Journal of Statistics Applications & Probability Letters
An International Journal

http://dx.doi.org/10.18576/jsapl/100103

Correction: Dichotomous Randomised Response Techniques

A. T. Adeniran * *and O. M. Akojenu*

Department of Statistics, University of Ibadan, Ibadan, Oyo State, Nigeria

Received: 24 Aug. 2022, Revised: 10 Oct. 2022, Accepted: 15 Oct. 2022
Published online: 1 Jan. 2023

Abstract: This study points out some mistakes in "Dichotomous Randomised Response Techniques", introduced by Hussain and Shabbir to estimate proportion of people with a sensitive trait. The same study was reparameterized, re-estimated, extended and cited by various researchers, though in literature no one has pointed out the mistakes. The study adopted same Binomial distribution as sampling model, Maximum Likelihood as estimation procedure, and Percentage Relative Efficiency (PRE) evaluated at varying values of design parameters as model evaluation criterion. In most cases, the PRE of the corrected-dichotomous Randomised Response Distributions (RRD) over the orthodox models are greater than 100, indicating its efficiency.

Keywords: Binomial distribution, maximum likelihood, proportion, relative efficiency, sensitive trait

1 Introduction

Obtaining truthful responses is challenging in all types of surveys, particularly, when sensitive subject matters are being investigated. The Randomised Response Technique (RRT) also known as Randomised Response Model (RRM) launched by [1] is a survey method specifically developed to enhance precision of answers to sensitive questions (induced-abortion, masturbation, rituals killing, occult-affiliation, kidnapping, banditry, cyber crime, raping, sexual assault etc) with the focus of deducing the proportion of subjects carrying the sensitive trait(s).

Warner technique require respondent to select (on random basis) and answer a sensitive question with probability p and its negation with probability $(1-p)$. Using the classical estimation procedure, Warner estimate of proportion of subject characterising the sensitive attribute is

$$\hat{\pi}_w = \frac{\frac{n'}{n}-(1-p)}{2p-1} = \frac{\hat{\phi}-(1-p)}{2p-1}; \quad \text{for } p \neq \tfrac{1}{2} \tag{1}$$

$\hat{\phi} = \frac{n'}{n}$ with variance as

$$Var(\hat{\pi}_w) = \frac{\pi(1-\pi)}{n} + \frac{p(1-p)}{n(2p-1)^2}. \tag{2}$$

where,
$n =$ number of observed respondents (sample size),
$n' =$ observed number of yes-answers,
$p =$ probability of selecting (and answering) sensitive question, and
$\pi =$ proportion of yes answers to the sensitive question.

[2] postulated an alternative RRM that requires respondent to answer, on the first stage, the sensitive question without randomiser. All respondents with no response then use probability device on the second stage with Warner statements. The Mangat estimate of proportion of people in stigmatizing class is

$$\hat{\pi}_m = \frac{\frac{n'}{n}-1+p}{p} = \frac{\hat{\theta}-1+p}{p}; \quad \text{where } \hat{\theta} = \frac{n'}{n} \tag{3}$$

* Corresponding author e-mail: at.adeniran@ui.edu.ng

with sample variance

$$Var(\hat{\pi}_m) = \frac{\pi(1-\pi)}{n} + \frac{(1-p)(1-\pi)}{np}. \tag{4}$$

π and n maintain their usual meaning. Mangat RRT is very efficient but less valid because of contamination of yes-answer [3].

[4] convoluted-Warner RRMs consists of random use of either of the two randomization devices R_1 and R_2 with probabilities $q = \frac{\alpha}{\alpha+\beta}$ and $1-q = \frac{\beta}{\alpha+\beta}$, respectively, where $\alpha, \beta \in \mathbb{Z}^+$. The randomisation devices R_1 and R_2 are same to the Warner's device but with different probability (p_i, $i = 1, 2$) of selecting the sensitive question. Then for the ith respondent among n simple random sample selected with replacement, the probability of a yes response is given by

$$\phi = \frac{\alpha}{\alpha+\beta}[p_1\pi + (1-p_1)(1-\pi)] + \frac{\beta}{\alpha+\beta}[p_2\pi + (1-p_2)(1-\pi)] \tag{5}$$

and the probability of no response as

$$1-\phi = \frac{\alpha}{\alpha+\beta}p_1(1-\pi) + \frac{\beta}{\alpha+\beta}p_2(1-\pi) \tag{6}$$

[4] later set $p_1 = 1-p_2$ in (5) to have

$$\phi = \frac{\pi[(2p_1-1)(\alpha-\beta)] + p_1\beta + p_2\alpha}{\alpha+\beta} \tag{7}$$

which consequently leads to dichotomous-Warner estimator of π as

$$\hat{\pi}_{dw} = \frac{\hat{\phi}(\alpha+\beta) - p_1\beta - p_2\alpha}{(2p_1-1)(\alpha-\beta)} \tag{8}$$

with variance

$$Var(\hat{\pi}_{dw}) = \frac{\pi(1-\pi)}{n} + \frac{(p_2\alpha + p_1\beta)(p_1\alpha + p_2\beta)}{n(2p_1-1)^2(\alpha-\beta)^2(\alpha+\beta)} \tag{9}$$

[5] and [6] noticed an error in (9), re-estimated it to be

$$Var(\hat{\pi}_{edw}) = \frac{\pi(1-\pi)}{n} + \frac{(p_2\alpha + p_1\beta)(p_1\alpha + p_2\beta)}{n(2p_1-1)^2(\alpha-\beta)^2} \tag{10}$$

and extended Hussain and Shabbir work to Mangat RRM with probability of yes and no-answer's as given in (11) and (12) below:

$$\theta = \pi + \frac{\alpha}{\alpha+\beta}(1-p_1)(1-\pi) + \frac{\beta}{\alpha+\beta}(1-p_2)(1-\pi) \tag{11}$$

and

$$1-\theta = \frac{\alpha}{\alpha+\beta}p_1(1-\pi) + \frac{\beta}{\alpha+\beta}p_2(1-\pi). \tag{12}$$

Also, in a similar manner, they obtained

$$\pi_{dm} = \frac{\hat{\theta}(\alpha+\beta) - p_2\alpha - p_1\beta}{p_1\alpha + p_2\beta} \tag{13}$$

and

$$Var(\pi_{dm}) = \frac{\pi(1-\pi)}{n} + \frac{(1-\pi)(p_2\alpha + p_1\beta)}{n(p_1\alpha + p_2\beta)^2} \tag{14}$$

by setting $p_1 = 1-p_2$ in (11).

The following errors were identified in [4] and [6] dichotomous RRMs.

(a) p denotes probability of selecting and answering sensitive question, it is a pre-determined, pre-assigned or preset value, set by the researcher before field work commenced. Therefore, altering p by setting $p_1 = 1-p_2$ after data collection or during estimation procedure is not a good practice.

(b)[4] set $p_1 = 1 - p_2$ in (5) to have (7). Without much mathematics, this result is wrong. The answer should be in terms of p_2 alone.

