Heliyon 10 (2024) e27252

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

A new improved randomized response model with application to compulsory motor insurance

Ahmad M. Aboalkhair [a,b,**], A.M. Elshehawey [c], Mohammad A. Zayed [a,b,*]

[a] *Department of Quantitative Methods, College of Business Administration, King Faisal University, Al-Ahsa, 31982, Saudi Arabia*
[b] *Department of Applied Statistics and Insurance, Faculty of Commerce, Mansoura University, Mansoura, 35516, Egypt*
[c] *Department of Applied, Mathematical & Actuarial Statistics, Faculty of Commerce, Damietta University, New Damietta, 34519, Egypt*

ARTICLE INFO

Keywords:
Sensitive attributes
Non-sampling error
Randomized response technique
Privacy measure
Third-party liability insurance
Insurance inclusion

ABSTRACT

One of the challenges, when investigating sensitive attributes, or information that people tend not to disclose, through surveys is the ethical obligation to preserve the privacy of respondents. Although the randomized response method, originally suggested by Warner, allows estimating the proportion of such attributes within the population while maintaining confidentiality, the variance of the estimate consistently increases if the likelihood of selecting the question about sensitive attribute increases. The purpose of this research is to introduce a new three-stage RR model which provides an efficient alternative to Warner's model that allows more credibility from a practical perspective and apply the model to estimate noncompliance ratio in compulsory motor insurance. For the two models, a measure of privacy protection was calculated and a relation between this measure and the efficiency of both models was introduced. Efficiency comparisons indicate that the proposed model can always be made more efficient than both Warner's and Mangat & Singh's RR models. The proposed model, with specific parameter selection, was applied on a selected population and proved a practical reliability. The noncompliance ratio to obtain compulsory motor insurance was estimated by both a point and a confidence interval. This estimate provides a basis to predict third party motor insurance inclusion.

1. Introduction

Surveys are widely used by researchers as a principal technique to assess attitudes and behaviors in various situations and in many areas. When survey questions deal with sensitive or embarrassing issues, (such as, the use of drugs, psychiatric conditions, cheating behavior, fraud in insurance, …etc.), then research and practical difficulties arise, mainly, due to the ethical obligation of the researchers to preserve the privacy of their respondents. Other issues of concern may also appear, among which is the non-sampling error bias resulting from "response effects" such as refusing to respond or untruthful reporting. Before the randomized response technique (RRT) was introduced, little progress has been made towards the solutions of these issues.

Randomized response (RR) is a method used in surveys in which the goal is to reduce or eliminate response errors when respondents are queried about illegal behaviors, sensitive or highly personal matters. A randomized response design indirectly acquires information from respondents by using a probabilistic random device through which the respondent selects a question from two or

* Corresponding author. Department of Quantitative Methods, College of Business, King Faisal University, Al-Ahsa, 31982, Saudi Arabia.
** Corresponding author. Department of Quantitative Methods, College of Business, King Faisal University, Al-Ahsa, 31982, Saudi Arabia.
E-mail addresses: aaboalkhair@kfu.edu.sa (A.M. Aboalkhair), a-elshehawey@du.edu.eg (A.M. Elshehawey), mzayed@kfu.edu.sa (M.A. Zayed).

https://doi.org/10.1016/j.heliyon.2024.e27252
Received 12 April 2023; Received in revised form 22 February 2024; Accepted 27 February 2024
Available online 29 February 2024

more questions, one of them, at least, is sensitive. This happens without revealing to the interviewer which question has been chosen. Since the types of responses are the same for each question, no respondent (or answer) can be classified with certainty a posteriori with respect to the sensitive characteristic. This leads to that, when using RR, it is assumed that the information received from respondents are truthful and sufficient for estimation purposes.

In addition, for any given probability distribution used in the design, it is possible to compute unbiased estimates of the parameters, such as the proportion, associated with the sensitive attribute. Therefore, RRT enables the researcher to assess attitudes and behaviors of populations which direct questions often fail to assess.

Randomized response (RR) was first proposed by Warner in 1965 [1]. He built his method on the premise that respondents' cooperation should be better if the questions allow answers that reveal less, even to the interviewer. The method is essentially based on a random device that makes the interviewee responds with answers that furnish information only on a probability basis. Although Warner's method allows collecting responses about sensitive issues while maintaining confidentiality, the estimate of the proportion of the population with a sensitive attribute has additional variance due to the random device.

Since Warner's proposal, the randomized response technique has been extended to a number of directions by several authors who focused on reducing the variance of the estimate and improving the efficiency of the model, whether by suggesting parameter selection according to specific criteria that ensures minimizing the variance, or through using different estimation methods, or, mostly, via suggesting a design modification to the original model of Warner.

In general, the efficiency of the randomized response estimate depends on all parameters involved in this estimate. However, choosing certain values for the parameters involved in the estimate could result in a minimum variance for the estimator, these same values could make respondents more suspicious and therefore the bias rising from incomplete or untruthful answers will dominate the mean square error, which in turn decrease the efficiency of the estimate.

Several authors have suggested that the efficiency of randomized response technique can be improved by methods based on alternative estimation procedures. Few researchers suggested using information about auxiliary characteristics (Covariates) to improve RRT estimate precision [2–8]. Other researchers have dealt with the idea of adopting a Bayesian approach for the RRT [9–15]. Some developments suggested a stratified sampling to improve the estimate of the sensitive attribute [16–24].

The main approach, adopted by most research, for increasing the efficiency of RRT is based on design modification. Different modifications of Warner's [1] RR were developed by various authors, just to name a few, Greenberg et al. [25]; Moors [26]; Raghavarao [27]; Mangat and Singh [28]; Kuk [29]; Mangat [30]; Singh et al. [31]; Bhargava and Singh [32]; Singh et al. [33]; Haung [34]; Chang et al. [35]; Gupta et al. [36]; Gjestvang and Singh [37]; Perri [38]; Abdelfatah and Mazloum [39]; Batool et al. [40]; Singh and Gorey [41]; Tarray and Singh [42]; Narjis and Shabbir [43]; Singh and Suman [44].

