Improving standards for case-control studies

Sjoerd Houwing, SWOV Institute for Road Safety Research, Leidschendam, The Netherlands

Abstract

Between 2007 and 2010, six case-control studies were conducted within the European research-project DRUID to estimate the relative risk of serious injury for psychoactive substance use. Guidelines and study protocols were prepared for the DRUID-case-control studies according to the ICADTS guidelines for epidemiological studies (Walsh et al., 2008)(Walsh et al., 2008) to stimulate that the outcomes of these case-control studies were comparable. Furthermore, equivalent cut-offs were applied to adjust for differences between studies that collected information on recent substance use by means of oral fluid samples and by studies that used blood samples. Despite the high comparability of the study designs the results still showed large variations between the calculated odds ratios.

It is assumed that these differences were likely to be caused by several kinds of random and systematic errors. To investigate this assumption an in-depth study on eleven indicators of potential study errors was conducted on the results of the six DRUID case-control studies on injury risk.

The most commonly detected types of errors were selection bias and lack of sufficient study power due to small sample size. These differences seem to explain the majority of the variance between the calculated odds ratios.

In order to avoid bias and confounding due to errors, future guidelines are recommended to more systematically include an overview of the sources of potential bias and instructions of how to avoid them. Furthermore, a-priori assessments on potential bias could reduce the effect of random and specific errors.

By increasing the comparability of study designs and decreasing the potential errors of case-control studies a good estimate of the risk of driving under the influence of psychoactive substances might be available in future.

Background

Between 2006 and 2010 six population based case-control studies were conducted during the European research-project DRUID (DRiving Under the Influence of Drugs, Alcohol and Medicines) in order to determine the risk of being seriously injured while driving with psychoactive substances (Hels et al., 2011). These case-control studies were performed in Belgium (BE), Denmark (DK), Finland (FI), Italy (IT), Lithuania (LT) and the Netherlands (NL). All six studies were screened for the same 23 substances applying uniform analytical cut-off levels. Cases were seriously injured drivers admitted to hospitals after a traffic crash and controls were randomly selected drivers from the general traffic. In epidemiological research, case-control studies are used to compare determinants (e.g. the presence of drugs) between injured and non-injured drivers. The main outcome measure of case-control studies

---

1 This paper is based on the following article: Houwing, S., Hagenzieker, M., Mathijssen, M.P.M., Legrand, S.-A., Verstraete, A. G., Hels, T., Bernhoft, I.M., Simonsen, K.W., Lillesunde, P., Favretto, D., Ferrara, D., Caplinskiene, M., Movig, K. & Brookhuis, K.A. Random and systematic errors in case-control studies calculating the injury risk of driving under the influence of psychoactive substances. This article is published online in Accident Analysis and Prevention (AAP) 2013; doi: 10.1016/j.aap.2012.12.034.
is the odds ratio, which estimates relative risk, since relative risk calculations cannot always be used (Schmidt and Kohlmann, 2008).

The results from the DRUID case-control studies showed large variations in the relative risks for driving under the influence of psychoactive substances. The differences between the odds ratios of the case-control studies could reflect actual differences in relative risk for driving under the influence in the different countries. However, it is hard to believe that drivers in different countries show such large differences for the relative risk of serious injury while driving under the influence of psychoactive substances. Therefore, the observed differences between the odds ratios could also be at least partially explained by several types of errors.

**Aims**
The main objective of this study is to provide insight in the different types of errors that could explain the variance in the results of the six DRUID case-control studies. The results of this assessment may clarify the large inter-country variation that was observed between the odds ratios.

**Methods**
In epidemiological literature different types of categorisations of bias exist. In this study we used the classifications of Kleinbaum et al. (1982) and Rothman (1986) who distinguish three main types of bias: selection bias, information bias and confounding. This categorization is supported by Wacholder et al. (1992) who wrote two companion papers on issues involved in selecting controls for case-control studies. They state that the selection of controls should be comparable to the selection of cases in three ways: the study base should be the same, confounding factors should be used to eliminate any distortion by other factors, and the measurement errors should be comparable. These principles should reduce the three previously mentioned types of bias in case-control studies: selection bias, information bias and confounding. However, they also state that the effectiveness of these principles is constrained by the availability of resources and time.