(c)In addition, the two RR designs ([4], [6]) are convolution of two Warner and two Mangat RRMs, respectively, such that whenever $p_1 = p_2 = p$, expected result should be conventional-Warner and Mangat, respectively. Evidence from (15-18), putting $p_1 = p_2 = p$ in (8-10, 13-14) doesn't produce conventional estimators as

$$\hat{\pi}_{dw} = \frac{\hat{\phi}(\alpha+\beta) - p\beta - p\alpha}{(2p-1)(\alpha-\beta)} = \frac{\hat{\phi}(\alpha+\beta) - p(\alpha+\beta)}{(2p-1)(\alpha-\beta)} = \frac{(\alpha+\beta)(\hat{\phi}-p)}{(2p-1)(\alpha-\beta)} \neq \hat{\pi}_w, \tag{15}$$

$$Var(\hat{\pi}_{dw}) = \frac{\pi(1-\pi)}{n} + \frac{(p\alpha+p\beta)(p\alpha+p\beta)}{n(2p-1)^2(\alpha-\beta)^2(\alpha+\beta)} = \frac{\pi(1-\pi)}{n} + \frac{p^2(\alpha+\beta)}{n(2p-1)^2(\alpha-\beta)^2} \neq Var(\hat{\pi}_w), \tag{16}$$

$$\hat{\pi}_{dm} = \frac{\hat{\theta}(\alpha+\beta) - p\alpha - p\beta}{p\alpha+p\beta} = \frac{\hat{\theta}(\alpha+\beta) - p(\alpha+\beta)}{p(\alpha+\beta)} = \frac{\hat{\theta}-p}{p} \neq \hat{\pi}_m, \tag{17}$$

and

$$Var(\hat{\pi}_{dm}) = \frac{\pi(1-\pi)}{n} + \frac{(1-\pi)(p\alpha+p\beta)}{n(p\alpha+p\beta)^2} = \frac{\pi(1-\pi)}{n} + \frac{(1-\pi)}{np(\alpha+\beta)} \neq Var(\hat{\pi}_m). \tag{18}$$

(d)Mathematically, putting $p_1 = 1 - p_2$ in (5), the reduced-equation should correctly be of the form

$$\phi = \frac{\pi(2p_2-1)(\beta-\alpha) + p_2\alpha + (1-p_2)\beta}{\alpha+\beta} \tag{19}$$

with corresponding estimator as

$$\pi = \frac{\phi(\alpha+\beta) - p_2\alpha - (1-p_2)\beta}{(2p_2-1)(\beta-\alpha)}, \tag{20}$$

which also in turn doesn't produce (1) on setting $p_1 = p_2 = p$. Either (8), (13) or (20) cannot be established for unbiasedness, yet, [4], [6] concluded without proof that their respective estimator is unbiased.

(e)The authors asserted that the reason for altering randomization parameter is to make their proposed estimator unbiased. Despite, the objective wasn't achieved. In fact, it is essential to understand that an estimator is biased does not depreciate it. Note that sample variance, $s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right]$, is unbiased and the population variance, $\sigma^2 = \frac{1}{n}\left[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right]$, is biased. Still, $Var(\sigma^2) < Var(s^2)$, meaning that a biased estimator can produce smaller deviation than an unbiased estimator. Though, one that is unbiased and has the smallest variance is the best.

The dichotomous RRM innovated by [4] is a powerful effective design that has been modified, extended and cited by various researchers. Unfortunately, their estimators and accompany properties (unbiasedness and variances) were wrongly derived. Thus, the purpose of this study is to obtain the correct estimators of dichotomous Warner and Mangat RRTs developed by [4] and [6], respectively.

2 Theoretical Justification

2.1 Correction of Hussain and Shabbir Dichotomous Warner RRM

Equations (5) and (6) can be equivalently presented as

$$\phi = \left[\frac{\alpha(2p_1-1) + \beta(2p_2-1)}{\alpha+\beta}\right]\pi + \left[\frac{\alpha(1-p_1) + \beta(1-p_2)}{\alpha+\beta}\right] \tag{21}$$

and

$$1-\phi = \frac{\alpha\, p_1 + \beta\, p_2}{\alpha+\beta} - \left[\frac{\alpha(2p_1-1) + \beta(2p_2-1)}{\alpha+\beta}\right]\pi. \tag{22}$$

That is,

$$X_i = \begin{cases} 1; & \text{if } i^{th} \text{ say yes with probability } \left[\frac{\alpha(2p_1-1)+\beta(2p_2-1)}{\alpha+\beta}\right]\pi + \left[\frac{\alpha(1-p_1)+\beta(1-p_2)}{\alpha+\beta}\right] \\ 0; & \text{if } i^{th} \text{ respondent say no with probability } \frac{\alpha p_1+\beta p_2}{\alpha+\beta} - \left[\frac{\alpha(2p_1-1)+\beta(2p_2-1)}{\alpha+\beta}\right]\pi. \end{cases} \tag{23}$$

Expression for the expectation of X is

$$E(X) = \sum_i X_i P(X_i) = \left[\frac{\alpha(2p_1-1)+\beta(2p_2-1)}{\alpha+\beta}\right]\pi + \left[\frac{\alpha(1-p_1)+\beta(1-p_2)}{\alpha+\beta}\right]. \tag{24}$$

Also,

$$E(X^2) = \sum_i X_i^2 P(X_i) = \left[\frac{\alpha(2p_1-1)+\beta(2p_2-1)}{\alpha+\beta}\right]\pi + \left[\frac{\alpha(1-p_1)+\beta(1-p_2)}{\alpha+\beta}\right]. \tag{25}$$

Hence,

$$\begin{aligned} Var(X) = & \left[\left\{\frac{\alpha(2p_1-1)+\beta(2p_2-1)}{\alpha+\beta}\right\}\pi + \left\{\frac{\alpha(1-p_1)+\beta(1-p_2)}{\alpha+\beta}\right\}\right] \\ & - \left[\left\{\frac{\alpha(2p_1-1)+\beta(2p_2-1)}{\alpha+\beta}\right\}\pi + \left\{\frac{\alpha(1-p_1)+\beta(1-p_2)}{\alpha+\beta}\right\}\right]^2 \\ & = \left[\left\{\frac{\alpha(2p_1-1)+\beta(2p_2-1)}{\alpha+\beta}\right\}\pi + \left\{\frac{\alpha(1-p_1)+\beta(1-p_2)}{\alpha+\beta}\right\}\right] \times \\ & \left[1 - \left(\left\{\frac{\alpha(2p_1-1)+\beta(2p_2-1)}{\alpha+\beta}\right\}\pi + \left\{\frac{\alpha(1-p_1)+\beta(1-p_2)}{\alpha+\beta}\right\}\right)\right] \end{aligned}$$

which finally produce

$$Var(X) = \frac{[\alpha(2p_1-1)+\beta(2p_2-1)]^2\pi(1-\pi) + (p_1\alpha+p_2\beta)[(\alpha+\beta-p_1\alpha-p_2\beta)]}{(\alpha+\beta)^2}. \tag{26}$$

In the design, outcome of one respondent has no influence on the other, only two possible response; ”yes” or ”no” with constant probability of each as presented in (21) and (22), respectively. It is an immediate consequence of the above conditions that the likelihood function $\prod_{i=1}^{n} f(y;\pi)$ be defined as:

$$\begin{aligned} L = \prod_{i=1}^{n} \binom{n}{n'} & \left[\left\{\frac{\alpha(2p_1-1)+\beta(2p_2-1)}{\alpha+\beta}\right\}\pi + \left\{\frac{\alpha(1-p_1)+\beta(1-p_2)}{\alpha+\beta}\right\}\right]^{n'} \times \\ & \left[\frac{\alpha p_1+\beta p_2}{\alpha+\beta} - \left\{\frac{\alpha(2p_1-1)+\beta(2p_2-1)}{\alpha+\beta}\right\}\pi\right]^{n-n'} \end{aligned} \tag{27}$$

