, with 29,000 children and adolescents up to the age of 18, and with 120,000 adults aged 18 and older. This is an average-sized town for the Netherlands.
Data were collected by the eight local organizations listed in Table 1.

**Instruments and procedures**

At the onset of the study, most of the participating organizations did not register family violence at all; some registered inadequately or used conflicting definitions. With help of these organizations, we thus developed a uniform one-page registration form. Among other items, the registration form included the following: victim, witness or perpetrator status; age; relation to the perpetrator or victim; and a distinction between current or recent violence (defined as a currently dangerous or potentially dangerous situation for the victim) and past violence (defined as any psychosocial problem of individuals due to exposure to violence in the past). Date of birth, postal code and gender were registered as personal identifiers.

The definition of family violence was adopted from that of the Dutch Ministry of Justice; in essence, it consisted of a list of specific examples of forms of violence.\(^{11}\) To enable the organizations participating in data collection to use it without ambiguity, it was carefully explained.

All cases of family violence that came to attention during the registration period were to be registered, whether or not they were already known before the start of the registration period. This created a period prevalence as a study outcome that required some sort of intervention by an agency (such as the police or health service). To register young witnesses of family violence, each participant organization was instructed to register all relevant members of each family with whom they were in contact.

The duration of the registration period was 7 months, from 1 October 2003 to 1 May 2004. At the end of the registration period, the personal identifiers were recoded by an independent third party into a unique identification code, which was impossible to trace to the person in question. This was done both to secure the collaboration of the data-collecting institutes and to stay well within the Dutch Law on privacy and the use of data. The data of the eight organizations with the unique code numbers and without the personal identifiers were then sent to the Municipal Public Health Department, which had no access to the coding key.

After checking and cleaning, the data of the eight organizations were combined. With the help of a computerized-matching algorithm, it was ascertained how often the same individual had been registered by one organization or by more than one (overlap). This led to the generation of lists containing counts \((1, 2, \ldots, K)\) of the number of times a person had been seen by any of the data-collecting organizations.

**Statistical analyses**

The data available made it possible to analyse four groups: young victims and young witnesses, each up to age 18; adult victims and adult perpetrators. For each of these four categories preferably three-sample CRCs were applied, whenever enough overlap existed between three organizations simultaneously. This was to obtain a summary estimate while at the same time controlling for statistical interaction between organizations. If this was not possible because of lack of overlap, two-sample CRCs were applied using the data of two separate organizations.

Three-sample estimates were calculated with Fienberg’s estimator,\(^{13}\) using log-linear analysis of incomplete data. This was done in two steps with SPSS 12.0. First, the best-fitting, most parsimonious model was determined using hierarchical log-linear analysis. Second, the SPSS generalized log-linear model procedure was used to extract the selected model’s constant, which is the natural logarithm of the unknown number, and its corresponding confidence interval.

For the two-sample method, the Petersen estimator was used to estimate the total number of people\(^{5,7}\); 95% CIs were also calculated.\(^{14}\)

The registration period was concluded at 7 months, which led to the calculation of 7-month period prevalence rates for being a victim of recent family violence. (Past violence, defined as any registered psychosocial problem due to exposure to violence in the past, was in itself not relevant for this article). However, comparison of study outcomes with other studies would be better served by 1-year prevalence rates. For adults, 1-year prevalence rates could be calculated by simple extrapolation of the 7-month rates (by multiplying by 12 divided by 7). The validity of this procedure was based on the calculation with the available data of a series of prevalence estimates of consecutive months. For the young victims, numbers were too small for such calculations and a straightforward extrapolation was not automatically
warranted. For this reason, we present for this age group both the 7-month prevalence rate and the calculated 1-year prevalence rate, assuming that the real results lie somewhere in between (see for details for these procedures in Supplementary data).

CIs of the 1-year prevalence rates were obtained by analogous linear extrapolation of the CIs of the 7-month prevalence rates, obtaining CIs with the same relative widths as with the 7-month prevalence.

Where possible, a one-sample analysis was performed using truncated Poisson mixture models applied to the data of people who appeared to have been registered more than once by a separate organization.15

Statistical assumptions
If a CRC study is to produce valid results, four key assumptions must be met: perfect identification, independence of samples, a closed population (closure) and a homogeneous population (homogeneity).5,7

Cross-validation
We cross-validated study outcomes mainly by comparison with the outcomes of the one-sample truncated Poisson mixture models referred to above, and by comparison with results from Dutch regional general health surveys.