Singh et al. [45] proposed a model in which three randomizing devices are given to the respondents, the first two devices carrying the same two statements: (i) I belong to sensitive group A1 and (ii) Go to next randomization device, and the third device is the same as in the model suggested by Singh et al. [33]. The latter was a modified version of Greenberg's unrelated question model [25] using a randomizing device that carries three statements: (i) I belong to sensitive group A1; (ii) I belong to non-sensitive group A2 and (iii) Report "No".

In this paper, a new model is considered, that differs from the Singh et al. [45] procedure in the sense that the third randomization device used in the proposed procedure is same as Warner's model [1].

As a suggested application of the proposed RR model, the case of obligatory motor liability insurance is examined. This latter type of insurance is looked at as an ongoing concern, especially in regions where insurance culture and awareness are still developing.

In 2009, a report by The World Bank shed some light on this issue in developing countries and emphasizes the importance of this type of insurance for road safety, personal responsibility, and safe transport systems in these countries. The report raises an issue of awareness pointing that car owners tend to think of motor insurance as a type of tax they can freely avoid rather than as a safeguard against personal liability, a concept that is unfamiliar to the public [46].

The lack of awareness and negative perceptions of auto liability insurance motor liability insurance among drivers has an impact, not only on seeking insurance coverage, but also on insurance claims [47]. The sufficiency of obligatory motor liability insurance premiums (contributions), which is a direct reflection of car owners' compliance to the law, is a significant factor that affects the functionality of the insurance system as a whole [48].

In general, insurance rates and loss ratios are widely affected by the sufficiency or insufficiency of premiums, and rates can periodically increase due to this factor. Examples of such scenarios have been discussed by some authors, considering rate changes due to this along with other factors [49,50].

2. Materials and methods

2.1. The original Warner's model

Warner [1] developed a method for estimating the proportion of persons with a sensitive attribute, A, (π), without requiring the individual respondent to report his actual classification, whether it be A or not-A to the interviewer. The respondent is provided with a randomizing device (such as, a spinner) in order to choose one of two statements of the form:

(a) - I belong to sensitive group A (selected with probability p_1)
(b) - I do not belong to sensitive group A (selected with probability q_1)

Without revealing to the interviewer which statement has been chosen, the respondent answers “yes” or “no” according to the statement selected and to his actual status with respect to the attribute A. With a random sample with replacement of n respondents, the maximum likelihood estimate of π as given in Warner’s design [1] with appropriate change of notations is:

$$\widehat{\pi} = [\widehat{\alpha} - q_1][1 - 2q_1]^{-1}, q_1 \neq 0.5 \tag{1}$$

where $\widehat{\alpha}$ is the observed proportion of “yes” answers, $\widehat{\alpha} = n'/n$, n':the number of “yes” answer in the sample. If all respondents answer the selected statement truthfully, the resulting estimate will be an unbiased estimate with variance given by:

$$V(\widehat{\pi}) = \pi(1-\pi)/n + q_1[1-q_1][1-2q_1]^{-2}/n \tag{2}$$

It is clear from Eq. (2) that the variance of π estimator, according to Warner’s model, is greater than the typical variance of the proportion estimate, and that this variance is negatively correlated with the value of $p = 1 - q$ (the second fraction in Eq. (2)).

Warner [1] has established that, the variance of π decreases as $|p - 0.5|$ increases, however, choosing big values for p leads to the loss of the advantages of confidentiality and that the bias rising from incomplete or untruthful answers dominates the mean square error. That is because the efficiency of the randomized response estimate depends on the psychological reaction of the respondents to the particular randomized response design that persuades them to cooperate or not to cooperate.

This was the motivation behind suggesting the modified RR model presented in this work, that allows increasing the probability of selecting the sensitive question without making respondents more suspicious which increases untruthful responses, and, therefore, the variance of the estimate.

This is done by increasing the efficiency of Warner’s model by proposing a random multi-stage tool that allows increasing the value of p to a reasonable extent without significantly affecting respondents’ confidence in the tool and their truthfulness.

2.2. Proposed RR model

The proposed model is based on a three-stage random tool that is distributed to the selected sample as shown in Fig. 1. In the first stage (S1), each respondent randomly chooses one of two alternatives: the first is a yes/no question which indicates whether the respondent possesses a specific sensitive attribute or not, and the second is an instruction to skip to the next stage. If the respondent moves to the second stage (S2), the choice is made between the same two alternatives, as in the first stage. For the third stage (S3), if the respondent reaches it, the choice is made between one of two alternatives (yes/no questions), exactly the same as the original Warner’s model. At the end of the process, all responses are collected, some of which are "yes" and the rest are "no".

The probability of getting a "Yes" (α) can be expressed as follows (Eq. (3)):

$$\alpha = p_3\,\pi + q_3\,[p_2\,\pi + q_2[p_1\,\pi + q_1\,(1-\pi)]] \tag{3}$$

where:

π: The proportion of individuals, within the population, who belong to the sensitive group.

p_{4-s}:The probability for the question that the individual belongs to the sensitive group shows up at stage s, where s = 1,2,3 and $p_{4-s} + q_{4-s} = 1$.

The proposed estimator for the proportion of individuals, within the population, who belong to the sensitive group ($\widehat{\pi^*}$) is:

$$\widehat{\pi^*} = [\widehat{\alpha} - Q][1 - 2Q]^{-1}, Q \neq 0.5 \tag{4}$$

where $\widehat{\alpha}$ is the proportion of ‘yes’ answer obtained from the n sampled respondents and $Q = \prod_1^3 q_{4-s}$. To check the validity of Eq. (4), if

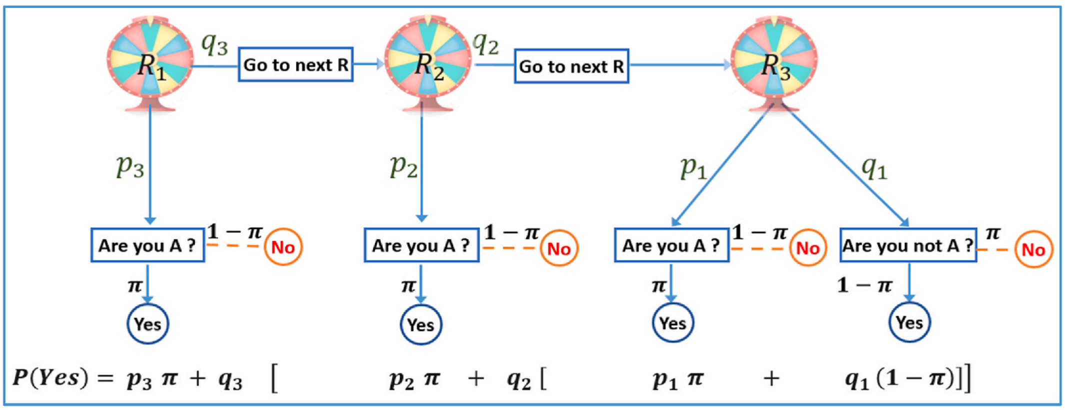

Fig. 1. Proposed model flowchart.

we replace Q with q_1, we can get the Warner's estimate as given by Eq. (1)

Properties of proposed estimator.