Setting derivative of the natural logarithm of (27) to zero gives

$$\frac{n'\left\{\frac{\alpha(2p_1-1)+\beta(2p_2-1)}{\alpha+\beta}\right\}}{\left\{\frac{\alpha(2p_1-1)+\beta(2p_2-1)}{\alpha+\beta}\right\}\pi + \left\{\frac{\alpha(1-p_1)+\beta(1-p_2)}{\alpha+\beta}\right\}} = \frac{(n-n')\left\{\frac{\alpha(2p_1-1)+\beta(2p_2-1)}{\alpha+\beta}\right\}}{\frac{\alpha p_1+\beta p_2}{\alpha+\beta} - \left\{\frac{\alpha(2p_1-1)+\beta(2p_2-1)}{\alpha+\beta}\right\}\pi}. \tag{28}$$

Succinctly, (28) can be presented as

$$\frac{n'A}{A\pi+B} = \frac{(n-n')A}{C-A\pi} \tag{29}$$

where, $A = \frac{\alpha(2p_1-1)+\beta(2p_2-1)}{\alpha+\beta}$, $B = \frac{\alpha(1-p_1)+\beta(1-p_2)}{\alpha+\beta}$ and $C = \frac{\alpha p_1+\beta p_2}{\alpha+\beta}$. From (29),

$$\pi = \frac{n'AC-(n-n')AB}{nA^2} = \frac{\frac{n'}{n}AC - \left(1-\frac{n'}{n}\right)AB}{A^2}. \tag{30}$$

Explicitly,

$$\pi = \frac{\frac{n'}{n}\left[\left\{\frac{\alpha(2p_1-1)+\beta(2p_2-1)}{\alpha+\beta}\right\}\left\{\frac{\alpha p_1+\beta p_2}{\alpha+\beta}\right\}\right] - \left(1-\frac{n'}{n}\right)\left[\frac{\alpha(2p_1-1)+\beta(2p_2-1)}{\alpha+\beta}\right]\left[\frac{\alpha(1-p_1)+\beta(1-p_2)}{\alpha+\beta}\right]}{\left[\frac{\alpha(2p_1-1)+\beta(2p_2-1)}{\alpha+\beta}\right]^2}$$

$$= \frac{\frac{n'}{n}\left[\frac{\alpha p_1+\beta p_2}{\alpha+\beta}\right] - \left(1-\frac{n'}{n}\right)\left[\frac{\alpha(1-p_1)+\beta(1-p_2)}{\alpha+\beta}\right]}{\frac{\alpha(2p_1-1)+\beta(2p_2-1)}{\alpha+\beta}}$$

$$= \frac{\frac{n'}{n}[\alpha p_1+\beta p_2] + \frac{n'}{n}[\alpha(1-p_1)+\beta(1-p_2)] - [\alpha(1-p_1)+\beta(1-p_2)]}{\alpha(2p_1-1)+\beta(2p_2-1)}$$

$$= \frac{\frac{n'}{n}[\alpha p_1+\beta p_2+\alpha-\alpha p_1+\beta-\beta p_2] - \alpha-\beta+\alpha p_1+\beta p_2}{\alpha(2p_1-1)+\beta(2p_2-1)}$$

which finally gives the corrected-dichotomous Warner of [4] as

$$\hat{\pi}_{cdw} = \frac{\frac{n'}{n}(\alpha+\beta) - [(\alpha+\beta)-(\alpha p_1+\beta p_2)]}{\alpha(2p_1-1)+\beta(2p_2-1)}. \tag{31}$$

The estimator in (31) is unbiased since

$$E(\hat{\pi}_{cdw}) = \frac{1}{\alpha(2p_1-1)+\beta(2p_2-1)}\left[\frac{(\alpha+\beta)}{n}E(n') - [(\alpha+\beta)-(\alpha p_1+\beta p_2)]\right]$$

$$= \frac{1}{\alpha(2p_1-1)+\beta(2p_2-1)}\left[\frac{n(\alpha+\beta)}{n}E\left(\sum_{i=1}^{n}E(X)\right) - [(\alpha+\beta)-(\alpha p_1+\beta p_2)]\right]$$

$$= \frac{(\alpha+\beta)\left(\left\{\frac{\alpha(2p_1-1)+\beta(2p_2-1)}{\alpha+\beta}\right\}\pi + \left\{\frac{\alpha(1-p_1)+\beta(1-p_2)}{\alpha+\beta}\right\}\right) - [(\alpha+\beta)-(\alpha p_1+\beta p_2)]}{\alpha(2p_1-1)+\beta(2p_2-1)}$$

$$= \frac{[\alpha(2p_1-1)+\beta(2p_2-1)]\pi+\alpha-\alpha p_1+\beta-\beta p_2-\alpha-\beta+\alpha p_1+\beta p_2}{\alpha(2p_1-1)+\beta(2p_2-1)}$$

$$= \pi$$

with variance as

$$Var(\hat{\pi}_{cdw}) = \frac{(\alpha+\beta)^2}{n^2[\alpha(2p_1-1)+\beta(2p_2-1)]^2}Var\left(\sum_{i=1}^{n}X_i\right) = \frac{n(\alpha+\beta)^2}{n^2[\alpha(2p_1-1)+\beta(2p_2-1)]^2}Var(X). \tag{32}$$

Substituting (26) in (32) and simplify to have the following results

$$Var(\hat{\pi}_{cdw}) = \frac{n(\alpha+\beta)^2}{n^2[\alpha(2p_1-1)+\beta(2p_2-1)]^2}\times \left[\frac{[\alpha(2p_1-1)+\beta(2p_2-1)]^2\pi(1-\pi)+(p_1\alpha+p_2\beta)[(\alpha+\beta-p_1\alpha-p_2\beta)]}{(\alpha+\beta)^2}\right]$$

$$= \frac{[(2p_1-1)\alpha+(2p_2-1)\beta]^2\pi(1-\pi)+(p_1\alpha+p_2\beta)[(\alpha+\beta-p_1\alpha-p_2\beta)]}{n[\alpha(2p_1-1)+\beta(2p_2-1)]^2}$$

which finally simplified to

$$Var(\hat{\pi}_{cdw}) = \frac{\pi(1-\pi)}{n} + \frac{(p_1\alpha+p_2\beta)[(\alpha+\beta)-(p_1\alpha+p_2\beta)]}{n[\alpha(2p_1-1)+\beta(2p_2-1)]^2}. \tag{33}$$

Putting $p_1 = p_2 = p$ in (31) and (33) one may check whether the results are original Warner estimator in (1) and (2), respectively as follows:

$$\begin{aligned}\hat{\pi}_{cdw} &= \frac{\frac{n'}{n}(\alpha+\beta) - [(\alpha+\beta) - (\alpha p + \beta p)]}{\alpha(2p-1) + \beta(2p-1)} \\ &= \frac{\frac{n'}{n}(\alpha+\beta) - [(\alpha+\beta) - p(\alpha+\beta)]}{(2p-1)(\alpha+\beta)} \\ &= \frac{\frac{n'}{n} - (1-p)}{(2p-1)} \\ &= \hat{\pi}_w\end{aligned}$$

and

$$\begin{aligned}Var(\hat{\pi}_{cdw}) &= \frac{\pi(1-\pi)}{n} + \frac{(p\alpha + p\beta)[(\alpha+\beta) - (p\alpha + p\beta)]}{n[\alpha(2p-1) + \beta(2p-1)]^2} \\ &= \frac{\pi(1-\pi)}{n} + \frac{p(\alpha+\beta)[(\alpha+\beta) - p(\alpha+\beta)]}{n[(2p-1)(\alpha+\beta)]^2} \\ &= \frac{\pi(1-\pi)}{n} + \frac{(\alpha+\beta)^2 p(1-p)}{n(2p-1)^2(\alpha+\beta)^2} \\ &= \frac{\pi(1-\pi)}{n} + \frac{p(1-p)}{n(2p-1)^2} \\ &= Var(\hat{\pi}_w).\end{aligned}$$