Information was gathered for each of the potential indicators from the relevant DRUID reports on prevalence (Houwing et al., 2011; Isalberti et al., 2011) and risk estimates (Hels et al., 2011). These prevalence reports included the national reports for the six participating countries, which provided detailed information concerning the hospital studies and the roadside surveys.
### Table 1. Indicators of potential errors.

<table>
<thead>
<tr>
<th>Type of error</th>
<th>Type of bias</th>
<th>Indicator</th>
<th>Short explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random error</td>
<td>Sample size</td>
<td>Influences accuracy</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low cell counts</td>
<td>Low frequency in a cell leads to less accurate odds ratios</td>
<td></td>
</tr>
<tr>
<td>Systematic error</td>
<td>Selection bias</td>
<td>Geographic area covered by cases and controls</td>
<td>Difference in area covered when sampling cases and controls may result in bias</td>
</tr>
<tr>
<td></td>
<td>Size of non-response</td>
<td>Large non-response may result in larger bias</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Age and gender distribution of response and non-response group</td>
<td>Differences may indicate non-response bias</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Non-random sampling</td>
<td>Over representation of specific groups may lead to bias</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Injury scale</td>
<td>Differences in inclusion criteria may lead to incomparable case populations</td>
<td></td>
</tr>
<tr>
<td>Information bias</td>
<td>Sampling method cases and controls</td>
<td>Different sampling methods in cases and controls may lead to bias</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Analytical method</td>
<td>Differences in sensitivity of the methods of analysis may lead to bias</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Time between accident and sampling</td>
<td>Differences in time between accident and sampling may lead to bias</td>
<td></td>
</tr>
<tr>
<td>Confounding</td>
<td>Confounding factors controlled for</td>
<td>Control for different confounding factors may lead to bias</td>
<td></td>
</tr>
</tbody>
</table>

Finally, the collected information from the DRUID reports was used to interpret the calculated odds ratios. A quantitative assessment of bias was not possible, since this would require a large amount of detailed data on e.g. the use of psychoactive substances in the general population for each study region, which were not available. Therefore, in this study bias is only discussed in qualitative terms based on the information derived from the reports on the prevalence and risk studies by searching for deviations between the six studies on each of the indicators.

### Results

Random errors were indicated by an assessment on sample size and on low cell counts. The sample size is the number \( n \) of individual samples in a study. The precision and thus statistical power of a study increases with the sample size. The sample size for the hospital cases varied between 54 for the Finnish study and 839 for the Danish study. The sample size of the control samples which were collected at the roadside ranged between 1,086 for the Italian study and 4,822 for the Dutch study. Only the Belgian and the Danish study included relatively high number of samples in both the case (BE:348 and DK:839) and the control populations (BE: 2,949 and DK:3,002) (Houwing et al., 2011; Isalberti et al., 2011).
In addition to the total sample size, the distribution of the samples over the four cells in a case-control study provides valuable information on the accuracy of the outcomes as well. In the report on risk estimates (Hels et al., 2011) the issue of low cell counts was explored by an alternative method to calculate overall odds ratios. The results of this analysis on low cell counts showed low cell counts for all substances in the Finnish case-control study. The Lithuanian, Italian and Dutch study each had sufficient cell counts for three of the substance groups only. In Denmark and Belgium the results of the case-control study seemed to have the highest accuracy with sufficient cell counts for five and six substance groups, respectively.

Based on the assessment on random errors it may be concluded that all studies have been subject to random errors. The Belgian and Danish studies were considered to have the highest study power and thus the largest precision. Both studies were also the only studies that were regarded as having a sufficiently large sample size of both cases and controls.

Systematic errors were indicated by an assessment on selection bias, information bias, and confounding. For selection bias an assessment was made on five different indicators: the geographical coverage by cases and controls, size of non-response, gender and age differences between the response and non-response group, non-random sampling and differences in injury scale. The geographical coverage by cases and controls and differences in injury scale did not seem to have resulted in large bias. However, the size of non-response in the roadside surveys showed large variations with a range of between 0 and 52%. In Italy, non-response was non-existent since participation was mandatory. In Lithuania, Belgium and Finland the proportion of non-respondents at the roadside was very high at 25%, 48% and 52%, respectively. Based on this information, we assess that there was likely to be an overestimation of the odds ratios for illicit drugs in these three countries.