Since the random variable $n\widehat{\alpha} \sim Bin(n, \alpha)$, therefore $\widehat{\alpha}$ is considered an unbiased estimator of α, and the variance of $\widehat{\pi}^*$ can be expressed as follows:

Theorem 1. *The* variance *of the proposed estimator* $V(\widehat{\pi^*})$ *is given by*

$$V(\widehat{\pi}^*) = \pi(1-\pi) / n + Q[1-Q][1-2Q]^{-2} / n \tag{5}$$

Proof of Theorem 1. Using Eq. (4), the variance of $\widehat{\pi}^*$ is

$$V(\widehat{\pi}^*) = \mathrm{V}\left([\widehat{\alpha} - Q][1-2Q]^{-1}\right) = \mathrm{V}(\widehat{\alpha})[1-2Q]^{-2} \tag{6}$$

Since $\widehat{\alpha}$ follows a binomial distribution with parameter n and α, then

$$\mathrm{V}(\widehat{\alpha}) = \alpha\,(1-\ \alpha) / n \tag{7}$$

Substituting in Eq. (6) by Eq. (7) we can get,

$$V(\widehat{\pi^*}) = \alpha(1-\alpha)[1-2Q\,]^{-2} / n \tag{8}$$

We can use Eq. (3) to calculate $\alpha(1-\alpha)$ as follow,

$$\alpha(1-\alpha) = \pi(1-\pi)[1-2Q\,]^2 + Q[1-Q] \tag{9}$$

then, it is easy to get Eq. (5) by inserting Eq. (9) in Eq. (8). □

To check the validity of Eq. (5), if we replace Q with q_1 , we can get the variance of Warner's estimate as given by Eq. (2).

Theorem 2. *An unbiased estimator of* $V(\widehat{\pi^*})$ *is*

$$\widehat{V}(\widehat{\pi}^*) = \widehat{\alpha}(1-\widehat{\alpha})[1-2Q]^{-2} / (n-1) \tag{10}$$

Proof of Theorem 2. Taking expectation on both sides of Eq. (10), the result is hold. □

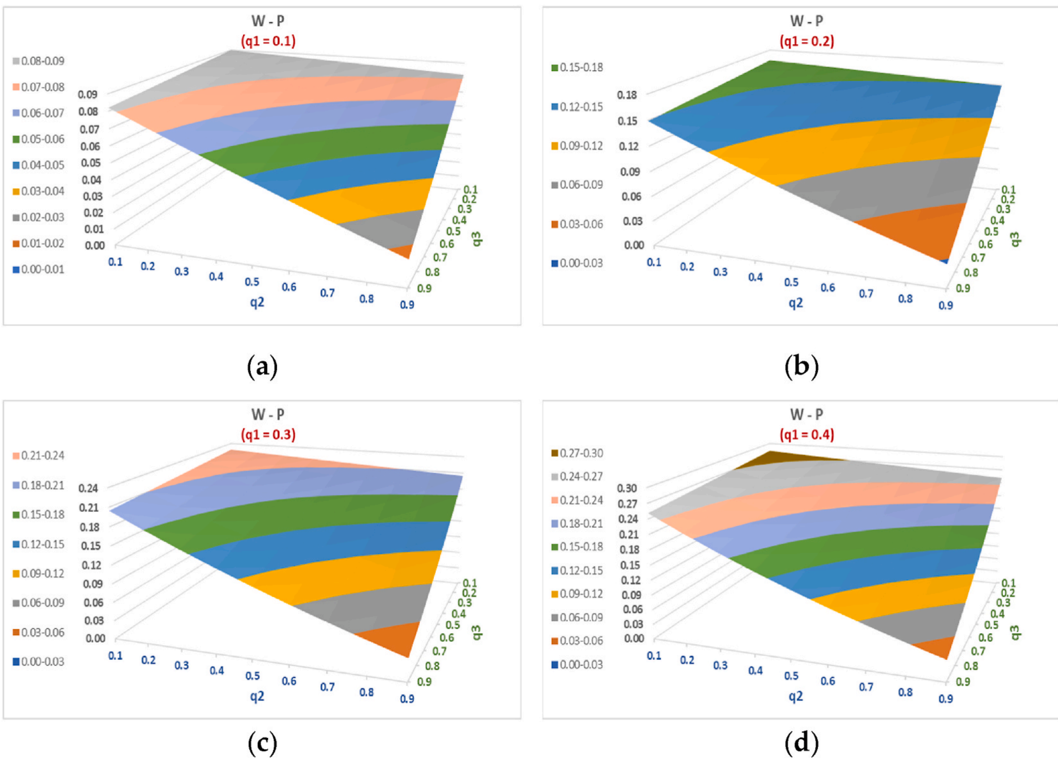

Fig. 2. The difference in efficiency between Warner's model and the proposed model. (a) For $q_1 = 0.1$ (b) For $q_1 = 0.2$ (c) For $q_1 = 0.3$ (d) For $q_1 = 0.4$.

2.3. Efficiency comparison

Here, our focus lies on exploring the specific conditions under which the efficiency of the proposed model, based on a three-stage random tool, outperforms both the original Warner's model [1], which relies on a one-stage random tool, and the well-known Mangat & Singh's model [28], which utilizes a two-stage random tool.

The proposed estimator will be more efficient than the original Warner's estimator iff:

$$q_2 q_3 < [1 - q_1][1 - q_1 q_2 q_3]^{-1} \tag{11}$$

The above inequality (Eq. (11)) shows that the proposed strategy can always be made more efficient than the usual Warner's strategy by choosing any values of $q_2 q_3$ for any suitable practicable value of q_1.

