Submission for the 2015 Amsterdam Privacy Conference (APC 2015)

PRIVACY AND SURVEY RESEARCH: GUARANTEEING ANONYMITY WHEN DISCLOSING SENSITIVE INFORMATION

MSc. Maikel Groenewoud

Researcher and data-analyst

2015

Regioplan (Policy Research Agency)

Amsterdam, The Netherlands

+3120 – 5 315 340

maikel.groenewoud@regioplan.nl

Track: Special paper session

Abstract

The main focus of this article is to illustrate how the anonymity of respondents' answers in survey research can be guaranteed by using the randomized response method. In this article we discuss the research approach and results of a study concerning rules violation. In this type of study respondents often tend to give socially desirable or generally accepted answers. Furthermore there is often an underrepresentation of offenders in the response because they are more likely to refuse cooperation to survey research. We discuss these effects with a study about violations by Dutch companies of a Dutch law concerning the legal circumstances under which it is allowed to employ foreigners, the so-called 'Wet arbeid vreemdelingen (WAV)'.

The most important aspect of the RR method is that respondents' answers depend on the outcome of a randomization device, a game of chance so to speak. This in theory gives the individual respondent the guarantee that he or she can answer truthfully to sensitive questions without any risk of being found out. Neither the investigator nor the interviewer knows if the respondent has answered truthfully or has given a fixed answer in accordance with the rules of the 'RR game'. In RR the anonymity of the respondents' answers is guaranteed by the introduction of the element of chance. That said, even though it is not possible to determine if individual answers are correct, it is still possible for the researcher to make an estimate for the entire group (population) concerning the total level of rules violation or engagement in the sensitive behavior. This is possible because based on the probability distribution of the randomization device, the percentage of respondents that were supposed to answer truthfully can be predicted.

The exponential increase in data being generated and collected in modern society has caused increasing concerns about privacy issues, especially in relation to the disclosure of sensitive information. Randomized response could potentially ease some of these concerns by guaranteeing the anonymity of data being collected.

Based on our research approach it is possible to derive control mechanisms which can be used to generate correction factors. The use of these correction factors leads to better estimates of the levels of rules violation. The outcomes of the study contribute to the further optimization of the randomized response design.

Keywords: sensitive topics, randomized response, anonymity, survey research, big data

1. INTRODUCTION

The main focus of this article is to show how the anonymity of respondents' answers in survey research can be guaranteed by using the randomized response (RR) method. In contemporary modern society people are often asked or obliged to disclose information, often sensitive information. As a result of technological advancements such as the internet, the process of information disclosure has changed and it has in many ways become easier to collect data about people. Mobile computing, cloud computing and the so-called 'internet of things' are just some examples of recent innovations. The ease of data collection and the exponential increase in data being generated and collected has caused increasing concerns about privacy issues, especially in relation to the disclosure of sensitive information (see for example Cranor et al. 2000). Randomized response could potentially ease some of these concerns by guaranteeing the anonymity of data being collected (see for example Sun et al. 2014).

Participating in studies concerning sensitive or fraudulent behavior can be threatening in various ways, this threat has an intrinsic and extrinsic component (Lensvelt-Mulders and De Leeuw 2002). Questions about sensitive topics are often very personal in nature and that poses an intrinsic threat. When answering these types of questions during a survey, respondents are confronted with their own unacceptable or undesirable behavior which can have a negative influence on the way they perceive themselves.
In addition to the intrinsic component of the threat there is also an extrinsic component as a result of perceived threats coming from the external environment (either real or imaginary). Certain behaviors can for instance result in penalties. In addition to criminal penalties, certain behaviors also carry the risk of social sanctions. For instance, one could be excluded from one's own peer-group if behavior is exhibited which is deemed socially undesirable or unacceptable.
Because of the intrinsic and extrinsic threat respondents often refuse to participate in research on sensitive topics and when they do, they have the tendency to skip the most sensitive questions or to only give socially desirable or generally accepted answers (Lee 1993).

