Universiteit Utrecht

# *Inquiry into methods to estimate the extent of online offences, especially by young people.*

(Summary, conclusions and recommendations)

Peter G.M. van der Heijden (Universiteit Utrecht)
Maarten J.L.F. Cruyff (Universiteit Utrecht)
Ger H.C. van Gils (BeleidsOnderzoek&Advies (BOA))
Utrecht, september 2017

# Summary, conclusions and recommendations

## *1. Introduction*

This is a report of a study into a number of methods to estimate the extent of online offences, especially by young people. The research question is whether there are any methods that could lead to improvements or additions to methodology of the Juvenile Crime Monitor 2015 (*Monitor Jeugdcriminaliteit*, MJC). The MJC is based on (1) respondents self-reporting to direct questions, as used in the 5-yearly Youth Delinquency Survey (*Monitor Zelfrapportage Jeugd*, MZJ plus (2) data from the Dutch criminal justice system. This study has a particular focus on innovative methodology.

The research question was: building on the insights from the Youth Crime Monitor (MJC), what research method or combination of methods can be used to estimate the percentage of juveniles in the Netherlands guilty of the following crimes:
(1) online threat,
(2) the online distribution of sexual images of minors and
(3) logging in to a computer/website without permission/notification, whether or not accompanied by a change of data.

’Juveniles’ refers to young people aged between 12 and 22.

Based on an initial exploration in the research proposal, four methods were selected for further research: methods for collecting survey data from respondents, namely the Network Scale-Up Method and randomised response, and two methods using data already available. The methods were first assessed on whether they can provide reliable and valid estimates of the size of hard-to-count populations, so-called 'dark numbers'. This question deserves a comprehensive answer in the case of the two methods with little or no experience in large-scale applied research, the Network Scale-Up Method (NSUM) and Social Media Text Profiling (SMTP). For the other two methods, randomised response and capture-recapture, which are already part of the standard arsenal of policy research, the assessment focuses on more practical questions. In the case of randomised response, the question is whether the technique produces better results than direct survey as used in the MZJ. In the case of capture-recapture, the main criterion is the availability of adequate data from the Dutch police registration. For each method, the types of questions that can be answered were investigated, such as questions about modus operandi and offender characteristics, and whether there are specific requirements for investigating online offences and for research among juveniles. Attention was also paid to the anticipated costs of using the methods and, where appropriate, to the availability of data.

In the following sections, we briefly summarise the main findings on the four methods in the study, and draw conclusions on the contributions that the methods can make to estimating online crime. Based on these conclusions, we make recommendations for possible follow-up actions. The final section gives a concise and coherent summary of the main conclusions and recommendations concerning the four different methods.

## *2. Network Scale-Up Method (NSUM)*

NSUM asks respondents in a survey to count the numbers of people they know in the sub-population whose size needs to be estimated. In addition, the number of contacts in a number of sub-populations of known size is also requested. These data make it possible to estimate the size of the respondents' network and the size of the sub-populations not yet known. Because information about the respondent's own behaviour is not asked for, it is assumed that, less socially desirable answers are given, producing better reporting results concerning matters that may be sensitive.

*Conclusions on potential contribution*

We feel that NSUM is an interesting method for estimating the size of hard-to-count populations. The method probably makes it easier for respondents to report deviant behaviour. However, the experience with NSUM in large-scale surveys and population research is still limited. In addition, we did not encounter any thorough validation study comparing NSUM with other methods. The literature studied does not mention any experiences with research into online crime or among juveniles. There are few reasons to assume particularities for research among young people. The method does not work well for surveying offences that are in principle easy to commit individually and about which little is known in a social circle. Computer intrusion may well fall into this category. On the positive side, the method is relatively simple to include in a survey, without many additional costs and facilities. No larger samples are needed than for regular surveys. However, questionnaires will be slightly longer, survey time will increase and the construction of the questionnaire will require extra time and attention because asking for subpopulations of known size requires careful preparation and selection. When investigating online offences, attention should be paid to the question of how to define a relevant and realistically existing network of respondents on social media.

