Direct Questioning of Sensitive Topics in Public health Studies: A Simulation Study

Jessica K. Fox, Evrim Oral*

LSUHSC School of Public Health Biostatistics Program, New Orleans, LA

LSU Health NEW ORLEANS School of Public Health

Louisiana State University Health Sciences Center School of Public Health

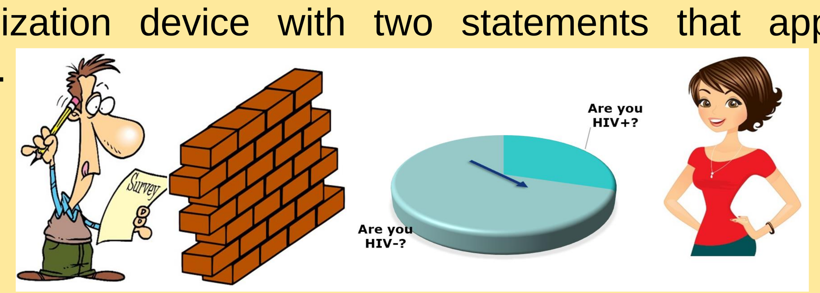
**Contact: eoral@lsuhsc.edu*

Motivation and Objectives

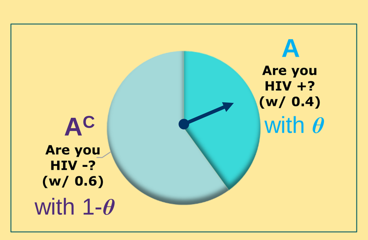
In health-related research, researchers often study sensitive topics such as sexually transmitted diseases, medical compliance, mental health, or drug and alcohol abuse. When such questions are asked directly, respondents show a tendency of answering them in a socially acceptable fashion rather than answering them truthfully, causing bias in estimates. Warner's randomized response technique (RRT) is the first, and most commonly used indirect surveying technique that has been developed to increase the respondents' confidentiality and obtain honest responses. Although it leads to an unbiased estimate of the sensitive characteristic of interest, its theoretical variance is larger than the one from the direct questioning technique (DQT). In this study, we evaluate the empirical mean square error (MSE) from DQT under different scenarios where cheating proportion in the population varies. We show that under some situations the empirical MSE from Warner's RRT becomes smaller than the MSE from DQT using an extensive Monte Carlo simulation study.

Background

Warner's randomized response technique (RRT) asks the sensitive question by providing respondents a randomization device with two statements that appear on it with known probabilities θ and $1-\theta$.



The wheel is divided into 2 areas with known probabilities for the sensitive question A, and its complement A^C. The respondent spins the wheel. The arrow lands on one of the two areas. The respondent answers truthfully with a "yes" or "no" to the question. Since the interviewer does not see where the arrow landed, he cannot know which question the respondent received. He only records what he hears from the respondent ➔ **Gives privacy protection to respondents.**



- Let π_A be the true proportion of the sensitive characteristic (A) in the population.
- Let π_λ be the proportion of "yes" responses.
- Let θ be the probability that the spinner points A.
- Then, the probability of getting a "yes" as response $\pi_\lambda = \pi_A\theta + (1-\pi_A)(1-\theta)$

Let $\lambda_i = \begin{cases} 1, \text{if the respondent i says "yes"} \\ 0, \text{if the respondent i says "no"} \end{cases}$ Warner's RRT estimator for proportion is obtained after solving the equation above:

$\hat{\pi}_{AW} = \frac{\hat{\pi}_\lambda + \theta - 1}{2\theta - 1}$, where π_λ can be estimated by $\hat{\pi}_\lambda = \frac{\sum_{i=1}^{n} \lambda_i}{n}$

and the theoretical variance of $\hat{\pi}_{AW}$ can be derived as $Var(\hat{\pi}_{AW}) = \frac{\pi_\lambda(1-\pi_\lambda)}{n(2\theta-1)^2}$