To gain insight into the level of violation by Dutch companies of a law concerning the legal circumstances under which it is allowed to employ foreigners, the so-called 'Wet arbeid vreemdelingen (WAV)', we conducted a study using randomized response (Groenewoud and Van Rij 2007). This study was based on a large dataset that deals with illegal workforces in the Netherlands and the social context thereof. We applied a heuristic approach for the detection of cheating and non-compliance and used this to adjust the estimated levels of rules violation. A unique aspect of this research study was the availability of semi-validation data, which made it possible to better evaluate the quality of the RR estimates.

Research studies have indicated that RR is superior in eliciting admissions to sensitive or fraudulent behaviors compared to direct questioning (van der Heijden et al. 2000). Having said that, there are certain complications when using RR, mainly as a result of cheating and non-compliance with the RR instructions (see for example Coutts and Jann 2011). Cheating mostly refers to respondents who engage in the sensitive behavior being investigated, but are unwilling to admit this when asked about it. Non-compliance usually refers to respondents who are not engaged in the sensitive behavior, but refuse to adhere to the RR instructions when they are forced to say that they have engaged in this behavior (when in reality they have not). The probability key involved in RR however, makes it possible to compensate for cheating and non-compliance.

2. DATA: ILLEGAL WORKFORCES IN THE NETHERLANDS

The data was collected to gain insight into the level in which Dutch companies violated a law concerning the legal circumstances under which it is allowed to employ foreigners. It is not easy to determine the number of illegal workers, because this is a so-called 'hidden population' of which the size and whereabouts are unknown (see also Sikkel et al. 2006). There is no list containing all employers who make use of illegal workers. There is also no list containing all illegal workers. There are certain list containing some of these people, such as police registrations and medical records, but these are far from complete.
The data were gathered via a computer-assisted randomized response online survey. Semi-validation data were available consisting of an anonymized sample of companies from seven high-risk sectors which were audited by the Dutch Labor Inspection a year prior to the study. This is a unique group, we did not know at which individual companies the Labor Inspection had established violations of the law, but we did know how many percent of them were fined by the Labor Inspection. We also knew the total number of audited companies that were invited to participate in this study and what percentage of them were fined. The control group is essential in our evaluation of the method used for correcting the initial RR estimates. Using the control group, we can ascertain whether fraudulent employers are less inclined to participate in research concerning violation of the WAV. This

control group also allows us to determine if employers who do participate are more inclined to give socially desirable or generally accepted answers.

For the study, a random sample was drawn from seven high-risk sectors. The sample was expanded with employers outside these seven sectors allowing us to generalize the results to the entire population consisting of all Dutch employers in 2006. The questionnaire was completed by 7.293 respondents on behalf of their company. For the control group of the Labor Inspection, 2.449 companies were contacted, of which 436 enterprises were willing to participate. The response rate in the various sectors ranged from 15-23 percent. The fieldwork took place in the period January 2007 - February 2007 and concerned rules violations in the year 2006.

3. RANDOMIZED RESPONSE METHOD

There are several techniques available to investigate sensitive topics. We here list some of the most well-known ones used for eliciting sensitive information:

- Direct questioning (see for example van der Heijden et al. 2000);
- Item or unmatched count technique (Droitcour et al. 1991);
- Three-card method (Droitcour et al. 2001);
- Randomized response (Warner 1965).

The last three techniques in the above list can only be used to generate population estimates. If you are interested in exactly classifying the individuals that have engaged in the sensitive behavior, one would require individual estimates. However, for many purposes it will suffice to generate population estimates of the prevalence of the sensitive behavior. Regardless of the technique used, there are of course also other factors to consider such as anonymity clauses and the wording of questions (Tourangeau and Yan 2007).