More details of the study methods are given in Supplementary data.

Results
During the registration period, a total number of 1051 complete registration forms were collected, relating to 886 individuals. By making use of the information obtained from the registration forms, and by identifying those individuals who had been registered more than once at any one of the participating organizations, it was possible to divide this set of data in different categories in several steps. The final categories of interest were those with which to calculate the 1-year prevalence rates of recent violence. These were adult victims and adult perpetrators of recent violence, and young victims and young witnesses of recent family violence. Table 1 shows an overview of the number of individuals in these four categories, by the organizations by which they were registered.

For the additional one-sample analyses with the Poisson models, subsets of data were split off of the police, being the only organization in the study, which appeared to have enough registrations for this type of analysis.

More details of the data generation mechanism are presented in Supplementary data.

Adult victims of family violence
Table 1 shows by which organization the adult victims of recent violence were registered, including the 18 adults who were registered by more than one organization.

As the table shows, there were overlaps between the police, the social services and the Family Violence Hotline & Referral Agency. No overlap existed among the three organizations simultaneously, which is necessary when testing statistical interaction (i.e. dependence) with three-sample analysis. Therefore, the data were combined with the category of adult victims of past violence. In this larger dataset, an overlap among the three organizations mentioned indeed was present, and three-sample analysis showed interaction between the social services and the referral agency (see for details Supplementary data). The existence of such an interaction (that is referral) was confirmed by the participating caregivers. The interaction was then incorporated into the three-sample analysis of recent violence only, whose results are presented in table 2.

The unknown number of adult victims of recent family violence was calculated as 1055. Together with the observed number of 318, this gave the total estimated number of 1373 (95% CI: 923–2157) for the 7-month period. Extrapolation to 1 year [expressed as a percentage of the total number of adults on 1 January 2004 (census data)], produced an estimate of the 1-year prevalence rate of 2.0% (95% CI: 1.3–3.1) of adult victims of recent family violence.

Young victims of family violence (child abuse)

With 69, the number of registrations obtained for children as victims of recent family violence (child abuse) was much lower. Because there was only one overlapping registration—between the police and the youth health care organization (table 1)—only a two-sample analysis could be performed. These results are presented in table 3. The estimated 1-year prevalence rate for child victims of recent family violence lay between 1.5% and 2.5%, being the 7-month prevalence rate and the extrapolation to the 1-year prevalence rate, respectively.

Young witnesses of family violence
Table 3 also presents our results for the number of children or teenagers who were registered as witnesses of family violence (spouse-abuse). The estimated 1-year prevalence rate of children who were recent witnesses of family violence was 1.2–2.1%, again depending on the extrapolation from 7 months to 1 year.

Perpetrators
The 1-year prevalence rate for adult perpetrators of recent family violence was estimated to be 1.1% (table 2).

Coverage
The total number of registered victims divided by the estimated totals gave the ‘percentages known’ (coverage). These varied between 16% and 28%; see the bottom rows of tables 2 and 3.

Additional analyses
See Supplementary data for the results of the one-sample truncated Poisson mixture models.

Discussion
Main findings
The estimated annual number of victims of recent family violence in Haarlem corresponds to a 1-year prevalence rate of 2.0% for adults, and a prevalence rate of somewhere in the range of 1.5–2.5% for children up to 18 years of age. Each year, another 1.2–2.1% children in Haarlem are witnesses of spouse-abuse. About 20% of the victims of family violence appeared to be registered (coverage).

Validity of definition and completeness of registration
The outcomes of this CRC study should be considered within the context of its strengths and limitations. The definition of family violence used in the study was specific and practical.
However, its overall validity depends largely on comparisons with the results available from other kinds of study (cross-validation; see below and see Supplementary data).

In this study, we registered not only established cases of family violence, but also cases that aroused serious concerns or suspicions. Based on their professional experience, the participating police officers and caregivers estimated that the vast majority of people who had been registered on the grounds of serious concerns would later turn out to be actual cases. However, this may have caused an upward bias in our results. We recommend that future studies establish the correctness of this estimate by following a sample of cases of serious concern over time.