Our conclusion is that NSUM offers sufficient opportunities to provide information on the prevalence of online offences. The information it provides is probably more limited than from other survey methods such as direct response and randomised response surveys. After all, we cannot assume that respondents are fully aware of the behaviour of everyone in their social circle. The data that can be asked for is probably limited to information on offenders and possibly some easily identifiable characteristics of offenders and offences. So there is probably little point in asking questions about whether an offence, for example, has a foreign victim.

*Recommendations*

A first step to gain more insight into the possibilities of the method could be to carry out a simulation study. In such a study, the method is applied to a population (self-generated) in which all properties are fully known. In this way, the reliability of NSUM estimates can be tested quickly, using limited resources and in a controlled manner. In addition, it is also possible to determine the influence of violating the main assumptions of the method and measures to limit that influence.

We also feel it is worth considering carrying out a validation study of NSUM, for example in the form of a comparison with an estimate based on a direct survey. This could be done by comparing the results of a study with NSUM with results from the MZJ. But it is better to use the two research methods under as many equal conditions as possible. This could be done at relatively limited cost, for example in an internet panel. We suppose that any differences between the methods will not be different when using different channels, i.e. an internet panel and face-to-face approach.

The literature makes it clear that there is potential for optimising the use of the method. The proposed data processing measures can be easily tested on readily available data from NSUM research or on generated data in a simulation study. The most promising provisions for data collection can be tested in a validation study.

Finally, the literature on NSUM does not contain any indications of how to answer the question of how to take into account the' virtual' network of respondents on social media. For a survey using NSUM, it is important to answer this question, for example by using recent literature on social networks or existing surveys among users of social media.

## *3. Randomised response*

Randomised response is also a method of asking questions about sensitive issues where respondents tend to conceal information on sensitive issues. It is now a relatively familiar and well-known method. The respondents response can be 'mixed' with answers decided using prespecified probabilities in various ways, for example by throwing dice, so that the true response of each individual respondent can no longer be determined, but can be estimated based on the known probability distribution for the whole group. Randomised response therefore offers extra protection of anonymity for respondents on sensitive subjects, compared to direct surveying, at the cost of a (possibly considerably) larger sample. The method has been used to estimate, among other things, the prevention of social security fraud, theft in the care sector and doping in sport.

*Conclusions on potential contribution*

Randomised response can, in principle, collect the same information about offences as ordinary direct questioning such as that used in the MZJ, for example. This means that there are few restrictions on the information that can be obtained. The questionnaire of a survey can be adapted relatively flexibly to the information needed. Questions can be asked about perpetrators, ways in which they have acted, the motivation and possible proceeds of the offences, possible victims abroad, etc. However, questions about very specific behaviours or characteristics may be particularly threatening for respondents. In addition, the probability of finding significant numbers of respondents in a sample that meet the requested characteristic is reduced. These are, however, limitations of any form of questionnaire-based research.

It is also possible to establish statistical relationships between different group characteristics, for example admitting an offence on the one hand and gender, age or education on the other. For now, reservations should be made with regard to asking the exact frequency in which an offence has been committed during a certain period of time. This specific technique should first be further developed. However, good experiences have been gained with the questioning of frequency classes ("You have never/1-3 times/... / done more than 12 times?"). There are no indications that particular problems or challenges can be expected in research with randomised response to online crimes or among juveniles.

An important consideration in the choice of randomised response is the size of the required samples. In order to obtain information comparable to the MZJ, a sample of 10,000 respondents for one of the age categories might be needed, an impossible task. If follow-up questions are used, the required sample size decreases substantially. Nevertheless, we expect that the use of randomised response in the MZJ will be a major intervention and will lead to a substantial increase in costs. It is not known whether juveniles find it difficult to admit to offences committed online or offline. It would therefore be premature to propose randomised response for the MZJ.

*Recommendations*

A comparative study of randomised response with a conventional direct survey could determine any underreporting of online offences by juveniles and how large or serious such underreporting is. Here the same questions about online offences would be asked directly and in a randomised way to two different groups of respondents from the same population.
Based on the results, it could be decided whether or not to include randomised response in the MZJ. Alternatively, it is possible to regularly compare randomised response with direct questions.