The original randomized response method was developed by Stanley Warner (1965). The basic idea is the introduction of an element of chance through a randomization device to guarantee the anonymity of individual respondents' answers. In RR the respondents' answers depend on the outcome of a game of chance. This in theory gives the respondent the guarantee that he or she can answer truthfully to sensitive questions without any risk of being found out. Neither the investigator nor the interviewer knows if the respondent has answered truthfully or has given a fixed answer in accordance to the rules of the 'RR game'. That said, even though it is not possible to determine if individual answers are correct, it is still possible for the researcher to make an estimate for the entire group (population) concerning the total level of rules violation. This is possible because based on the probability distribution of the randomization device, the percentage of respondents that were supposed to answer truthfully can be estimated.

Forced randomized response method with virtual dice

Since the inception of the basic RR method, several other forms of this technique have been developed (see for example Lensvelt-Mulders et al. 2004; Fox and Tracy 1986). In this study (Groenewoud and Van Rij 2007), we have applied the so-called *forced randomized response method* with virtual dice. The questions were asked via a computer-assisted randomized response online survey. The respondents are asked to throw the two dice without the interviewer or researcher knowing the outcome. Depending on the sum of the dice, the respondent must answer 'yes' or 'no' or answer truthfully. It for instance can go as follows:

- Throw the dice.
- If the outcome is 2, 3 or 4, always answer yes (regardless the content of the question).
- If the outcome is 11 or 12, always answer no (regardless the content of the question).
- If the outcome is 5, 6, 7, 8, 9 of 10, always answer truthfully (regardless the content of the question).

At most of our RR-questions, the above RR-instructions were used but we also varied this a bit resulting in different probability distributions for the various RR-questions..

The fraction of rules violators can be computed as follows (Fox and Tracy 1986):

(1) $$\hat{\pi} = \frac{\lambda - \theta}{p}$$

With variance:

(2) $$\hat{\text{var}}(\hat{\pi}) = \frac{\lambda(1-\lambda)}{np^2}$$

with $\hat{\pi}$ denoting the estimated proportion of rules violators, λ denoting the proportion given yes-answers, θ denoting the proportion forced yes-answers and p the probability of being asked to answer truthfully.

4. DIFFICULTIES WHEN USING RANDOMIZED RESPONSE

The RR method is generally accepted as one of the best currently available methods for estimating levels of rules violation, but there are a number of concerns when applying this method. Due to the RR instructions the questions are formulated in a more complex fashion than is the case when using direct questioning. This may lead to higher numbers of respondents who do not understand the questions and as a result possibly answer incorrectly. There is also the problem of cheaters, people who are in fact 'guilty' of the sensitive behavior being investigated yet will never say so when asked about it. When these cheaters are supposed to answer truthfully in accordance to the RR instructions or are 'forced' to say that they are engaged in the sensitive behavior, they deliberately ignore the RR instructions because they do not want any record of their undesirable or unacceptable behavior. Cheating is not a problem unique to RR though but also applies to other methods such as direct questioning, it is a problem related to inquiries into sensitive behavior in general. There are also people who find it difficult to answer that they have engaged in any sensitive behavior when that is not the case. However, sometimes the RR instructions would force them to do so, but they then choose not to comply with those instructions.
So whether it happens consciously or unconsciously, deliberately or by accident, whether it concerns actual violators or not, it is likely that certain people will not always adhere to the RR instructions. Furthermore, despite the guarantee of anonymity, violators are still probably less likely to participate in research concerning the level of rules violation than people who have not engaged in the behavior being investigated. As a result data collected using RR will be biased to a certain extent. However, the probability key involved in the implementation of RR, allows for several cheating detection methods which can be used to correct the initial RR estimates. When using other data collection methods it is far less possible to accurately correct for cheating/evasive answers.

5. CORRECTION FOR CHEATING AND NON-RESPONSE

In this study conducted in 2007, we used an intuitive or heuristic approach to classify potential cheaters and non-compliers. In the survey several RR questions were included which were used to compensate for cheating and non-compliance.