2.2 Correction of Ewemooje et al. (2018) Dichotomous Mangat RRM

Without loss of generality, from (11),

$$\begin{aligned}\theta &= \pi + \frac{\alpha(1-p_1)}{\alpha+\beta} - \frac{\alpha\pi(1-p_1)}{\alpha+\beta} + \frac{\beta(1-p_2)}{\alpha+\beta} - \frac{\beta\pi(1-p_2)}{\alpha+\beta} \\ &= \left[1 - \frac{\alpha(1-p_1)}{\alpha+\beta} - \frac{\beta(1-p_2)}{\alpha+\beta}\right]\pi + \frac{\alpha(1-p_1)}{\alpha+\beta} + \frac{\beta(1-p_2)}{\alpha+\beta} \\ &= \left[\frac{\alpha p_1 + \beta p_2}{\alpha+\beta}\right]\pi + \frac{\alpha+\beta-\alpha p_1 - \beta p_2}{\alpha+\beta}\end{aligned}$$

which finally gives

$$\theta = \left[\frac{\alpha p_1 + \beta p_2}{\alpha+\beta}\right]\pi + \left[1 - \left\{\frac{\alpha p_1 + \beta p_2}{\alpha+\beta}\right\}\right] \quad (34)$$

Similarly, probability of no-answer in (12) can be rewritten as

$$1-\theta = \frac{\alpha p_1}{\alpha+\beta} - \frac{\alpha\pi p_1}{\alpha+\beta} + \frac{\beta p_2}{\alpha+\beta} - \frac{\beta\pi p_2}{\alpha+\beta} = \frac{\alpha p_1 + \beta p_2}{\alpha+\beta} - \left[\frac{\alpha p_1 + \beta p_2}{\alpha+\beta}\right]\pi \quad (35)$$

so that the dichotomous random variable

$$Y_i = \begin{cases} 1; & \text{if ith respondent say yes with probability } \left[\frac{\alpha p_1+\beta p_2}{\alpha+\beta}\right]\pi + \left[1 - \left\{\frac{\alpha p_1+\beta p_2}{\alpha+\beta}\right\}\right] \\ 0; & \text{if ith respondent say no with probability } \frac{\alpha p_1+\beta p_2}{\alpha+\beta} - \left[\frac{\alpha p_1+\beta p_2}{\alpha+\beta}\right]\pi. \end{cases} \quad (36)$$

which consequently defined

$$E(Y) = \sum_{i=0}^{n} Y_i P(Y_i) = \left[\frac{\alpha p_1 + \beta p_2}{\alpha+\beta}\right]\pi + \left[1 - \left\{\frac{\alpha p_1 + \beta p_2}{\alpha+\beta}\right\}\right], \quad (37)$$

$$E(Y^2)=\sum_{i=0}^{n}Y_i^2P(Y_i)=\left[\frac{\alpha p_1+\beta p_2}{\alpha+\beta}\right]\pi+\left[1-\left\{\frac{\alpha p_1+\beta p_2}{\alpha+\beta}\right\}\right], \tag{38}$$

and

$$\begin{aligned}Var(Y)&=\left[\left\{\frac{\alpha p_1+\beta p_2}{\alpha+\beta}\right\}\pi+\left\{1-\frac{\alpha p_1+\beta p_2}{\alpha+\beta}\right\}\right]-\left[\left\{\frac{\alpha p_1+\beta p_2}{\alpha+\beta}\right\}\pi+\left\{1-\frac{\alpha p_1+\beta p_2}{\alpha+\beta}\right\}\right]^2\\&=\left[\left\{\frac{\alpha p_1+\beta p_2}{\alpha+\beta}\right\}\pi+\left\{1-\frac{\alpha p_1+\beta p_2}{\alpha+\beta}\right\}\right]\times\left[1-\left[\left\{\frac{\alpha p_1+\beta p_2}{\alpha+\beta}\right\}\pi+\left\{1-\frac{\alpha p_1+\beta p_2}{\alpha+\beta}\right\}\right]\right].\end{aligned}$$

After little Algebra, this gives

$$Var(Y)=\frac{(\alpha p_1+\beta p_2)^2\pi(1-\pi)+[(\alpha+\beta)(\alpha p_1+\beta p_2)-(\alpha p_1+\beta p_2)^2](1-\pi)}{(\alpha+\beta)^2}. \tag{39}$$

Throughout the survey-experiment, only two possible outcomes "yes" and "no"; trials are independent; and prob of yes as given in (34) is constant from trial to trial. Therefore, probability of obtaining n'-yes out of n-subject follows binomial distribution with likelihood function

$$L=\prod_{i=1}^{n}\binom{n}{n'}\left[\left\{\frac{\alpha p_1+\beta p_2}{\alpha+\beta}\right\}\pi+\left\{1-\frac{\alpha p_1+\beta p_2}{\alpha+\beta}\right\}\right]^{n'}\left[\left\{\frac{\alpha p_1+\beta p_2}{\alpha+\beta}\right\}-\left\{\frac{\alpha p_1+\beta p_2}{\alpha+\beta}\right\}\pi\right]^{n-n'}. \tag{40}$$

Setting $\frac{d}{d\pi}\ln L=0$, where ln is a natural logarithm function, we have

$$\frac{n'\left[\frac{\alpha p_1+\beta p_2}{\alpha+\beta}\right]}{\left[\frac{\alpha p_1+\beta p_2}{\alpha+\beta}\right]\pi+\left[1-\frac{\alpha p_1+\beta p_2}{\alpha+\beta}\right]}=\frac{\left[\frac{\alpha p_1+\beta p_2}{\alpha+\beta}\right](n-n')}{\left[\frac{\alpha p_1+\beta p_2}{\alpha+\beta}\right]-\left[\frac{\alpha p_1+\beta p_2}{\alpha+\beta}\right]\pi}. \tag{41}$$

