On the influence of data source in aggregated data studies: a comparative study of suicide information based on death certificates and judicial files

H van de Voorde, PH Hoof, UM Mulders

Abstract
Study objective—The aim was to assess the differences in suicide data obtained from different database sources.

Design—Death certificates based on suicidal deaths were compared with the information obtained from the non-natural death investigation files at the Public Prosecutor’s Office.

Setting—The study was confined to the time period of 1981–1984 and to the residents of the district of Leuven (Belgium) who died in the district.

Main results—Differences were found in the number of suicidal deaths reported as well as in the demographic variables, the distribution of suicide methods, and the place of death. Sixty percent of the records in both databases could be matched for all variables studied (age, gender, civil state, place of death, and suicide method), and another 10% if only place of death was allowed to differ. For 4.5% of the cases in the judicial files and for 8% of the records in the official statistics database no corresponding record from the other data source could be found. Based on simple probability statistics \[ P(\text{missing record}) + P(\text{different information/present}) \] the possible minimum global information bias could be estimated to range from 7 to 42%.

Conclusions—Different epidemiological pictures of suicide mortality may result from studying different data sources.

J Epidemiol Community Health 1993; 47: 73–75

In Belgium the diagnosis of death and, subsequently, death certification is restricted to the medical profession. In addition, as the cause of death falls under medical confidentiality, it cannot be disclosed to anyone. For that reason, the part of the death certificate listing the primary and related causes of death is sealed by the certifying physician and may not be opened by the local registrar. It is detached from the administrative parts containing identification variables prior to transmission to the Health Authorities for encoding.

Belgian law, however, prescribes that the Public Prosecutor should be informed of all non-natural deaths occurring in his district. To comply with the law, the certifying physician marks on the administrative part of the certificate if death is due to natural, violent, or suspicious causes. In the latter two cases the local registrar will inform the Public Prosecutor’s Office, whereas in the first case he will most probably authorise the burial or cremation of the body. The judicial investigation following certification of a violent or suspicious death may range from simply obtaining confirmation that death was not caused by a third party, to interrogation of relatives and investigation of the circumstances, and ultimately to the appointment of a forensic expert and a full investigation.

From this procedure, it is clear that the certifying physician plays a key role both in the ultimate usefulness of suicide mortality statistics and in the degree of judicial interference with suicidal deaths, since for different reasons he might choose not to report a death as being suicide to the Health Authorities, to the Judicial Authorities, or to both.\(^1\) The true circumstances and the intention of the deceased to die may be unknown to the certifying doctor, in which case he might choose to mention that death was accidental or of undetermined intention, resulting in the underreporting and misclassification of suicidal deaths in official mortality statistics.\(^2\) He may also experience more overt pressure from the relatives of the deceased, who may try to convince the physician not to declare to the local authorities that death was due to violent causes, resulting in an underreporting in the judicial files. The Nationale Raad of the Orde der Geneesheren, the highest medical disciplinary organ in Belgium, had in 1984 already accepted the principle that the certifying physician should in some cases of apparent suicide be allowed to declare incorrectly to the local registrar that death was due to natural causes, and repeated this advice in its journal in 1989.\(^3\,4\)

So, although data obtained from the files at the Public Prosecutor’s office and from the official mortality statistics concerning suicidal deaths should theoretically deal with the same cases and lead to the same epidemiological picture of suicidal death, there is some reasonable doubt as to whether this is true in practice. The present study was designed to investigate this problem and to see if the degree of difference, if any, between the two databases could be used as an indicator of reporting bias.

Methods
All information currently available for scientific research from the official mortality databases on suicide for the period 1981–1984 (source: National Institute of Statistics) was compared to the non-natural deaths investigation files kept with the Public Prosecutor’s office of Leuven. For practical reasons, only residents of the district of Leuven who died in the district were retained for study and ICD code E959 (death from sequelae
after suicide attempt) was omitted. The judicial files were reviewed by two forensic doctors, who classified each case as being suicidal or not suicidal. Only the former category was retained for study.

In a first step, aggregated data were analysed. In order to investigate the differences found between the two data sources, each record in the official mortality database was then compared to each judicial file, putting a different emphasis on the comparability sought.

**Results**

Table I gives the number of suicide cases recorded for the Leuven district in the official mortality database (NIS) and in the Public Prosecutor’s office files, according to gender and civil state. There seems to be an underreporting of 12 male suicide cases in the judicial files. As in the whole of Belgium, suicide in the Leuven district is more frequent in the male population: age-standardised suicide rates according to the NIS data were 27 per 100 000 men-years and 12 per 100 000 women-years (RR = 2.25, 95% CI = 1.81-2.79, p < 0.001). In both sexes suicide rates rise with age, the mean age not being significantly different between sexes [51.5 (SD 18.0) for men and 55.5 (17.7) for women] and between both data sources.

Divorced people and widows or widowers are represented significantly more often than would be expected from their presence in the population (according to NIS data, 7.8%; and 20.7%, against 1.2% and 6.6% in the population respectively, p < 0.001). The underreporting of male suicide deaths in the judicial files seems to have taken place in the subgroups of unmarried and divorced men. In addition, the civil state distribution for women differs slightly between the two data sources, regardless of the same total number of female suicide cases.