Fig. 2 (a – d) shows the difference, in terms of efficiency, between Warner's model and the proposed model at practicable values of q_1 and the different values of q_2 and q_3. Positive values are in favor of the proposed model.

From Fig. 2, it may be noted that.

1 For all different values of q_2 and q_3, and practicable values of q_1 ($q_1 < 0.5$), the proposed estimate is more efficient than Warner's estimate.
2 For fixed values of q_2 and q_3, the efficiency of the proposed estimate against Warner's estimate increases as q_1 increases from 0.1 to 0.4.
3 For fixed values of q_1 and q_2, the efficiency of the proposed estimate against Warner's estimate increases as q_3 decreases from 0.9 to 0.1 (since the variance of the proposed estimate decreases as q_3 decreases from 0.9 to 0.1, but the variance of Warner's estimate is fixed).
4 For fixed values of q_1 and q_3 the efficiency of the proposed estimate against Warner's estimate increases as q_2 decreases from 0.9 to 0.1 (since the variance of the proposed estimate decreases as q_2 decreases from 0.9 to 0.1, but the variance of Warner's estimate is fixed.

The proposed estimator will be more efficient than the original Mangat & Singh's estimator iff:

$$q_3 < [1 - q_1 q_2][1 - q_1 q_2 q_3]^{-1} \tag{12}$$

The above inequality (Eq. (12)) shows that the proposed strategy can always be made more efficient than the Mangat & Singh's

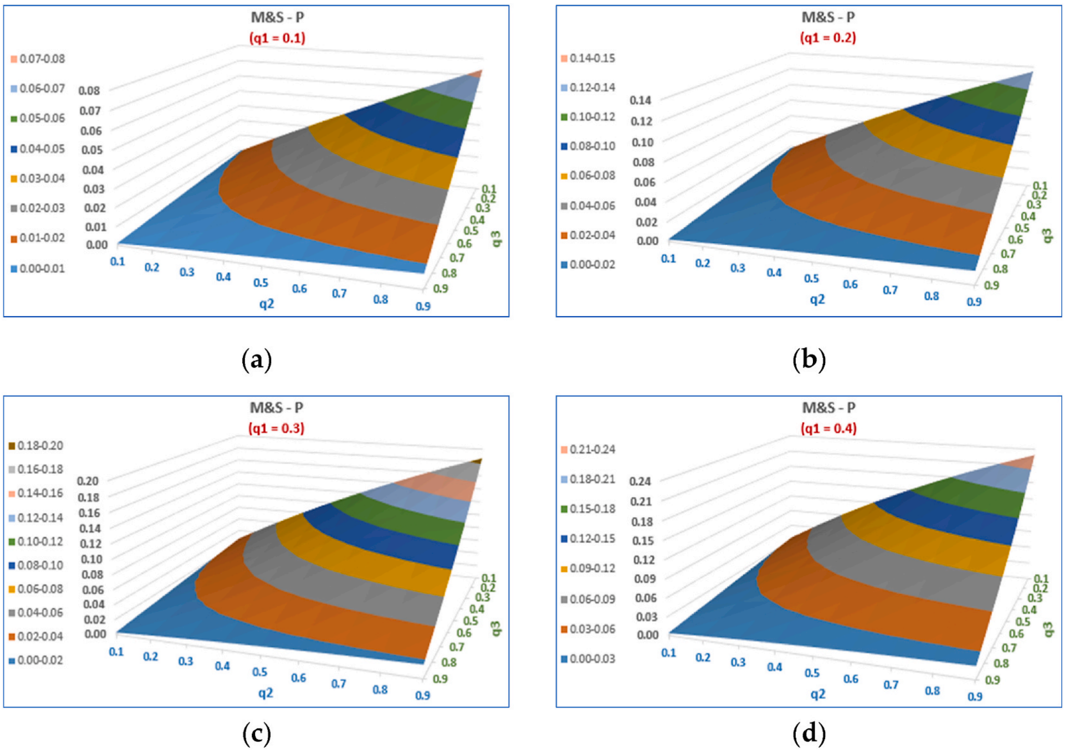

Fig. 3. The difference in efficiency between Mangat & Singh's model and the proposed model. (a) For $q_1 = 0.1$ (b) For $q_1 = 0.2$ (c) For $q_1 = 0.3$ (d) For $q_1 = 0.4$.

strategy by choosing any values of q_3 for any suitable practicable value of q_1 and q_2.

Fig. 3 (a – d) shows the difference, in terms of efficiency, between Mangat & Singh's model and the proposed model at practicable values of q_1 and the different values of q_2 and q_3. Positive values are in favor of the proposed model.

From Fig. 3, it may be noted that.

1 For all different values of q_3, and practicable values of q_1 and q_2 , the proposed estimate is more efficient than Mangat & Singh's estimate.
2 For fixed values of q_2 and q_3, the efficiency of the proposed estimate against Mangat & Singh's estimate increases as q_1 increases from 0.1 to 0.4.
3 For fixed values of q_1 and q_3, the efficiency of the proposed estimate against Mangat & Singh's estimate increases as q_2 increases from 0.1 to 0.4.
4 For fixed values of q_1 and q_2, the efficiency of the proposed estimate against Mangat & Singh's estimate increases as q_3 decreases from 0.9 to 0.1 (since the variance of the proposed estimate decreases as q_3 decreases from 0.9 to 0.1, but the variance of Mangat & Singh's estimate is fixed).

2.4. Measure of privacy protection

One of the basic characteristics of a RR model is protecting the privacy of respondents. Different measures of privacy protection for RR models have been proposed (Anderson [51], Lanke [52], Leysieffer and Warner [53], Zhimin and Zaizai [54]).

Applying the latter, the measure of protection for Warner's model are given as

$$M_W(R) = \frac{(1-2q_1)^2}{2q_1(1-q_1)} \tag{13}$$

and the design probabilities for the proposed RR model can be obtained as:

$P(y|A) = 1 - Q$ and $P(y|\overline{A}) = Q$.