In order to curtail the costs of such a study, it could be included in an internet panel. It may be sufficient to answer a limited number of questions in an omnibus questionnaire that combines several small studies. In a comparison via an internet panel, we also assume that any difference between direct questioning and randomised response in an internet panel is no different from a face-to-face survey. It would also be possible to make a rough first comparison, for example only for the offence with the highest prevalence. This comparison would only be aimed at investigating whether young people in general react differently to the different methods.
Another option to reduce costs is to test only differences in the reporting of groups of offences, e. g. digitised offences taken together. If a randomised response does indeed result in a significantly higher reporting, the subsequent investigation will result in smaller samples being sufficient. It will therefore become easier to differentiate the underreporting for different parts of the population.

## *4. Social Media Text Profiling (SMTP)*

SMTP consists of text analysis (text mining) of social media messages that automatically assigns certain qualifications to the messages, for example whether there is a threat. SMTP is of course only suitable for analysing and estimating online offences that produce sufficient text, i.e. generally online offences with a clear social or interactive component, such as threats, extortion, possibly the distribution of sexual images of minors, and less for, for example, computer hacking. There are also techniques for analysing visual material. These were not included in this study.

*Conclusions on potential contribution*

SMTP techniques can determine the roles of various stakeholders, e. g. perpetrator, victim, accomplice, spectator, and can identify the authors of messages. In principle, the results of these analyses make it possible to count and/or estimate the number of perpetrators and victims of, for example, the threat. Various characteristics can be attributed to the different stakeholders, such as the usual background characteristics such as age, gender and education, but also attributes such as personality traits and political orientation that are more difficult to identify. We did not encounter any statements about whether cybercrime can be traced in, or entering from, other countries in the literature consulted.  There are no reasons to assume that juveniles present a particular analytical challenge for the method. It is possible that the analyses need to be carried out more often because of a rapid evolution of specific language usage among juveniles on social media.

Social media analysis offers the most direct way to analyse online offences. The analyses can be characterised as a form of digital investigation into traces of online crimes. Other methods ask questions to potential perpetrators or victims, or use data about reports to the police. In all cases, there are more links between the concept of cybercrime and research data.
In addition, the analysis of social media data promises to be a relatively cheap (and therefore attractive) alternative or supplement to the usual methods. In a minimum variant, a study could be carried out by a team of a senior researcher and two junior researchers, each equipped with a powerful PC.

However, the approach is still primarily a promising idea. Based on our research, we conclude that there is as yet no method available in this area that can realistically be used to estimate the extent of an online offence. Parts of the required approach have been developed to a large extent and are already being used operationally. However, there are no examples of research that has integrated the various components into a census or estimation of a size of a population. According to the literature, a crucial component, (open class) authorship attribution, has not yet been sufficiently developed to perform realistic analyses of authorship of online communication with good results. Another limitation is that there are no data publicly available to make national estimations of online crime with SMTP. In principle, the data from social media are not publicly available and are not available for analysis. However, the National Police is able to obtain access to the data from Twitter.

*Recommendations*
The approach should be re-assessed by an expert in SMTP in consultation with an expert in estimation methods. The individual components could then be combined into one reliable estimation method. With the limited data available, it would then be possible to map part of the research area. The National Police has access to Twitter data to carry out this exercise. Although an estimate with these data is not representative of the population as a whole, repeated estimates over a number of years with additional analyses could be taken as an indication of a national trend.
One way to get a more complete national picture would be to make a fully developed estimation method available to providers and to encourage them to report the results periodically.

### *5. Capture-recapture and mining police data*

Capture-recapture methods provide an effective, efficient and proven method for estimating the number of offenders or victims of crime and are an attractive alternative to surveys. This is a group of statistical methods which, based on the number of times a person appears in a record, can be used to estimate the number of persons belonging to the same group but not registered. Often police records or registrations of other enforcement authorities are used as a source for the data. Previous examples include estimates of drink-driving, firearm possession, illegal residence in the Netherlands and domestic violence.

*Conclusions on potential contribution*
These methods could be used to obtain information on the number of perpetrators, offenders and characteristics of the offence, such as direct financial damage. In principle, the unknown part ('dark

number') can be estimated for all data registered. A limitation is that it is not possible to estimate phenomena or characteristics which are not recorded in a standardised way. Examples include the motives of offenders or modus operandi, perpetrators or victims abroad. Nor is it possible to estimate the number of incidents or offences.