Make π subject of the formula gives

$$\pi=\frac{n'A^2-(n-n')AB}{nA^2}=\frac{\frac{n'}{n}A^2-(1-\frac{n'}{n})AB}{A^2} \tag{42}$$

where $A=\frac{\alpha p_1+\beta p_2}{\alpha+\beta}$ and $B=1-\frac{\alpha p_1+\beta p_2}{\alpha+\beta}$. Explicitly,

$$\begin{aligned}\pi_{cdm}&=\frac{\frac{n'}{n}\left[\frac{\alpha p_1+\beta p_2}{\alpha+\beta}\right]^2-(1-\frac{n'}{n})\left[\frac{\alpha p_1+\beta p_2}{\alpha+\beta}\right]\left[1-\left\{\frac{\alpha p_1+\beta p_2}{\alpha+\beta}\right\}\right]}{\left[\frac{\alpha p_1+\beta p_2}{\alpha+\beta}\right]^2}\\&=\frac{\frac{n'}{n}\left[\frac{\alpha p_1+\beta p_2}{\alpha+\beta}\right]-(1-\frac{n'}{n})\left[\frac{\alpha+\beta-\alpha p_1-\beta p_2}{\alpha+\beta}\right]}{\frac{\alpha p_1+\beta p_2}{\alpha+\beta}}\\&=\frac{\frac{n'}{n}[\alpha p_1+\beta p_2]-(1-\frac{n'}{n})[\alpha+\beta-\alpha p_1-\beta p_2]}{\alpha p_1+\beta p_2}\\&=\frac{\frac{n'}{n}[\alpha p_1+\beta p_2]+\frac{n'}{n}[\alpha+\beta-\alpha p_1-\beta p_2]-[\alpha+\beta-\alpha p_1-\beta p_2]}{\alpha p_1+\beta p_2}\\&=\frac{\frac{n'}{n}[\alpha p_1+\beta p_2+\alpha+\beta-\alpha p_1-\beta p_2]-[\alpha+\beta-\alpha p_1-\beta p_2]}{\alpha p_1+\beta p_2}\end{aligned}$$

which finally gives

$$\hat{\pi}_{cdm}=\frac{\frac{n'}{n}(\alpha+\beta)-(\alpha+\beta)+\alpha p_1+\beta p_2}{\alpha p_1+\beta p_2}. \tag{43}$$

The corrected estimator is unbiased as

$$\begin{aligned}
E(\hat{\pi}_{cdm}) &= \frac{1}{\alpha p_1+\beta p_2}\left[(\alpha+\beta)E\left(\frac{n'}{n}\right)-(\alpha+\beta)+\alpha p_1+\beta p_2\right] \\
&= \frac{1}{\alpha p_1+\beta p_2}\left[\frac{(\alpha+\beta)}{n}\sum_{i=1}^{n}E(Y)-(\alpha+\beta)+\alpha p_1+\beta p_2\right] \\
&= \frac{\left[(\alpha+\beta)\left\{\pi\left[\frac{\alpha p_1+\beta p_2}{\alpha+\beta}\right]+\frac{\alpha+\beta-\alpha p_1-\beta p_2}{\alpha+\beta}\right\}-(\alpha+\beta)+\alpha p_1+\beta p_2\right]}{\alpha p_1+\beta p_2} \\
&= \frac{[\pi(\alpha p_1+\beta p_2)+\alpha+\beta-\alpha p_1-\beta p_2-\alpha-\beta+\alpha p_1+\beta p_2]}{\alpha p_1+\beta p_2} \\
&= \frac{[\pi(\alpha p_1+\beta p_2)]}{\alpha p_1+\beta p_2} \\
&= \pi.
\end{aligned}$$

with variance as

$$Var(\hat{\pi}_{cdm}) = \frac{\frac{(\alpha+\beta)^2}{n^2}\sum_{i=1}^{n}Var(Y)}{[\alpha p_1+\beta p_2]^2}. \tag{44}$$

Putting (39) in (44) we have the following

$$\begin{aligned}
Var(\hat{\pi}_{cdm}) &= \frac{\frac{n(\alpha+\beta)^2}{n^2}\left[\frac{(\alpha p_1+\beta p_2)^2\pi(1-\pi)+[(\alpha+\beta)(\alpha p_1+\beta p_2)-(\alpha p_1+\beta p_2)^2](1-\pi)}{(\alpha+\beta)^2}\right]}{[\alpha p_1+\beta p_2]^2} \\
&= \frac{(\alpha p_1+\beta p_2)^2\pi(1-\pi)+[(\alpha+\beta)(\alpha p_1+\beta p_2)-(\alpha p_1+\beta p_2)^2](1-\pi)}{n[\alpha p_1+\beta p_2]^2} \\
&= \frac{(\alpha p_1+\beta p_2)^2\pi(1-\pi)}{n[\alpha p_1+\beta p_2]^2}+\frac{[(\alpha+\beta)(\alpha p_1+\beta p_2)-(\alpha p_1+\beta p_2)^2](1-\pi)}{n[\alpha p_1+\beta p_2]^2}
\end{aligned}$$

which after simplification yields

$$Var(\hat{\pi}_{cdm}) = \frac{\pi(1-\pi)}{n}+\frac{[(\alpha+\beta)-(\alpha p_1+\beta p_2)](1-\pi)}{n(\alpha p_1+\beta p_2)}. \tag{45}$$

Then, setting $p_1 = p_2 = p$ in (43) and (45) return conventional Mangat estimators as

$$\hat{\pi}_{cdm} = \frac{\frac{n'}{n}(\alpha+\beta)-(\alpha+\beta)+p(\alpha+\beta)}{p(\alpha+\beta)} = \frac{\frac{n'}{n}-1+p}{p} = \hat{\pi}_m$$

and

$$\begin{aligned}
Var(\hat{\pi}_{cdm}) &= \frac{\pi(1-\pi)}{n}+\frac{[(\alpha+\beta)-(\alpha p+\beta p)](1-\pi)}{n(\alpha p+\beta p)} \\
&= \frac{\pi(1-\pi)}{n}+\frac{[(\alpha+\beta)-p(\alpha+\beta)](1-\pi)}{np(\alpha+\beta)} \\
&= \frac{\pi(1-\pi)}{n}+\frac{(1-p)(1-\pi)}{np} \\
&= Var(\hat{\pi}_m).
\end{aligned}$$

2.3 Relative Efficiency

Suppose $\hat{\pi}_1$ and $\hat{\pi}_2$ are two unbiased estimators of population parameter π with $Var(\hat{\pi}_1) < Var(\hat{\pi}_2)$ or Relative Efficiency (RE) defined as $\frac{Var(\hat{\pi}_2)}{Var(\hat{\pi}_1)} > 1$, then we say $\hat{\pi}_1$ is more efficient than $\hat{\pi}_2$. To facilitate interpretation, relative efficiency can be multiplied by 100 to obtain Percent Relative Efficiency (PRE). Using (2) and (33), corrected-dichotomous-Warner estimator is more efficient than earliest Warner's model if

$$\frac{\pi(1-\pi)}{n}+\frac{p(1-p)}{(2p-1)^2n}-\left[\frac{\pi(1-\pi)}{n}+\frac{(p_1\alpha+p_2\beta)[(\alpha+\beta)-(p_1\alpha+p_2\beta)]}{n[\alpha(2p_1-1)+\beta(2p_2-1)]^2}\right]>0$$

which when simplified, it establish the condition for efficiency as

$$\frac{p(1-p)}{(2p-1)^2}>\frac{(p_1\alpha+p_2\beta)(\alpha+\beta-p_1\alpha-p_2\beta)}{[\alpha(2p_1-1)+\beta(2p_2-1)]^2}. \tag{46}$$

Furthermore, using (4) and (45), the corrected-convoluted-Mangat performs better than conventional Mangat-improved-two-step RRM if $Var(\hat{\pi}_m)>Var(\hat{\pi}_{cdm})$. That is, if

$$\frac{\pi(1-\pi)}{n}+\frac{(1-p)(1-\pi)}{np}-\left[\frac{\pi(1-\pi)}{n}+\frac{[(\alpha+\beta)-(\alpha p_1+\beta p_2)](1-\pi)}{n(\alpha p_1+\beta p_2)}\right]>0.$$

This is equivalent to

$$(1-p)(\alpha p_1+\beta p_2)-p(\alpha+\beta-\alpha p_1-\beta p_2)>0.$$

Expand and collect like terms,

$$(\alpha+\beta)p<p_1\alpha+p_2\beta \quad \text{or} \quad p<\frac{p_1\alpha+p_2\beta}{(\alpha+\beta)}. \tag{47}$$

The inequality in (47) holds $\forall\ \alpha,\beta\in\mathbb{Z}^+$ and $p,p_1,p_2\in(0,1)$.