$P(n|A) = Q$ and $P(n|\overline{A}) = 1 - Q$.

and

$$P(A|y) = \frac{\pi}{\pi + (1-\pi)(Q)/(1-Q)}$$

$$P(A|n) = \frac{\pi}{\pi + (1-\pi)(1-Q)/(Q)}$$

And, hence, the measure of privacy protection is obtained as:

$$M_P(R) = \left| 1 - \frac{1}{2}\left\{\tau(y) + \tau(n)\right\} \right|$$

where

$$\tau(y) = \frac{1-Q}{Q}, \tau(n) = \frac{Q}{1-Q}$$

$$M_P(R) = \frac{(1-2\text{Q})^2}{2\text{Q}(1-\text{Q})} \tag{14}$$

To check the validity of Eq. (14), if we replace Q with q_1 , we can get the measure of protection for Warner's model as given by Eq. (13).

The relation between the previous measure of privacy protection and the efficiency of both Warner's model and the proposed model can be presented as follows:

$$V(\widehat{\pi}) = \pi(1-\pi) \Big/ n + \frac{1}{2nM_w(R)} \tag{15}$$

$$V(\widehat{\pi}) = \pi(1-\pi) \Big/ n + \frac{1}{2nM_p(R)} \tag{16}$$

It is obvious from Eq. (15) and Eq. (16) the smaller the values of $M_P(R)$, the less the efficiency of $\widehat{\pi}^*$. On the other hand, Zhimin and Zaizai [54] showed that, the smaller the values of their measure of privacy protection the more the privacy of respondents is being protected. A balancing act is obviously necessary.

3. Empirical study and results

Saudi traffic law, Chapter 2, Article 8 [55], stated that *"C. Each vehicle's owner shall insure his/her vehicle and the regulations determine the rules regulating this."*. A Unified Compulsory Motor Insurance Policy has been issued by Saudi Central Bank, the financial supervisory authority in said Arabia [56]. The policy covers the damage caused by the vehicle to a third party, but it does not cover the damage inflicted on the own vehicle. By law, third party motor insurance is a requirement in cases of car ownership transfer and car periodic technical inspection. At any other time, mostly if the car was involved in an accident or was randomly checked at a traffic point, if there is no valid vehicle insurance, the car owner will be fined a max of around $40 (Saudi traffic law, Violations Table No. (1) [55]), which, in fact, does not guarantee an acceptable level of compliance by car owners. Recently, obligatory (third party) car insurance rates have, substantially, increased within the Saudi insurance market, despite the declining number of car accidents [57]. One of the possible reasons for this is the insufficiency of insurance contributions, an issue that is closely related to our empirical study.

In addition to that not having a third-party liability motor insurance is considered an illegal act, even if the insurance exists, there are other factors which may make a person hesitant to reveal this info when asked; among which are.

- If another person, other than the owner of the vehicle, is using it, his/her name must be stated as a driver in the insurance policy, which is not often the case. This usually happens within families and among friends in the local community.
- The vehicle may be used for a purpose other than that mentioned in the insurance policy. For example, when people use their cars to transport others for a fee, while the purpose of using the car in the insurance policy is private not commercial use.

The proposed model was applied through an experimental study in which the population was final year male undergraduate students within a college of business who 1. own or drive a car (which is very common among university male students) and 2. have knowledge of the basics of risk and insurance. The rationale behind this selection was that the individuals within this 'selected' population are expected to have more awareness of the importance of 'third party car insurance' regardless of the 'assumed' compliance of buying such insurance. Moreover, according to the latest published official statistics by the Saudi Ministry of Health, males within the age group 19–30 are the most vulnerable to injury and death due to road traffic accidents [58]. Fig. 4 (a – d) shows the classification of injuries and deaths of car accidents in Saudi Arabia in 2020 by age (Fig. 4 (a,b)) and gender (Fig. 4 (c,d)).

A random sample of 40 students was selected, then, few days before the experiment, they have been notified and agreed about

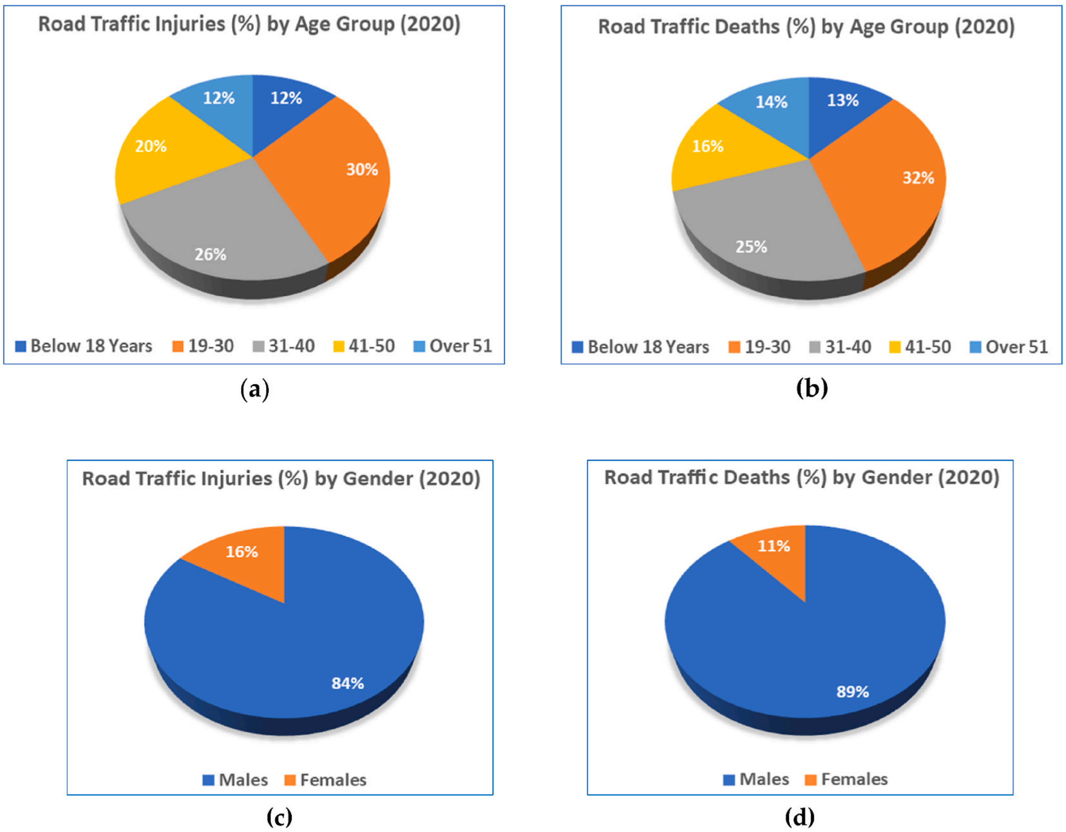

Fig. 4. Traffic injuries and deaths in Saudi Arabia (2020) by age group and gender. (a) Injuries by age (b) Deaths by age (c) Injuries by gender (b) Deaths by gender.

location (in the college), date and time of the experiment. At the beginning of the experiment, a short presentation was delivered explaining the whole process and emphasizing that, by design, their privacies are well-preserved. Each respondent goes through the experiment, behind a partition, without being seen by anyone else in the room and leaves the room when finished.