Online offences are a relatively new form of crime. Police records have not yet been set up sufficiently to adequately record this form of crime. In this investigation, we examined whether the police registration contains sufficient data on the three selected offences to make estimates based on capture-recapture. We developed a classifier by text-mining police data to automatically retrieve data on online offences from police registration. The dataset used comprises statements from the Dutch Police's National Law Enforcement Database (BVH), using reports from the years 2013-2015. These were obtained by an extraction from the complete BVH registration of the years in question, using a very broad query consisting of a large number of terms related to online phenomena and activities. A sample from the data extraction is annotated. Text mining of the annotated data was used to develop classifiers. Using the best classifiers for each of the three selected online offences, all available records were classified, i. e. we determined whether one of the three offences had been committed. A count was then made per registered suspect, per offence, per year of the number of times that an official report had been drawn up against the suspect. The census shows that for two of the three selected offences, online threat and online distribution of sexual images of minors, there are sufficient' captures' and' recaptures'  to make estimates for the selected years.

*Recommendations*
The outcome of the text mining of BVH offers the possibility to carry out an investigation for the two offences in order to make an initial estimate with capture-recapture. However, the classification of the offences recorded is still very broad. No distinction has been made according to the seriousness or different manifestations of the offences. If it is decided to make estimates of capture-recapture, it is important to first examine whether a further classification of the offence in question is appropriate and whether it can also be registered by means of annotation. It is wise to annotate a larger sample than was possible within the scope of the present investigation. A second important conclusion of this exercise is that text mining appears to be an interesting method to efficiently and effectively unlock non-standardised data from BVH and to make it suitable for capture-recapture estimates.

## *6. In closing, our main conclusions and recommendations are as follows*

Our preliminary investigation has not identified any ‘ready-to-go’ method of estimating online juvenile offences. Important questions remain open for the two newer methods, NSUM and SMTP. The two more familiar methods, capture-recapture and randomised response, can be used if desired. Our main conclusions and recommendations are as follows:

*Capture-recapture*
The text mining study shows that police registration for two of the three selected offences, online threat and online distribution of sexual images of minors, contains sufficient data to make estimates of capture-recapture. However, the division into three online offences is still very rough and does not distinguish between serious offences or different manifestations. It is advisable to first determine whether a meaningful further classification of the offences can be made. Subsequently, a sufficiently wide sample of cases should be annotated on the basis of this classification.

*Randomised response*

Randomised response offers extra protection of anonymity for respondents regarding answering questions that may be sensitive.  Larger samples are needed in comparison to direct questioning. Both methods can answer similar questions.  It is not clear whether there is sufficient reason to incur the additional costs for larger randomised response samples. First of all, the question is whether randomised response compared to the direct questions in the MZJ results in a significant and interesting improvement of the reporting. A first, cost-saving indication for this could be obtained by including the comparison in an internet panel.

*NSUM*

NSUM is likely to be able to provide relatively simple estimates of online crime, without drastic or costly modifications to existing research such as the MZJ. The method appears to be particularly suitable for estimating offences with a social component, i. e. online threat and online distribution of sexual images of minors. The method makes it possible to estimate the number of offenders and possibly a number of easily identifiable characteristics of offenders and offences. NSUM is therefore not a substitute for other methods, but may serve as part of a triangulation. We think it would be wise to first try out the achievements and some of the more interesting modalities of the method in a comparative study. Such a study should provide insight into whether NSUM provides significantly better estimates of selected online juvenile offences and what would be the best way of doing so. In the shorter term, a simulation study would provide a better insight into the possibilities of the method and the validity and reliability of the estimates that can be made with it, with relatively limited resources.

*SMTP*

Finally, SMTP. Although this approach has its charms, it cannot yet be called a real estimation method. It still needs to be assessed and tried out to explore its potential for making reliable estimates. Data files with national coverage are not yet available. This means that an SMTP estimate can only map a limited proportion of online offences. However, if an SMTP analysis is repeated over time, it could be used to identify trends.