Classification of cheaters (internal validity)
We examined the response patterns at the four most sensitive questions. Based on the statistical distribution of the dice, we determined the likely percentage of respondents that can be expected to give a socially acceptable or desirable answer at each RR question. This was done by multiplying the likely statistical probabilities that someone should give a socially acceptable answer at the individual RR questions.

The estimated or likely percentage of respondents with only socially acceptable answers (Pct_{es}) was then subtracted from the observed percentage of respondents with only socially acceptable answers (Pct_{os}) to determine the percentage of cheaters (Pct_{cheat}).

(3) $$Pct_{cheat} = Pct_{os} - Pct_{es}$$

For the analysis we focused on the four most sensitive RR-questions in the survey. It is import that the RR-questions included in this type of analysis are about equally sensitive. If for instance one of them was considerably less sensitive than the rest, one would have far less difficulty giving a socially undesirable answer to that question. As a result, the percentage of people with only socially desirable answers to all four of these questions would be lower.

Comparison to control group (external validity)
In order to assess both the internal and external validity of the results, the study contained a control group. It is an anonymized sample of companies from seven high-risk sectors who a year prior to the survey were inspected by the Dutch Labor Inspection. This is a unique group. We knew not at which individual companies the Labor Inspection established violations of the WAV, we only knew in total for how many percent of them a penalty report was made. The control group forms the key to our analysis of the correction factors. Firstly, by using this group we can establish whether fraudulent employers are less likely to participate in research concerning the

WAV and secondly whether employers who do participate give socially desirable answers. By making these two comparisons, we have determined correction factors . These correction factors are ultimately used to correct the answers of all employers (n = 7293) and that leads, in this case, to higher estimates. The correction factors are defined as follows:

$$C = \frac{Pct_{\text{orv}}}{Pct_{\text{erv}}} \tag{4}$$

With:
C = Correction factor
Pct_{orv} = Real or observed percentage of rules violation in control group (as established by the Dutch Labor Inspection)
Pct_{erv} = Estimated percentage of rules violation in control group

6. RESULTS

In this paragraph we discuss the analysis of the cheaters and other people who don't adhere to the rules of the RR-game (internal validity). Secondly, we look at potential non-response bias, using the control group (external validity). Then we present the estimated levels of rules violations before and after applying correction factors.

Analysis of cheaters

The percentage of respondents with only socially desirable answers at the four most sensitive RR-questions, is 61.5 percent. However, the expected percentage with only socially desirable answers, is 21.3 percent. That means that 40.2 percent of the respondents can most likely be considered cheaters. This is a significant group for which the estimated levels of rules violation needed to be corrected. In theory it is of course possible that not a single company employs illegal workers. If anyone was 'forced' by the dice to answer truthfully, the chance of a socially desirable response would be 100 percent in this hypothetical scenario. The chance that someone would have a socially desirable answer at all four RR-questions would then be 43.5 percent. This is a lot higher than the number previously found (21.3%), but still nearly 20 percent lower than the observed percentage of respondents who gave socially desirable answers at each RR-question (61.5%). In this hypothetical case, it is more likely that 61.5 percent would only have socially desirable answers. However, the underlying assumption here is that zero percent of respondents directly or indirectly employ illegal employees. Given the findings of the Dutch Labor Inspection, it seems highly unlikely that this assumption is correct.

Analysis of non-response

Of the 436 companies in the control group it is known that at 76 of them, violations of the WAV have been detected. This amounts to 17.4 percent. The percentage of cheaters in the control group is 35.6 percent. If we apply the correction for cheaters to the control group, this leads to an estimated percentage of 17.8 percent that violate the WAV. This means that our estimate only exceeds the threshold of 17.4 percent by 0.4 percentage points resulting in a correction factor of 0.98 (=17.4/17.8). This still somewhat underestimates the level of rules violation though since we know that 22 percent of the gross sample have violated the WAV. This means that offending companies have participated less often in this study which is consistent with the assumption that companies that have violated the rules will be less willing to take part in studies like this. Taking this into account leads to a correction factor of 1.24 (= 22/17.8).