It is possible that the participating relief workers and caregivers sometimes forgot to fill in a registration form for a case of family violence. Because the CRC method views these cases as unknown numbers, any such mistakes are unlikely to have had a serious effect on our study outcomes; only the percentage of cases known to the participating organizations (coverage) might have been underestimated.

The results concerning children are based largely on registrations at the participating youth healthcare organization responsible for children aged ≥5 years. Unfortunately, because the healthcare organization for children aged <5 years did not participate in the study, numbers for this age group were very low. Some reservation is therefore necessary as to the validity of the results for this age group. All other age groups were represented fairly well in the study, except for ages of ≥65 years, for which the same caveat is necessary.

**Table 2** Prevalences and prevalence rates, adult victims of recent family violence and adult perpetrators, three-sample estimates

<table>
<thead>
<tr>
<th></th>
<th>Prevalences and prevalence rates, adult victims</th>
<th>95% CIs</th>
<th>Prevalences and prevalence rates, adult perpetrators</th>
<th>95% CIs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>7-month period</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowna</td>
<td>309</td>
<td>605–1839</td>
<td>552</td>
<td>198–1533</td>
</tr>
<tr>
<td>Unknowna</td>
<td>1055</td>
<td>923–2157</td>
<td>751</td>
<td>397–1732</td>
</tr>
<tr>
<td>Total calculated number (prevalence)</td>
<td>1373</td>
<td>0.8–1.8%</td>
<td>0.6%</td>
<td>0.3–1.5%</td>
</tr>
<tr>
<td>Prevalence rate</td>
<td>1.2%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extrapolation to 1 year</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total number (prevalence)</td>
<td>2354</td>
<td>1582–3698</td>
<td>1287</td>
<td>681–2969</td>
</tr>
<tr>
<td>Prevalence rate</td>
<td>2.0%</td>
<td>1.3–3.1%</td>
<td>1.1%</td>
<td>0.6–2.5%</td>
</tr>
<tr>
<td>Percentage known (coverage)b</td>
<td>24%</td>
<td>15–36%</td>
<td>28%</td>
<td>12–53%</td>
</tr>
</tbody>
</table>

a: Numbers under ‘known’ exclude the registrations of participating organizations that were not used for the CRC-calculations, while those under ‘unknown’ include these registrations

b: Total number registered excluding overlap (table 1) divided by the total calculated number multiplied by 100%

**Table 3** Prevalences and prevalence rates, childhood victims and childhood witnesses of recent family violence, two-sample estimates

<table>
<thead>
<tr>
<th></th>
<th>Prevalences and prevalence rates, childhood victims</th>
<th>95% CIs</th>
<th>Prevalences and prevalence rates, childhood witnesses</th>
<th>95% CIs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>7-month period</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowna</td>
<td>40</td>
<td>143–617</td>
<td>308</td>
<td>106–510</td>
</tr>
<tr>
<td>Unknowna</td>
<td>380</td>
<td>183–657</td>
<td>360</td>
<td>158–562</td>
</tr>
<tr>
<td>Total calculated number (prevalence)</td>
<td>420</td>
<td>0.6–2.3%</td>
<td>1.2%</td>
<td>0.5–1.9%</td>
</tr>
<tr>
<td>Population number</td>
<td>28.829</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prevalence rate</td>
<td>1.5%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extrapolation to 1 year</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total number (prevalence)</td>
<td>720</td>
<td>314–1126</td>
<td>617</td>
<td>271–963</td>
</tr>
<tr>
<td>Prevalence rate</td>
<td>2.5%</td>
<td>1.1–3.9%</td>
<td>2.1%</td>
<td>0.9–3.3%</td>
</tr>
<tr>
<td>Percentage known (coverage)b</td>
<td>16%</td>
<td>10–37%</td>
<td>16%</td>
<td>10–36%</td>
</tr>
</tbody>
</table>

a: Numbers under ‘known’ exclude the registrations of participating organizations that were not used for the CRC-calculations, while those under ‘unknown’ include these registrations

b: Total number registered excluding overlap (table 1) divided by the total calculated number multiplied by 100%

Statistical assumptions

The assumption of perfect identification is especially dependent on the accuracy of record linkage, that is the determination of the overlap between data sources. In our case, bias seems unlikely, given that the record linkage used simple identifiers; in addition, sensitivity analysis showed that imperfect record linkage with adults would have had only negligible effects (data not shown). For the results concerning children, the effects of any imperfect record linkage would of course have been greater; the validity of these results depends largely on the cross-validation.