Realize that the variance above can be re-written as

$$Var(\hat{\pi}_{AW}) = \underbrace{\frac{\pi_A(1-\pi_A)}{n}}_{\text{Theoretical variance from DQT}} + \underbrace{\frac{\theta(1-\theta)}{n(2\theta-1)^2}}_{\text{Penalty for using randomization device}}$$

Thus, theoretical variance from the Warner's RRT is larger than the one from the direct questioning technique (DQT). However, the empirical variance from DQT will be different than the theoretical variance

$$Var(\hat{\pi}_{DQT}) = \frac{\pi_A(1-\pi_A)}{n}$$

since there will be participants in the sample that will not respond truthfully. To our knowledge, empirical variance and mean square error (MSE) from the DQT have not been studied under different scenarios where cheating proportion in the population varies. In the next section, we will show that, although Warner's RRT is known to have larger theoretical variance compared to the one from DQT, in practice, if there are many participants in the sample that do not respond truthfully, the empirical variance (and MSE) from Warner's RRT can be found smaller.

Simulation Study

To be able to compare the empirical variances and the MSEs of Warner's RRT with DQT, we performed a Monte Carlo simulation study using k=10,000 iterations as follows:

Simulation Study (cont.)

We assumed that true proportion of the sensitive characteristic (A) in the population= π_A=0.4. We initially considered the sample size to be 50. We considered various different values for the known probability θ changing from 0.1 to 0.4 in increments of 0.05. We also considered various proportions of not telling the truth (π_{UT}); more specifically we considered π_{UT}=0, 0.05, 0.1, 0.2, 0.3 and 0.4, π_{UT}=0 representing the case where all participants are responding truthfully. We calculated the estimates from Warner's RRT and DQT with the following formulas.

$$\hat{\pi}_{AW} = \frac{\hat{\pi}_\lambda + \theta - 1}{2\theta - 1} \quad \text{and} \quad \hat{\pi}_{DQT} = \frac{\sum_{i=1}^{n} D_i}{n}$$

where $D_i = \begin{cases} 1, \text{if the respondent i says "yes"} \\ 0, \text{if the respondent i says "no"} \end{cases}$ from DQT.

We calculated the empirical biases, variances and MSEs from the formulas given below:

$$EVar_w = \frac{\sum_{k=1}^{10,000}(\hat{\pi}_{AW} - \pi_A)^2}{10,000}, \quad EVar_{DQT} = \frac{\sum_{k=1}^{10,000}(\hat{\pi}_{DQT} - \pi_A)^2}{10,000}$$

$$MSE_W = EVar_w + (Bias_W)^2, \quad MSE_{DQT} = EVar_{DQT} + (Bias_{DQT})^2$$

We also calculated relative efficiencies using from the formula: RE= $\frac{MSE_W}{MSE_{DQT}}$