An empty box, 100 'Yes' cards, 100 'No' cards, and three spinner devices (smart phones with a spinner app (Spin The Wheel) installed) have been used for the purpose. The spinner app in the first two devices was set, with $q_1 = q_2 = 0.4$), to show one of two options:

- Do you comply with the third-party insurance requirements and terms on your car at all time?;
- Go to the next device.

So, if option 1 appeared at any of the first two devices, the experiment ends, and the respondent drops a Yes or No card in the box. If the experiment extends to the third spinner device, the respondent will have to answer one of the following two questions, each having an equal chance to show up ($q_3 = 0.5$):

- Do you comply with the third-party insurance requirements and terms on your car at all time?;
- Do you NOT comply with third-party insurance requirements and terms on your car at all time?

Based upon the results obtained from the sample and applying Eq. (4) the estimate of the proportion of individuals, within the selected population, who are not complying with buying obligatory third-party car insurance, $\widehat{\pi^*}$, is 0.1875 and its estimated variance $\widehat{V}(\widehat{\pi^*})$ is 0.007512 (Eq. (10)). Thus, a 95% confidence interval for $\widehat{\pi^*}$ is (0.018, 0.357).

4. Discussion

The suggested RR model provides an efficient alternative to both Warner's model and Mangat & Singh's model that allows more credibility from a practical perspective. Setting the values for the probabilities q_1, q_2, q_3 to $0.4, 0.4, 0.5$, respectively, seem to be a rational choice in the sense that, both the efficiency and the privacy for the proposed model are equal to those of Warner's model at $p_1 = 0.9$ and those of Mangat & Singh's model at $p_1 = 0.8$. Furthermore, this selection maintains a good balance between increasing the likelihood of sensitive question appearance and reducing the level of suspicion by respondents, and, hence, raising their cooperation. Fig. 5 shows that the efficiency of the proposed model is better than that of both Warner's and Mangat & Singh's models at $q_2 = q_3 = 0.5$ for all values of q_1. Also, at $q_2 = q_3 = 0.5$ and $q_1 = 0.4$ the efficiency of the proposed model corresponds to that of Warner's model at $p_1 = 0.9$ and to that of Mangat & Singh's model at $p_1 = 0.8$.

As per the specific application of the suggested model, for estimating obligatory motor insurance noncompliance ratio, the resulting estimate (0.1875) may be looked at as a close-to-minimum value for the whole car owners' population in Saudi Arabia. This is because the selected sub-population on which the RR model was applied represents those car owners with 'high' awareness of insurance. Furthermore, this estimate can be considered a basis to predict the level of inclusion for this type of insurance which is a matter of importance in the context of both social and financial sustainability. Also, it would be beneficial for insurance companies to estimate this ratio within different areas or communities from time to time to adjust insurance rates and terms accordingly. Since gross premiums in insurance include a loadings component, this ratio can be taken into consideration to adjust this component if necessary.

In general, RR models, including the one presented in this article, can be an efficient data collection tool to investigate sensitive attributes in various areas [59].

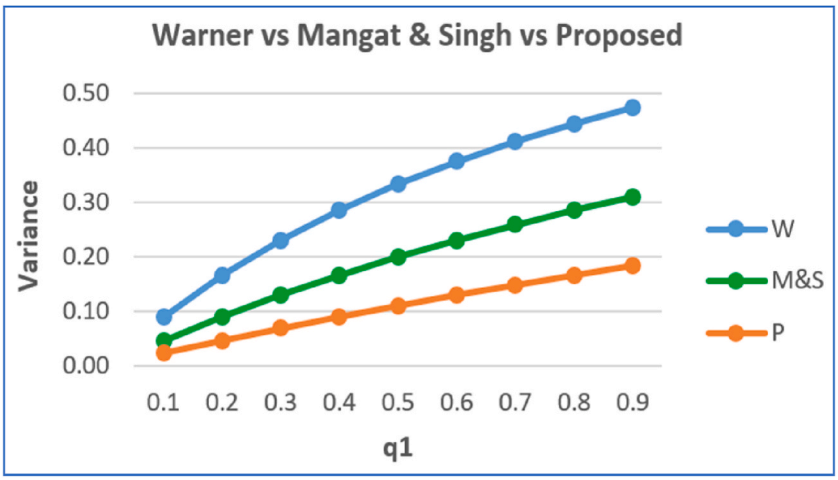

Fig. 5. Efficiency Comparison between Warner's model, Mangat & Singh model and the proposed model at selected values for q_1, q_2, q_3.

Funding

This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University (KFU), Saudi Arabia (GRANT5951).

Ethics statement

The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the deanship of the scientific research ethical committee, King Faisal University (date of approval: 23 February 2022).

Data availability statement

The authors declare that the data supporting the findings of this study are available within the article. The raw/derived data supporting the findings of this study are available from the corresponding author at request and with the permission of all authors.

CRediT authorship contribution statement

Ahmad M. Aboalkhair: Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **A.M. Elshehawey:** Writing – review & editing, Writing – original draft, Validation, Methodology, Formal analysis. **Mohammad A. Zayed:** Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors gratefully acknowledge KFU for technical and financial support.