Violation of the assumption of independence can severely affect study results. In our study, dependence in adult victims could be controlled for statistically under the log-linear model, which incorporated the dependence through the interaction terms. In children, this was not possible, because only two data sources overlapped, and we therefore had to rely on Peterson’s two-sample CRC estimator. We nonetheless believe that the assumption of independence was not violated.
as knowledgeable workers in the field stated that there was no dependence (e.g., referral) between the police and the youth health care organization.

As to the assumption of homogeneity, in most fields of health care and criminal justice the data are often likely to be heterogeneous.\(^5,7\) To some extent this is probably also the case with family violence. While heterogeneity can partly be controlled for by stratified analyses or by capture–recapture case with family violence. While heterogeneity can partly be controlled for by stratified analyses or by capture–recapture

Cross-validation

As stated above, we cannot be completely certain about the validity of the definition of family violence and of the statistical assumptions. Wherever possible, the validity of each specific result should therefore be cross-validated with other data sources.

The outcomes of the one-sample truncated Poisson mixture models appeared more or less in line with the two- and three-sample results, for adults more than for children. See Supplementary data for more details and for a discussion of the cross-validation with the other available data.

Comparison with other studies

Although Boyle et al.\(^10\) studied the application of the CRC method to adults in the study of family violence, the prevalence rate of such violence could not be established, because it was impossible to establish the size of the total study population.

Sibert and colleagues\(^9\) used CRC to study the incidence of severe physical abuse with babies under 1 year of age in Wales, producing an estimated incidence rate of 0.1%. However, we could not determine whether these investigators had actually measured the incidence or the period prevalence. The validity of the outcome is also questionable, as the authors did not mention the possibility of dependence between those two sources, while using only two sources of data, and while dependency was suggested by the relatively large overlap between the sources.

Several general health surveys have been conducted in the Netherlands that included questions on experiences with family violence over the past year. In five different regional studies, the 1-year prevalence rates ranged from 1% to 1.6%.\(^16–20\) Given the probability that victims of spouse-abuse will be underrepresented in studies of general health surveys, this is consistent with the results of our CRC study.

Some recent general health surveys in the Netherlands included questions on child abuse over the previous year. The most interesting of these was a survey in Haarlem and surrounding communities: the question to 13– to 16-year-old school children whether they had been a victim of physical or emotional abuse over the previous year produced a 1-year prevalence of 2.5%.\(^21\)

All in all, it can be concluded that, in terms of their validity, both research methods (the CRC and the general health survey) seem to support each other rather well.

Conclusions and recommendations

Based on our results, we conclude that, each year, about one person in 50 is a victim of family violence in Haarlem. Approximately, another 2% of young people witness spouse-abuse in their family each year.

The validity of this CRC study is supported by the results of the health surveys in Haarlem and in the Netherlands to which we refer above.

To our knowledge, our study is the first in which prevalence rates for spouse-abuse are presented using CRC, and the second in which the method is used for child abuse. The results should therefore be viewed with some caution, and it is recommended that the study be replicated. Caution is especially necessary regarding the study results on children, as available numbers were small.

In itself, CRC may be a useful and feasible tool for assessing and monitoring the prevalence and health-care coverage of family violence and its changes over time. In addition, the method may be the only one available for obtaining valid information on the annual prevalence of family violence affecting children aged approximately <12 years.

Supplementary data

Supplementary data are available at EURPUB online.

Acknowledgements

The authors wish to thank all those at the participating organizations for their cooperation, and W.A.H.J. van Stiphout, H.C.W. de Vet and M.J.J.C. Poos for their thoughtful comments on study analyses and interpretation.

Funding

Dutch Ministry of Justice, partial.

Conflicts of interest: None declared.

Key points

- Reliable research on the prevalence of family violence and child abuse is difficult because of the risk of bias. The CRC may solve this problem.
- The 1-year prevalence rate was 2% for adult victims of family violence and also 2% for victims of child abuse.
- These estimates were in line with results of regional general health surveys. Therefore, the CRC and the general health survey method may support each other’s validity.
- The CRC can also estimate to what extent family violence is covered by the health-care system. About 20% of cases appeared to be known by the health care institutions and other organizations involved.
- Health-care institutions can use their combined registrations to estimate prevalence and health-care coverage of family violence and child abuse in their region, provided they use a standardized method.

References