3 Results and Discussion

This section gives a comprehensive results of all the competing models. The performance of the model were compared with one another using Percentage Relative Efficiency criterion, $PRE=\frac{\text{conventional model}}{\text{corrected-dichotomous model}}\times 100\%$. $PRE>100$ indicating the efficiency of corrected-dichotomous over the orthodox RRT. The numerical computation was carried out for values of $\pi=\{0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9\}$ and $p=p_1,p_2\subset(0,1)$. The results are numerically tabulated and visually presented as follow.

Table 1: PRE of corrected-dichotomous-Warner over Warner conventional RRM for $\alpha<\beta(\alpha=2.00,\beta=90.00)$

		π								
p_2	$p=p_1$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.30	0.30	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
0.40		24.06	24.97	25.60	25.97	26.09	25.97	25.60	24.97	24.06
0.60		20.05	20.84	21.40	21.73	21.84	21.73	21.40	20.84	20.05
0.70		90.61	91.02	91.28	91.44	91.49	91.44	91.28	91.02	90.61
0.30	0.40	423.79	408.75	398.83	393.18	391.35	393.18	398.83	408.75	423.79
0.40		100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
0.60		91.28	91.37	91.44	91.48	91.49	91.48	91.44	91.37	91.28
0.70		403.41	389.96	381.07	375.99	374.33	375.99	381.07	389.96	403.41
0.30	0.60	403.41	389.96	381.07	375.99	374.33	375.99	381.07	389.96	403.41
0.40		91.28	91.37	91.44	91.48	91.49	91.48	91.44	91.37	91.28
0.60		100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
0.70		423.79	408.75	398.83	393.18	391.35	393.18	398.83	408.75	423.79
0.30	0.70	90.61	91.02	91.28	91.44	91.49	91.44	91.28	91.02	90.61
0.40		20.05	20.84	21.40	21.73	21.84	21.73	21.40	20.84	20.05
0.60		24.06	24.97	25.60	25.97	26.09	25.97	25.60	24.97	24.06
0.70		100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Table 2: PRE of corrected-dichotomous-Warner over original Warner RRM for $\alpha > \beta(\alpha = 90.00, \beta = 2.00)$

		π								
p_2	$p_1 = p$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.30	0.30	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
0.40		97.59	97.70	97.78	97.82	97.83	97.82	97.78	97.70	97.59
0.60		92.90	93.21	93.42	93.54	93.58	93.54	93.42	93.21	92.90
0.70		90.61	91.02	91.28	91.44	91.49	91.44	91.28	91.02	90.61
0.30	0.40	104.51	104.46	104.42	104.40	104.39	104.40	104.42	104.46	104.51
0.40		100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
0.60		91.28	91.37	91.44	91.48	91.49	91.48	91.44	91.37	91.28
0.70		87.09	87.22	87.31	87.36	87.38	87.36	87.31	87.22	87.09
0.30	0.60	87.09	87.22	87.31	87.36	87.38	87.36	87.31	87.22	87.09
0.40		91.28	91.37	91.44	91.48	91.49	91.48	91.44	91.37	91.28
0.60		100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
0.70		104.51	104.46	104.42	104.40	104.39	104.40	104.42	104.46	104.51
0.30	0.70	90.61	91.02	91.28	91.44	91.49	91.44	91.28	91.02	90.61
0.40		92.90	93.21	93.42	93.54	93.58	93.54	93.42	93.21	92.90
0.60		97.59	97.70	97.78	97.82	97.83	97.82	97.78	97.70	97.59
0.70		100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Table 3: PRE of corrected-convoluted-Mangat over existing Mangat RRM for $\alpha = 2.00$ and $\beta = 90.00$

		π								
p_2	$p_1 = p$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.30	0.30	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
0.40		150.79	147.83	145.19	142.83	140.71	138.78	137.03	135.43	133.96
0.50		217.74	208.07	199.87	192.82	186.71	181.34	176.61	172.39	168.62
0.60		309.89	286.26	267.35	251.92	239.10	228.28	219.01	211.00	203.10
0.70		445.22	391.83	352.74	322.88	299.34	280.29	264.56	251.36	240.12
0.30	0.40	66.41	67.75	68.98	70.13	71.19	72.18	73.11	73.97	74.78
0.40		100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
0.50		144.31	140.64	137.54	134.87	132.56	130.54	128.75	127.16	125.74
0.60		205.44	193.44	183.89	176.12	169.66	164.21	159.56	155.53	152.01
0.70		295.20	264.79	242.59	225.65	212.31	201.53	192.64	185.18	178.84
0.30	0.50	46.11	48.28	50.28	52.13	53.85	55.45	56.94	58.33	59.64
0.40		69.34	71.15	72.77	74.21	75.51	76.68	77.75	78.72	79.61
0.50		100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
0.60		142.35	137.50	133.64	130.51	127.91	125.71	123.84	122.22	120.81
0.70		204.65	188.24	176.27	167.16	160.00	154.22	149.45	145.45	142.06
0.30	0.60	32.45	35.20	37.72	40.06	42.23	44.25	46.13	47.89	49.54
0.40		48.74	51.80	54.52	56.95	59.13	61.10	62.89	64.52	66.02
0.50		70.25	72.75	74.86	76.66	78.23	79.60	80.80	81.88	82.83
0.60		100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
0.70		143.83	136.91	131.87	128.05	125.04	122.62	120.62	118.95	117.53
0.30	0.70	22.59	25.77	28.69	31.39	33.90	36.22	38.39	40.42	42.32
0.40		33.89	37.87	41.40	44.55	47.38	49.93	52.25	54.36	56.30
0.50		48.82	53.14	56.80	59.92	62.62	64.99	67.07	68.91	70.57
0.60		69.49	73.03	75.84	78.12	80.00	81.59	82.94	84.11	85.13
0.70		100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