Results

θ		π_{UT} 0 (n_{UT}=0)	0.05 (n_{UT}=3)	0.1 (n_{UT}=5)	0.2 (n_{UT}=10)	0.3 (n_{UT}=15)	0.4 (n_{UT}=20)
0.1	BiasW	0.00146250	0.00050750	0.00053750	0.00091750	0.00114500	0.00071250
	BiasD	0.00055000	0.02029200	0.04150800	0.08106800	0.12062600	0.15979000
	EvarW	0.00758569	0.00764269	0.00753869	0.00751494	0.00764088	0.00752581
	EVarD	0.00489212	0.00505704	0.00632712	0.01086768	0.01852116	0.02914284
	RE	1.55093441	1.39755297	**0.93651485**	**0.43095805**	**0.23107868**	**0.13765388**
0.15	BiasW	0.00018000	0.00020571	0.00143429	0.00069429	0.00001714	0.00030286
	BiasD	0.00007200	0.02042600	0.03912600	0.08052600	0.11994600	0.15972800
	EvarW	0.01005437	0.00995576	0.00993486	0.01003363	0.00962090	0.01000392
	EVarD	0.00483064	0.00513180	0.00616316	0.01090228	0.01841500	0.02922504
	RE	2.08137824	1.79415370	1.29151408	**0.57711383**	**0.29330180**	**0.18276146**
0.2	BiasW	0.00036667	0.00052000	0.00103667	0.00076667	0.00045000	0.00213333
	BiasD	0.00037600	0.01994000	0.04079800	0.08041600	0.11988600	0.16153200
	EvarW	0.01381867	0.01381044	0.01368767	0.01370089	0.01329678	0.01349778
	EVarD	0.00473480	0.00511400	0.00623044	0.01078536	0.01835964	0.02971608
	RE	2.91847354	2.50575256	1.73386771	**0.79419214**	**0.40623430**	**0.24193964**
0.25	BiasW	0.00101200	0.00164800	0.00052400	0.00192400	0.00098800	0.00132800
	BiasD	0.00048200	0.02005800	0.04088800	0.08039800	0.12011800	0.16051000
	EvarW	0.01997152	0.01958496	0.02012032	0.01913824	0.01993024	0.01993888
	EVarD	0.00473356	0.00512492	0.00627376	0.01080428	0.01848812	0.02939540
	RE	4.21914245	3.54384177	2.53229757	1.10851347	**0.60550921**	**0.36151297**
0.3	BiasW	0.00160500	0.00028000	0.00170500	0.00277000	0.00171500	0.00111000
	BiasD	0.00010400	0.02045800	0.04036200	0.08037000	0.11954400	0.16008000
	EvarW	0.03227975	0.03135150	0.03072325	0.03113500	0.03032675	0.03054200
	EVarD	0.00483912	0.00513388	0.00626172	0.01080292	0.01829872	0.02924128
	RE	6.67110009	5.64648139	3.89391621	1.80409046	**0.93065872**	**0.55667879**
0.35	BiasW	0.00490000	0.00225333	0.00098000	0.00184000	0.00177333	0.00242667
	BiasD	0.00103200	0.01962800	0.03963800	0.08029200	0.12017800	0.15947200
	EvarW	0.05447733	0.05411822	0.05638489	0.05612267	0.05638400	0.05686133
	EVarD	0.00483064	0.00508184	0.00618084	0.01083984	0.01848156	0.02908592
	RE	11.27994012	9.89982179	7.27370601	3.24678684	1.71262942	1.04310532
0.4	BiasW	0.00020000	0.00077000	0.00088000	0.00396000	0.00056000	0.00143000
	BiasD	0.00004800	0.02034400	0.03968800	0.08006000	0.11951200	0.15954200
	EvarW	0.12496800	0.12423700	0.12435600	0.12503400	0.12335400	0.12406500
	EVarD	0.00466648	0.00510488	0.00617992	0.01075368	0.01821848	0.02911148
	RE	26.77992369	22.51187411	16.03557123	7.28588331	3.79533071	2.27374232

It can be seen that, when $n_{UT}\neq 0$, all biases from the Warner's RRT are smaller than the biases from DQT, which was expected. We see that for lower θ values and higher π_{UT} values, empirical variances and MSEs from the Warner's RRT are smaller than the ones from DQT.

Results (cont.)

We repeated the same simulation study for a larger sample size, i.e. for n=500. We report our results below.