References

[1] S.L. Warner, Randomized response: a survey technique for eliminating evasive answer bias, J. Am. Stat. Assoc. 60 (1965) 63–69, https://doi.org/10.2307/2283137.
[2] A. Abul-Ela, H. Dakrouri, Randomized response model: a ratio estimator, in: Proceedings of the Survey Research Methods Section, American Statistical Association, USA, 1980.
[3] N.J. Scheers, C.M. Dayton, Covariate randomized response models, J. Am. Stat. Assoc. 83 (1988) 969–974, https://doi.org/10.2307/2290121.
[4] G.S. Maddala, Limited-Dependent and Qualitative Variables in Econometrics, Econometric Society Monographs; Cambridge University Press, Cambridge, 1983, 978-0-521-33825-7.
[5] S.H. Hsieh, S.M. Lee, P.S. Shen, Semiparametric analysis of randomized response data with missing covariates in logistic regression, Comput. Stat. Data Anal. 53 (2009) 2673–2692, https://doi.org/10.1016/j.csda.2009.01.011.
[6] S.H. Hsieh, S.M. Lee, P.S. Shen, Logistic regression analysis of randomized response data with missing covariates, J. Stat. Plann. Inference 140 (2010) 927–940, https://doi.org/10.1016/j.jspi.2009.09.020.
[7] P.-C. Chang, K.-H. Pho, S.-M. Lee, C.-S. Li, Estimation of parameters of logistic regression for two-stage randomized response technique, Comput. Stat. 36 (2021) 2111–2133, https://doi.org/10.1007/s00180-021-01068-5.
[8] S.-M. Lee, P.-L. Tran, T.-N. Le, C.-S. Li, Prediction of a sensitive feature under indirect questioning via Warner's randomized response technique and latent class model, Mathematics 11 (2023) 345, https://doi.org/10.3390/math11020345.
[9] R.L. Winkler, L.A. Franklin, Warner's randomized response model: a bayesian approach, J. Am. Stat. Assoc. 74 (1979) 207–214, https://doi.org/10.2307/2286752.
[10] H.S. Migon, V.M. Tachibana, Bayesian approximations in randomized response model, Comput. Stat. Data Anal. 24 (1997) 401–409, https://doi.org/10.1016/S0167-9473(96)00075-8.
[11] N.K. Unnikrishnan, S. Kunte, Bayesian analysis for randomized response models, Sankhya: The Indian Journal of Statistics, Series B 61 (1999) 422–432 (1960-2002).
[12] S.K. Bar-Lev, E. Bobovich, B. Boukai, A common conjugate prior structure for several randomized response models, Test 12 (2003) 101–113, https://doi.org/10.1007/BF02595813.
[13] L. Barabesi, M. Marcheselli, Bayesian estimation of proportion and sensitivity level in randomized response procedures, Metrika 72 (2010) 75–88, https://doi.org/10.1007/s00184-009-0242-7.
[14] Z. Hussain, J. Shabbir, M. Riaz, Bayesian estimation using Warner's randomized response model through simple and mixture prior distributions, Commun. Stat. Simulat. Comput. 40 (2010) 147–164, https://doi.org/10.1080/03610918.2010.532897.
[15] H. Xin, J. Zhu, T.-R. Tsai, C.-Y. Hung, Hierarchical bayesian modeling and randomized response method for inferring the sensitive-nature proportion, Mathematics 9 (2021) 2518, https://doi.org/10.3390/math9192518.
[16] H.H. Ki, K.Y. Jun, Y.L. Hwa, A stratified randomized response technique, Korean J. Appl. statistics 7 (1994) 141–147.
[17] J.-M. Kim, W.D. Warde, A stratified Warner's randomized response model, J. Stat. Plann. Inference 120 (2004) 155–165, https://doi.org/10.1016/S0378-3758(02)00500-1.
[18] J.-M. Kim, M.E. Elam, A two-stage stratified Warner's randomized response model using optimal allocation, Metrika 61 (2005) 1–7, https://doi.org/10.1007/s001840400319.
[19] S. Ghufran, S. Khowaja, M.J. Ahsan, Compromise allocation in multivariate stratified sample surveys under two stage randomized response model, Optim Lett 8 (2014) 343–357, https://doi.org/10.1007/s11590-012-0581-6.