Estimates of violation of the WAV by sector

In order to estimate the extent of non-compliance to the Wav, the following steps need to be taken:
- Step 0: Results before corrections. This gives the 'standard' RR-estimates.
- Step 1: Correction for cheaters on the basis of socially desirable response patterns.
- Step 2: Correction for non-response based on comparison with the control group.

Table 1 shows the levels of rules violation after each of the steps described above.

Table 1 **Estimated levels of rules violation in 2006 (of the Wav)**

	Step 0: Initial estimates	Step 1: Correction for cheaters	Step 2a: Correction for non-response (factor = 0.98)	Step 2b*: Correction for non-response (factor = 1.24)	N (number of companies in survey)
Dutch companies	-2.9%	7.6%	7.4%	9.4% (±1.7%)	7293

* The 95%-confidence intervals are in brackets.

7. DISCUSSION

Various studies have shown that the RR method gives better insight into hidden populations than direct questioning when conducting survey research. There are, however, a number of drawbacks to this method. Without any control mechanisms built into the research design, the results are not always useful and there are even negative estimates possible as was initially the case in this study. By correcting for cheaters we did arrive at positive estimates, but this came at the expense of a large group of respondents. It's very important to stress here that positive estimates by themselves are not a sufficient guarantee for valid results. A control group as included in this study offers a way of testing the validity of the results. The outcomes for the control group made it plausible that the corrected estimates are more valid than the estimates before correction. It should be noted here that it can't be completely ruled out that this particular group of respondents (the control group) answered differently from other respondents. The fact that they have been inspected by the Dutch Labor Inspection might influence their response process. Furthermore the Labor Inspection figures relate to a snapshot, while our study looks at the whole of 2006. Thus it is possible that the Labor Inspection on the day of their inspection finds no illegal workers or irregularities, but the employer the next day employs an illegal worker or otherwise violates the rules. The number of offenders at any given moment is usually less than the total number of offenders during a year.

Each correction step adds a certain degree of uncertainty to the estimates. Step 1 adds uncertainty through the smaller response numbers after correction for cheaters and step 2 through the fact that the control group may be different from the other groups. However, since the estimates for the control group after step 2 are very close to the actual percentage of rules violation and we have corrected for non-response, a strong case can be made that the estimates have become more accurate. Under the assumption that the other groups behave similarly to the control group, the confidence intervals would actually become smaller.

Randomized response and the 'big data' revolution

In light of the rapid developments in internet and communications technology, we briefly discuss the potential relationship between randomized response and so-called 'big data'. Big data can be very useful for gaining (new) insights into people, to target people for instance for marketing purposes and also to help people make better or more effective decisions. However, the exponential increase in data being generated and collected has also caused increasing concerns about privacy issues, especially in relation to the disclosure of sensitive information. Randomized response could potentially ease some of these concerns by guaranteeing the anonymity of individual characteristics being collected. Traditionally randomized response is used in the data collection process to give respondents a guarantee of privacy. Potentially this techniques can also be applied to data that has already been collected, for instance through direct questioning or from registration systems. This data can than be randomized by applying a randomization device as used in randomized response. The privacy of individuals would be guaranteed than because it isn't possible to determine which individual answers are 'truthful' and which are the result of the randomization device. It however would still be possible to generate population estimates about the behavior or characteristic of interest, based on the probability distribution of the randomization device.

Concluding remarks

We are aware that it is not always possible to include a control group into a study. But even without a control group, the study presented here serves as an example that under reasonable assumptions, corrections for cheaters and non-response (can) lead to more valid results. Our research contributes to the further optimization of the randomized response design. We believe that the RR method with control mechanisms as discussed in this article is a step forward for mapping hidden populations such as illegal workers.