In 243 out of the 311 files (78%) from the Public Prosecutor’s office that were classified as suicides, a general practitioner filled in the death certificate. In 64 cases (21%) it was filled in by a medical specialist. In only 13 cases (4%) was a forensic doctor appointed to investigate the cause of death, in four of which he also completed the death certificate. In four out of the other nine cases, the physician first called in to diagnose death had originally written a statement that death was accidental (one intoxication, one drowning, and two falling from windows).

Table II gives the distribution of suicide methods according to gender and data source. The differences between the two databases are alike for both sexes: more deaths by poisoning and jumping from a height in the judicial files and more hangings and drownings in the official mortality statistics. The elucidation in the judicial files of a great deal of the 30 “other” or “unprecisely described” methods (E958) still present in the official statistics cannot account completely for these differences.

Table III shows the distribution of the place of death according to the suicide method. Significant differences were found between both databases (p < 0.05). In the official mortality statistics substantially fewer people were reported to have died at home, in favour of the categories “died during transportation” and “died in the hospital”. From the investigation of the judicial files it was clear, however, that many of these people obviously or at least very probably must have been dead before any medical intervention reached them.

Table IV shows the results of individual record linking for each suicidal case in both databases, according to the comparability strain used.
Influence of data source in aggregated data studies

Table IV Number of corresponding records of suicidal death in official mortality statistics and judicial suicide files, according to different strains of comparability (Leuven district, 1981-1984)

<table>
<thead>
<tr>
<th>Gender</th>
<th>Age</th>
<th>Civil state</th>
<th>Place of death</th>
<th>Method</th>
<th>Male (n)</th>
<th>Female (n)</th>
<th>Total (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>142</td>
<td>54</td>
<td>196</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>+</td>
<td>19</td>
<td>14</td>
<td>33</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>14</td>
<td>6</td>
<td>20</td>
</tr>
<tr>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>26</td>
<td>14</td>
<td>40</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>5</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>206</td>
<td>91</td>
<td>297</td>
</tr>
</tbody>
</table>

PPO file without corresponding NIS record

<table>
<thead>
<tr>
<th>Gender</th>
<th>Age</th>
<th>Civil state</th>
<th>Place of death</th>
<th>Method</th>
<th>Male (n)</th>
<th>Female (n)</th>
<th>Total (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>8</td>
<td>6</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(3.75)</td>
<td>(2.25)</td>
<td>(5.0)</td>
</tr>
</tbody>
</table>

NIS record without corresponding PPO file

<table>
<thead>
<tr>
<th>Gender</th>
<th>Age</th>
<th>Civil state</th>
<th>Place of death</th>
<th>Method</th>
<th>Male (n)</th>
<th>Female (n)</th>
<th>Total (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>20</td>
<td>6</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(8.0)</td>
<td>(2.0)</td>
<td>(10.0)</td>
</tr>
</tbody>
</table>

The 297 cases for which at least some degree of correspondence in the two databases could be found, 66% matched for all variables studied, and another 11%, for all variables but the place of death. However, ultimately 14 cases (4.5%) in the judicial files and 26 cases (8.1%) in the NIS database could not be linked to a corresponding record in the other data source, even if the obviously very generous comparability strains of "method" only or "method and gender" only were used. The sex ratio (male:female) of the percentages of non-matches for judicial files was 1:1.7 and for NIS database records 1:4:1. These non-matching suicidal deaths are summarised in Table V.

Discussion

The usability of official mortality statistics in suicide research has been the subject of long and intense discussion. Some have argued that under-reporting and misclassification rendered them useless or at least of debatable use.6 Others supplied some evidence that, despite such under-reporting and misclassification, they were valuable in comparative studies and time trend analyses, provided some precautions were taken.7 8

In the present study it has been shown that suicide mortality data from two different official sources, which theoretically should deal with the same cases, may contain a number of cases that seem to be unique to one or other of the sources. In the sources studied here, this incidence reporting bias was of the order of magnitude of 4-8%. This percentage has to be added to the number of cases known to neither of the data sources consulted. In addition, even where the same cases were dealt with, substantial differences could be found in the variables that describe them. Depending on the strength of equality looked for, this contents reporting bias in the present study was of an order of magnitude ranging from 3%, where only the suicide method has to be equal in the two databases, to 34%, where all variables have to be equal. This percentage has also to be added to the number of cases in which the information in both data sources is equally incorrect.

Using these figures, the probability that a certain record is missing from one of the data sources or that one or other of them contains possibly false information can be estimated theoretically using probability statistics. It is given by the sum of the probability that a record is missing and the probability that an existing record contains variables that are different between data sources: P(missing) + P(different present). In the present study this minimal global information bias probability ranged from 7% to 42%.

The authors wish to express their sincere thanks to Dr Jur I Carmen, Public Prosecutor at the district of Leuven, for his collaboration and the authorisation to study the judicial files from his office.