The assessment on age and gender distribution indicate an under representation of illicit drug use by car drivers in Belgium and Finland, due to a higher share of young male drivers in the non-response group of the roadside surveys. Differences were detected in Denmark and the Netherlands as well, but since the total non-response rate was relatively low the effect is likely to be small. In Lithuania, it may be assumed that due to the over representation of female drivers in the non-response group the prevalence of illicit drugs in traffic is overestimated whereas the prevalence of medicinal drugs is underestimated. In Italy no non-response was observed in either the hospital study or the roadside survey.

The process of random sampling can be endangered by detection bias. Detection bias is a form of selection bias, which can occur due to oversampling of certain sub-populations. The presence of detection bias is hard to reveal. However, in the DRUID report on the case-control studies (Hels et al., 2011) it was noted, based on personal communication with the Italian researchers, that in the Italian roadside survey drivers who showed clinical signs of alcohol impairment had partially been preselected. Therefore, alcohol use in Italian traffic is expected to be overestimated, causing an underestimation of the odds ratios for alcohol use. This detection bias may also have been present for users of psychoactive substances other than alcohol, in the event that they showed signs of impairment as well. We assess that the Italian odds ratios for alcohol and other psychoactive substances were likely to be underestimated. For all other countries no indication was found for bias due to non-random sampling.
Information bias may be indicated by differences in sampling methods between the cases and the controls, by differences in the analytical methods and by differences in the time between accident and sampling between the national DRUID case-control studies. According to the assessment on these indicators we assume only a small likelihood of the presence of information bias. This likelihood is mainly in present in the Dutch and Danish study in which respectively, another collection method for oral fluid sampling and another injury coding method were used. The likelihood of information bias was mainly limited because of a priori agreements on a uniform study design and because of applying equivalent cut-offs between blood and oral fluid concentrations (Verstraete et al., 2011). Furthermore, four separate rounds of proficiency testing were performed. The results of the proficiency testing show that both qualitative and quantitative performance improved during the testing program (Pil et al., 2010). Therefore, bias due to differences in the analytical methods is not likely.

Deconfounding is, like the presence of an identical study base for both cases and controls, one of the main principles of comparative case and control selections (Wacholder et al., 1992). Confounding factors are variables that co-vary both with substance use and crash risk. Taking into account different types of confounding factors can cause variation between the results of case-control studies. In the DRUID project all the case-control studies’ results were calculated by the Technical University of Denmark (DTU), ensuring a uniform method of statistical analysis. The results of all countries were based on a similar set of variables (age and gender). Furthermore, the data from all six roadside surveys were weighted for the volume of traffic flow in the different time periods. There are probably other confounding factors that were not detected in these studies. Age and gender were included as confounding variables in all calculations and since the data were adjusted for differences in traffic volume between time periods, and therefore we assume that the potential effect of confounding has been reduced. However, since confounding has only been eliminated for those factors that we had data available for, any influence of confounding on the estimated odds ratios cannot be ruled out.

**Discussion and conclusions**

It is clear from the results of this study that the presence of uniform guidelines was not sufficient in excluding differences in the design and protocol of the six national DRUID case-control studies. Deviations from the guidelines such as those mentioned in the present study were caused by practical, financial and legal limitations. Such limitations are difficult to overcome for researchers. Therefore, it would be utopic to expect that future studies will be fully comparable with each other.

The results of this study reflect the importance for future review studies or meta-analyses of epidemiological studies that estimate the risk of driving under the influence of psychoactive substances to include assessments of potential errors. These assessments are essential for better understanding the relationship between observed and actual risk estimations. Furthermore, we advise that future case-control studies in the field of driving under the influence include a pilot study with an assessment on potential bias. This would allow identification of the presence of potential limitations in the study design that could result in bias. The list of potential indicators that was used in this study could be used as a guidance, as long as it is kept in mind that this list is tailored to the DRUID case-control studies. Therefore, including additional variables in this list might be necessary. Finally, an a priori calculation of sample size could provide valuable information in how to maximize the precision of the study given certain limitations regarding resources and time. The most
commonly detected types of errors were selection bias and lack of sufficient study power due to small sample size. These differences seem to explain the larger part of the variance between the calculated odds ratios.

By increasing the comparability of study designs and decreasing the potential errors of case-control studies a good estimate of the risk of driving under the influence of psychoactive substances might be available in future.

References