[20] H.P. Singh, T.A. Tarray, A stratified tracy and osahan's two-stage randomized response model, Commun. Stat. Theor. Methods 45 (2016) 3126–3137, https://doi.org/10.1080/03610926.2014.895839.
[21] S. Abdelfatah, R. Mazloum, An efficient two-stage randomized response model under stratified random sampling, Math. Popul. Stud. 23 (2016) 222–238, https://doi.org/10.1080/08898480.2016.1222222.
[22] T. Tarray, H. Singh, A proficient two-stage stratified randomized response strategy, J. Mod. Appl. Stat. Methods 17 (2018), https://doi.org/10.22237/jmasm/1544453468.
[23] Z. Hussain, S.A. Cheema, I. Hussain, A stratified randomized response model for sensitive characteristics using non identical trials, Commun. Stat. Theor. Methods 49 (2020) 99–115, https://doi.org/10.1080/03610926.2018.1530791.
[24] N. Gupta, S. Gupta, Mohd Tanwir Akhtar, Multi-choice stratified randomized response model with two-stage classification, J. Stat. Comput. Simulat. 92 (2022) 895–910, https://doi.org/10.1080/00949655.2021.1980880.
[25] B.G. Greenberg, A.-L.A. Abul-Ela, W.R. Simmons, D.G. Horvitz, The unrelated question randomized response model: theoretical framework, J. Am. Stat. Assoc. 64 (1969) 520–539, https://doi.org/10.2307/2283636.
[26] J.J.A. Moors, Optimization of the unrelated question randomized response model, J. Am. Stat. Assoc. 66 (1971) 627–629, https://doi.org/10.2307/2283543.
[27] D. Raghavarao, On an estimation problem in Warner's randomized response technique, Biometrics 34 (1978) 87–90, https://doi.org/10.2307/2529591.
[28] N.S. Mangat, R. Singh, An alternative randomized response procedure, Biometrika 77 (1990) 439–442, https://doi.org/10.2307/2336829.
[29] A.Y.C. Kuk, Asking sensitive questions indirectly, Biometrika 77 (1990) 436–438, https://doi.org/10.2307/2336828.
[30] N.S. Mangat, An improved randomized response strategy, J. Roy. Stat. Soc. B 56 (1994) 93–95.
[31] S. Singh, R. Singh, N.S. Mangat, Some alternative strategies to Moors' model in randomized response sampling, J. Stat. Plann. Inference 83 (2000) 243–255, https://doi.org/10.1016/S0378-3758(99)00092-0.
[32] M. Bhargava, R. Singh, A modified randomization device for Warner's model, Statistica 60 (2000) 315–322, https://doi.org/10.6092/issn.1973-2201/6941.
[33] S. Singh, S. Horn, R. Singh, N.S. Mangat, On the use of modified randomization device for estimating the prevalence of a sensitive attribute, Statistics in Transition 6 (2003) 515–522.
[34] K.-C. Huang, A survey technique for estimating the proportion and sensitivity in a dichotomous finite population, Stat. Neerl. 58 (2004) 75–82, https://doi.org/10.1046/j.0039-0402.2003.00113.x.
[35] H.-J. Chang, C.-L. Wang, K.-C. Huang, On estimating the proportion of a qualitative sensitive character using randomized response sampling, Qual Quant 38 (2005) 675–680, https://doi.org/10.1007/s11135-005-8105-4.
[36] S. Gupta, J. Shabbir, R. Iembo, Modifications to Warner's model using blank cards, Am. J. Math. Manag. Sci. 26 (2006) 185–196, https://doi.org/10.1080/01966324.2006.10737666.
[37] C.R. Gjestvang, S. Singh, A new randomized response model, J. Roy. Stat. Soc. B 68 (2006) 523–530.
[38] P.F. Perri, Modified randomized devices for simmons' model, Model Assisted Statistics Appl. 3 (2008) 233–239, https://doi.org/10.3233/MAS-2008-3307.
[39] S. Abdelfatah, R. Mazloum, Efficient estimation in a two-stage randomized response model, Math. Popul. Stud. 22 (2015) 234–251, https://doi.org/10.1080/08898480.2014.953897.
[40] F. Batool, J. Shabbir, Z. Hussain, An improved binary randomized response model using six decks of cards, Commun. Stat. Simulat. Comput. 46 (2017) 2548–2562, https://doi.org/10.1080/03610918.2015.1053922.
[41] H.P. Singh, S.M. Gorey, A new efficient unrelated randomized response model, Commun. Stat. Theor. Methods 46 (2017) 12059–12074, https://doi.org/10.1080/03610926.2017.1291972.
[42] T.A. Tarray, H.P. Singh, A randomization device for estimating a rare sensitive attribute in stratified sampling using Poisson distribution, Afr. Mat. 29 (2018) 407–423, https://doi.org/10.1007/s13370-018-0550-z.
[43] G. Narjis, J. Shabbir, Estimation of population proportion and sensitivity level using optional unrelated question randomized response techniques, Commun. Stat. Simulat. Comput. 49 (2020) 3212–3226, https://doi.org/10.1080/03610918.2018.1538453.
[44] G.N. Singh, S. Suman, A modified two-stage randomized response model for estimating the proportion of stigmatized attribute, J. Appl. Stat. 46 (2019) 958–978, https://doi.org/10.1080/02664763.2018.1529150.
[45] C. Singh, G.N. Singh, J.-M. Kim, A randomized response model for sensitive attribute with privacy measure using Poisson distribution, Ain Shams Eng. J. 12 (2021) 4051–4061, https://doi.org/10.1016/j.asej.2021.03.006.
[46] S.O. Gönülal, Motor Third-Party Liability Insurance in Developing Countries: Raising Awareness and Improving Safety, World Bank, Washington, DC, 2009.
[47] J. Namukasa, J. Ssekakubo, B. Bagenda, Assessment of the level of awareness and perception of motor third party insurance in kampala, Uganda, The Journal of Risk Management and Insurance 21 (2017).
[48] A. Jedrzychowska, E. Poprawska, Is the minimum amount of coverage from motor third party liability insurance enough to cover personal injuries? Int. J. Risk Assess. Manag. 19 (2016) 125–141, https://doi.org/10.1504/IJRAM.2016.074439.
[49] C. Van Schoubroek, European Harmonised Rules on Motor Vehicles Liability Insurance Reviewed (2022) 133–144.
[50] B. Tomeski, Development of motor third party liability insurance market in terms of changing regulation, Procedia - Social and Behavioral Sciences 44 (2012) 200–209, https://doi.org/10.1016/j.sbspro.2012.05.021.
[51] H. Anderson, Estimation of a proportion through randomized response, Int. Stat. Rev./Rev. Int. Stat. 44 (1976) 213–217, https://doi.org/10.2307/1403279.
[52] J. Lanke, On the degree of protection in randomized interviews, Int. Stat. Rev./Rev. Int. Stat. 44 (1976) 197–203, https://doi.org/10.2307/1403277.
[53] F.W. Leysieffer, S.L. Warner, Respondent jeopardy and optimal designs in randomized response models, J. Am. Stat. Assoc. 71 (1976) 649–656, https://doi.org/10.2307/2285595.
[54] H. Zhimin, Y. Zaizai, Measure of privacy in randomized response model, Qual Quant 46 (2012) 1167–1180, https://doi.org/10.1007/s11135-012-9698-z.
[55] Saudi. Ministry of Interior, The Executive Regulations of the Traffic Law, Available online: https://www.moi.gov.sa/wps/wcm/connect/5c50d3bd-dbb8-49a4-b3482ec4445ed883/The+Executive+Regulations+of+the+Traffic+Law.doc?MOD=AJPERES&CVID=nJk8Zqm (accessed on 15 February 2023)..
[56] Saudi Central Bank, Insurance Rules and Regulation, Available online: https://www.sama.gov.sa/enUS/Laws/InsuranceRulesAndRegulations/The_Unified_Compulsory_Motor_Insurance_Policy.pdf (accessed on 15 February 2023)..
[57] Traffic Accidents Are Declining and Insurance Prices Are Rising in Saudi Arabia. Available online: https://www.independentarabia.com/node/4... (accessed on 15 February 2023).
[58] Saudi Ministry of Health, Road Traffic Injuries and Deaths, Available online: https://www.moh.gov.sa/en/Ministry/Statistics/Pages/Traffic-accidents.aspx (accessed on 15 February 2023)..
[59] Ayat Da'seh, Mohammad Rababa, Military nurses perspectives towards research utilization barriers, Heliyon 7 (2021) e08181, https://doi.org/10.1016/j.heliyon.2021.e08181, 1167–1180.