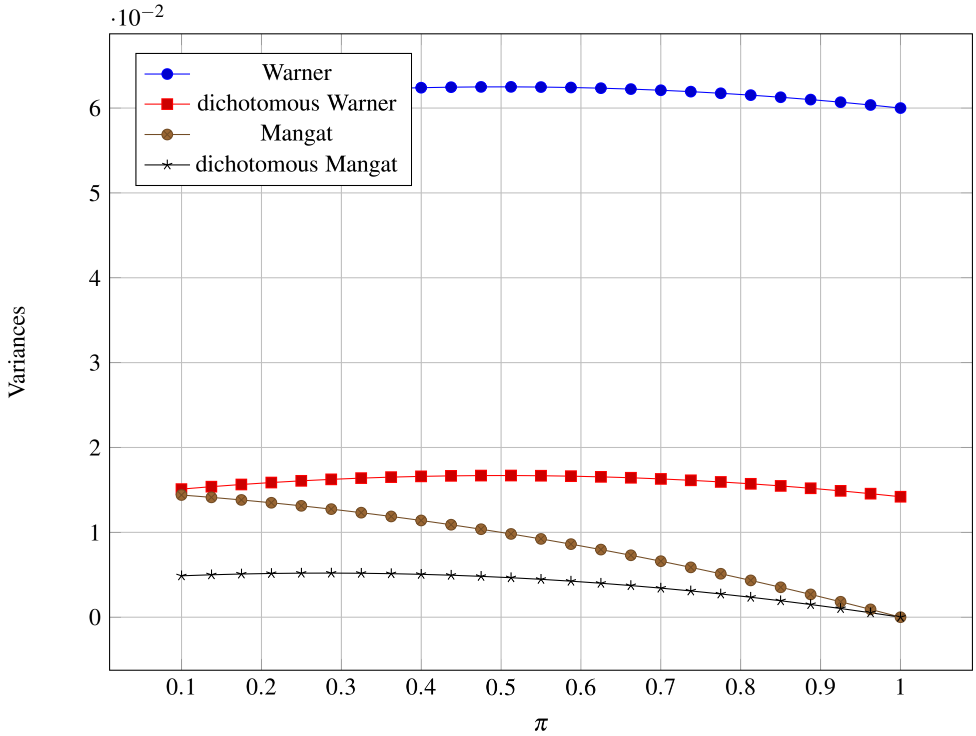

Fig. 1: Efficiency comparison of all the competing models for $n = 100, p = p_1 = 0.4, p_2 = 0.7$ and $\alpha = 2 < \beta = 90$.

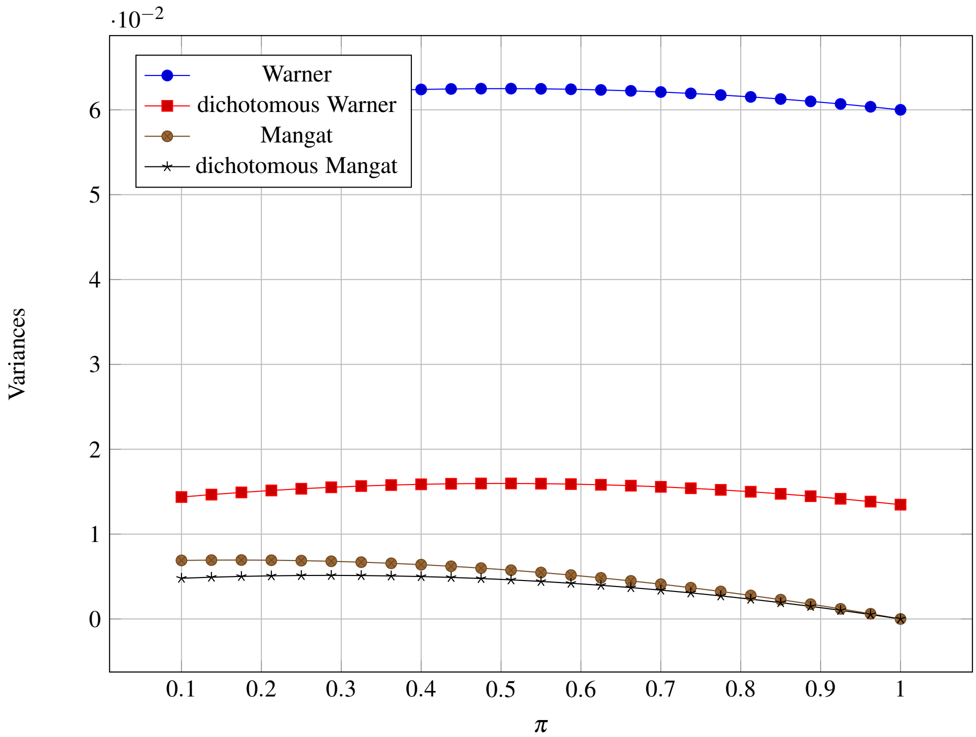

Fig. 2: Efficiency comparison of all the competing models for $n = 100, p = p_1 = 0.6, p_2 = 0.7$ and $\alpha = 2 < \beta = 90$

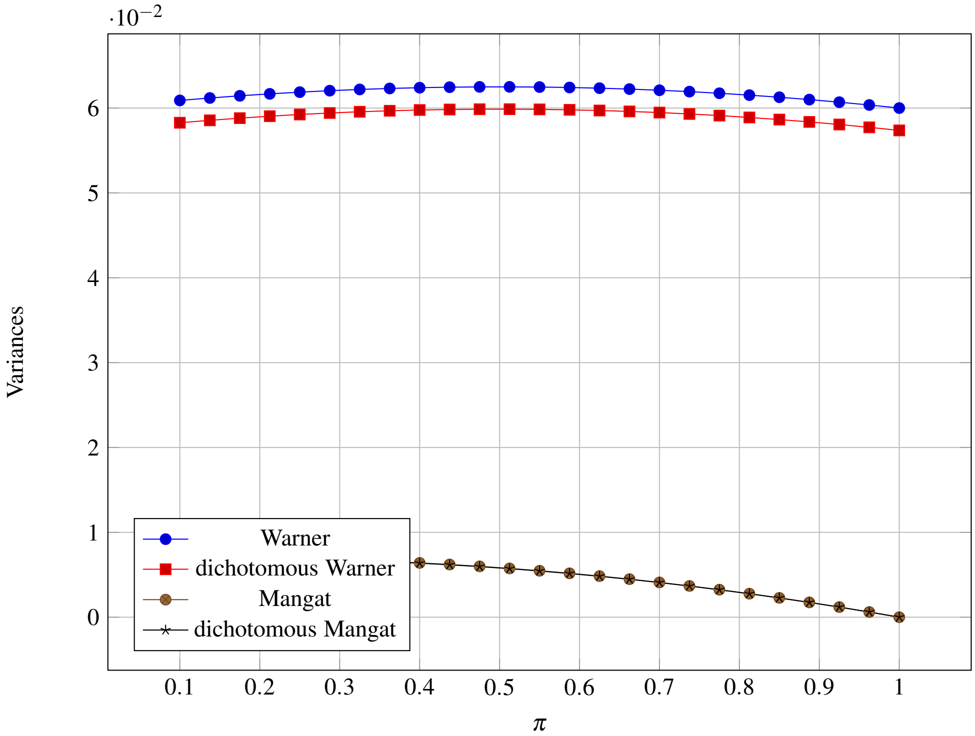

Fig. 3: Efficiency comparison of all the competing models for $n = 100, p = p_1 = 0.6, p_2 = 0.7$ and $\alpha = 90 > \beta = 2$