LITERATURE

Coutts, Elisabeth and Ben Jann (2011). 'Sensitive questions in online surveys: Experimental results for the randomized response technique (RRT) and the Unmatched Count Technique (UCT). Sociological Methods & Research 2011 40: 169-193

Cranor, Lorrie F, Joseph Reagle and Mark S. Ackerman, Beyond Concern: Understanding Net Users' Attitudes about Online Privacy, MIT Press, Cambridge, Mass, USA, 2000

Droitcour, Judith, Caspar, Rachel A., Hubbard, Michael L., Parsely, Teresa L., Visscher, Wendy., and Ezzati, Trena M. (1991). The item count technique as a method of indirect questioning: A review of its development and a case study application. In Paul Biemer, Robert Groves, Lars Lyberg, Nancy Mathiowetz, and Seymour Sudman (Eds.), Measurement errors in surveys (pp. 185–210). New York: Wiley.

Droitcour, Judith, Larson, Eric M., and Scheuren, Fritz J. (2001). The three card method: Estimation sensitive survey items—with permanent anonymity of response. In Proceedings of the Section on Survey Research Methods, American Statistical Association. Alexandria, VA: American Statistical Association.

Fox, James Alan and Paul E. Tracy (1986). *Randomized Response. A Method for Sensitive Surveys.* Beverly Hills: Sage.

Groenewoud, Maikel and Coen van Rij (2007). *Naleving van de Wet arbeid vreemdelingen 2006. Onderzoek onder werkgevers.* Amsterdam: Regioplan Beleidsonderzoek.

Heijden, Peter G.M. van der, Ger van Gils, Jan Bouts and Joop Hox (2000). "A Comparison of Randomized Response, Computer-Assisted Self Interview and Face to Face Direct Questioning: Elicting Sensitive Information in the Context of Welfare and Unemployment Benefit", in: *Sociological Methods & Research*, Sage Publications, vol.28, no.4, p.505-537.

Lee, Raymond M. (1993). *Doing Research on Sensitive Topics*. London: Sage.

Lensvelt-Mulders, Gerty and Edith de Leeuw (2002), Beschermd door een dobbelsteen, *Facta,* 10, september, 2002, p.28-30.

Lensvelt-Mulders, Gerty, Joop Hox, Peter van der Heijden and Cora Maas (2004). Meta-Analysis of Randomized Response Research: thirty five years of validation, in: *Sociological Methods & Research*, Sage Publications, vol.33, no.3, p.1-30.

Lensvelt-Mulders, Gerty, Peter van der Heijden, Olav Laudy and Ger van Gils (2006). A validation of computer-assisted randomized response survey to estimate the prevalence of fraud in social security, in: *Journal Royal Statistical Society*, , vol.169.

Rij, Coen van en Groenewoud, Maikel (2009), 'Fraude gemeten?!' in: Boekenoogen c.s. *Methoden van beleidsonderzoekers: creatief en oplossingsgericht*, Den Haag: Lemma, p.59-68

Sikkel, Dirk, Peter G.M. van der Heijden and Ger van Gils (2006). *Methoden voor het schatten van de omvang van verborgen populaties, in het bijzonder illegale vreemdelingen.* Den Haag: Ministry of Security and Justice, WODC.

Sun, Chongjing, Fu, Yan, Zhou, Junlin and Gao, Hui (2014), Personalized Privacy-Preserving Frequent Itemset Mining Using Randomized Response, The Scientific World Journal, Volume 2014 (2014), Article ID 686151

Tourangeau, Roger and Yan, Ting, "Sensitive questions in Surveys", Psychological Bulletin, 2007, Vol. 133, No. 5, 859–883

Warner, Stanley L., 1965. Randomized response: a survey technique for eliminating evasive answer bias. Journal of the American Statistical Association 60, 63–69.