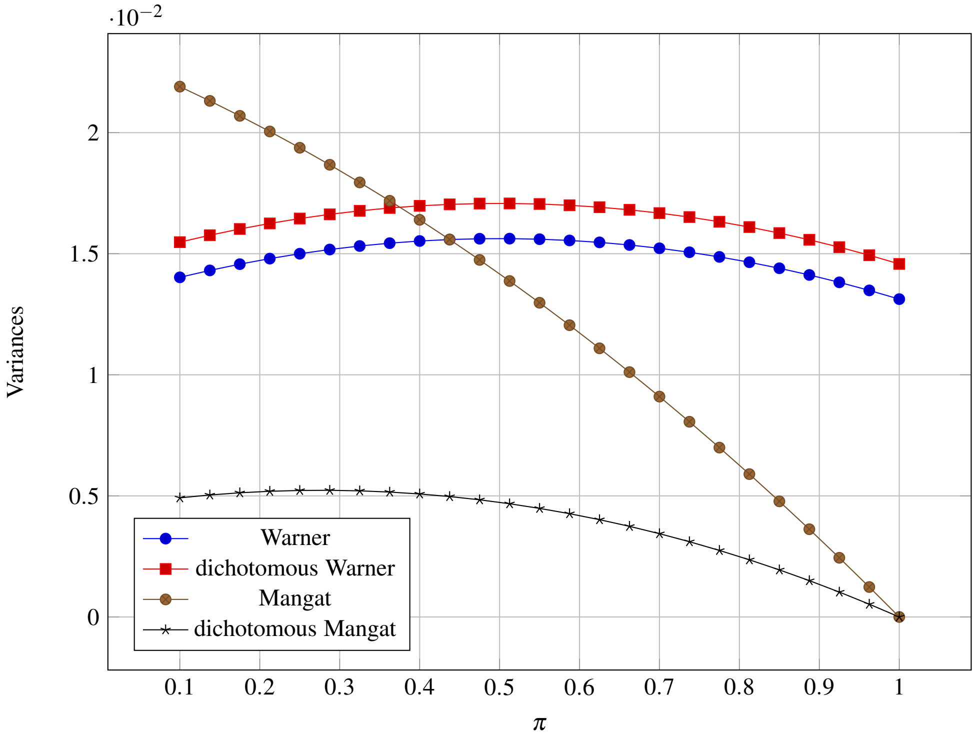

Fig. 4: Efficiency comparison of all the competing models for $n = 100, p = p_1 = 0.3, p_2 = 0.7$ and $\alpha = 2 < \beta = 90$

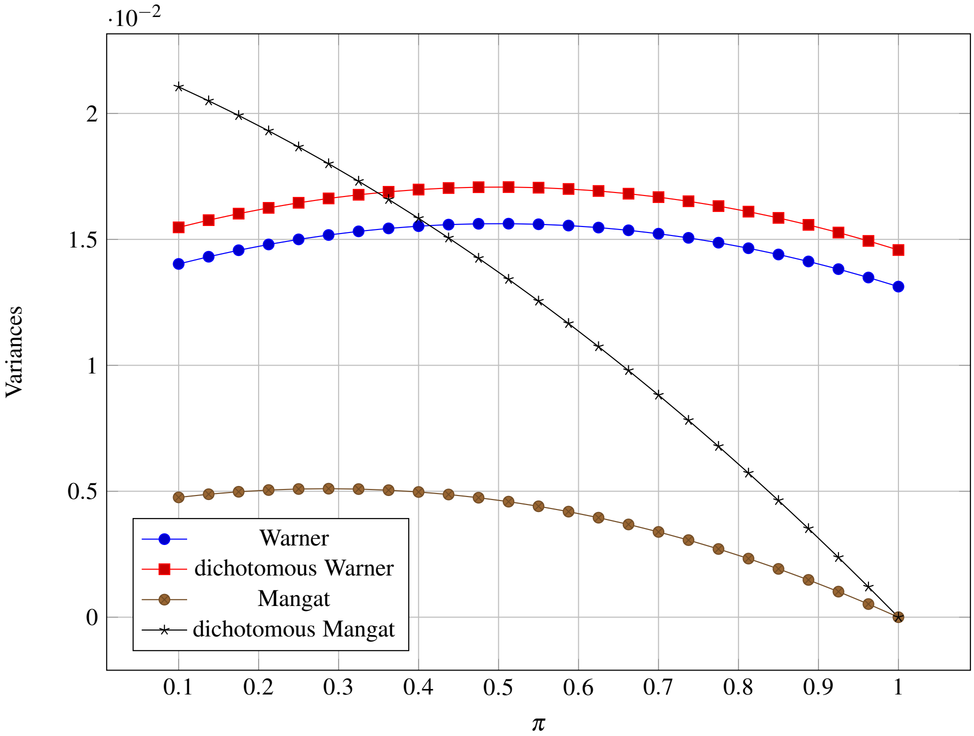

Fig. 5: Efficiency comparison of all the competing models for $n = 100, p = p_1 = 0.7, p_2 = 0.3$ and $\alpha = 2 < \beta = 90$

Empirical results from Table 1 when $\alpha < \beta$ revealed that Warner dichotomous conditionally performs better than the original Warner at $p = p_2 = (0.4, 0.6)$ for $p_2 = (0.3, 0.7)$, respectively. Similar inferences were drawn from Table 2 for $\alpha > \beta$. Evidence from (47) and Table 3, the corrected-convoluted-Mangat is more efficient than the orthodox [2] RRM $\forall \alpha, \beta > 0$ and $\forall\ p, p_1, p_2$ such that $p_1 = p < p_2$. If $p_1 = p_2 = p$, there is parallel efficiency, i.e., relative efficiency $= 1$ or $PRE = 100\%$ for the two pairs; corrected-dichotomous-Warner versus conventional Warner model and corrected-dichotomous-Mangat against orthodox Mangat model. The study deduced that corrected-dichotomous Mangat has least variance among the four competing models regardless of whether $\alpha > or < \beta$ in as much as $p = p_1 < p_2$. This condition is in agreement with (47). Figure 1-5 graphically illustrates the performance of all the four competing models for varying values of π (prevalence of sensitive attribute under study), p (randomization parameters), and real-value constants α and β. Evidence from Figure 5, reverse is the inference when $p = p_1 > p_2$. Observation also depicts that as π (prevalence-rate of the sensitive variable) decreases, the efficiency of corrected-dichotomous-Mangat increases. When $\alpha < \beta$ and $p = p_1 < p_2$, corrected-dichotomous Warner performs better that the conventional Warner, similarly, the corrected-dichotomous Mangat compare to the original Mangat.

The convoluted RRMs suggested by [4] and [6] are new RRMs that are more efficient and secure (keep privacy) than the conventional [1] and [2] models. Unfortunately, mathematical derivation and accompany statistical properties of [4] and [6] estimators are wrong. It is essential to note that their wrong estimation does not affect the validity of the designs and motivation of their respective work.

Acknowledgements

The second author acknowledges the academic support by the first author for his mentorship and supervision.
The authors are passionately grateful to the chief-editor and anonymous referee for immeasurable suggestions, comments and contributions that greatly improved the manuscript.

References

[1] S. L. Warner, Randomised response: A survey technique for eliminating evasive answer bias, *Journal of American Statistical Association*, **60**(309), 63-69, DOI: 10.2307/2283137, (1965).

[2] N. S. Mangat, An improved randomised response strategy, *Journal of Royal Statistical Society (Series B)*, **56**(1), 93-95, DOI:10.0035-9246/94/56093, (1994).
[3] A. T. Adeniran, A. A. Sodipo and C. G. Udomboso, A modified forced randomised response model, *Journal of the Nigerian Society of Physical Sciences (JNSPS)*, **2**(1), 36-50, DOI: https://doi.org/10.46481/jnsps.2020.1, (2020).
[4] Z. Hussain and J. Shabbir, Randomized use of Warner's randomized response model, "interstat", *http://interstat.statjournals.net/YEAR/2007/articles/0704007.pdf*, (2008).
[5] A. O. Adepetun and F. B. Adebola, On the relative efficiency of the proposed reparameterized randomized response model, *International Journal of Mathematics and Statistics Research*, **1**(1), 42-52, (2014).
[6] O. S. Ewemooje, F. B. Adebola and G. N. Amahia, Alternative estimator in dichotomous randomized response technique, *Communications in Mathematics and Statistics*, https://doi.org/10.1007/s40304-018-0145-x, (2018).