θ		π_{UT} 0 (n_{UT}=0)	0.05 (n_{UT}=3)	0.1 (n_{UT}=5)	0.2 (n_{UT}=10)	0.3 (n_{UT}=15)	0.4 (n_{UT}=20)
0.1	BiasW	0.00012025	0.00016550	0.00005050	0.00019775	0.00032800	0.00020525
	BiasD	0.00007580	0.02030120	0.04001400	0.08004080	0.11994380	0.15989120
	EvarW	0.00075239	0.00076318	0.00078093	0.00075317	0.00075790	0.00075991
	EVarD	0.00047633	0.00088152	0.00207062	0.00683624	0.01478578	0.02593312
	RE	1.57955289	0.58995662	0.21268703	0.05687679	0.02598394	0.01475677
0.15	BiasW	0.00005143	0.00023771	0.00038943	0.00009457	0.00011257	0.00012143
	BiasD	0.00012640	0.01993860	0.03994000	0.08017200	0.12015240	0.16000060
	EvarW	0.00102541	0.00102430	0.00099806	0.00100276	0.00098568	0.00099139
	EVarD	0.00049305	0.00087741	0.00205620	0.00685273	0.01483364	0.02596835
	RE	2.07965900	0.80344384	0.27337714	0.07550826	0.03367548	0.01922490
0.2	BiasW	0.00003200	0.00037567	0.00011967	0.00038400	0.00018633	0.00006833
	BiasD	0.00029760	0.01988640	0.04006060	0.08003240	0.12021740	0.16012340
	EvarW	0.00135615	0.00135861	0.00140188	0.00139133	0.00138109	0.00136911
	EVarD	0.00049341	0.00086234	0.00206565	0.00684248	0.01485104	0.02600761
	RE	2.74806563	1.08024890	0.38193427	0.10503544	0.04713203	0.02650903
0.25	BiasW	0.00015600	0.00016560	0.00078800	0.00012000	0.00054640	0.00064120
	BiasD	0.00013280	0.01981260	0.03991360	0.08036780	0.11976880	0.15989720
	EvarW	0.00200772	0.00201549	0.00194901	0.00201588	0.00198481	0.00199559
	EVarD	0.00048910	0.00087192	0.00204519	0.00688557	0.01475694	0.02592671
	RE	4.10478561	1.59397740	0.53586645	0.15106534	0.06821313	0.03876199
0.3	BiasW	0.00100200	0.00046800	0.00066100	0.00042900	0.00015850	0.00016150
	BiasD	0.00032520	0.02026560	0.04020300	0.07973640	0.11993100	0.15954180
	EvarW	0.00312633	0.00310157	0.00311248	0.00300643	0.00307171	0.00317569
	EVarD	0.00047357	0.00086663	0.00208485	0.00678918	0.01477697	0.02582587
	RE	6.60225069	2.42834309	0.84107113	0.22869075	0.10533913	0.06192965
0.35	BiasW	0.00112333	0.00016267	0.00148667	0.00018667	0.00002067	0.00005400
	BiasD	0.00035580	0.02004520	0.04003560	0.07966920	0.12018620	0.16006100
	EvarW	0.00541736	0.00563916	0.00543250	0.00547354	0.00561290	0.00548992
	EVarD	0.00049465	0.00086689	0.00207086	0.00678202	0.01485324	0.02598126
	RE	10.95163000	4.44485750	1.47935296	0.41690052	0.19158000	0.10639233
0.4	BiasW	0.00035300	0.00031400	0.00027500	0.00073700	0.00071800	0.00145800
	BiasD	0.00019220	0.02024680	0.03993260	0.08009880	0.12000480	0.15987620
	EvarW	0.01230399	0.01250636	0.01245555	0.01257553	0.01212540	0.01222522
	EVarD	0.00047742	0.00087426	0.00205381	0.00684841	0.01480958	0.02593021
	RE	25.76995000	9.73878200	3.41397113	0.94811920	0.41511854	0.23746749

We see that when the sample size was increased to 500, for all higher values of π_{UT} (i.e. when $\pi_{UT}\geq$0.2), empirical MSEs from the Warner's RRT become smaller than the ones from DQT. All biases from the Warner's RRT are smaller than the biases from DQT when there is no cheating, which was expected.

Discussion

Social desirability bias (SDB) has long been recognized as a serious problem in surveying sensitive topics. Various indirect questioning methods have been developed to reduce SDB, one of them being Warner's RRT. When respondents' privacy are guaranteed, their tendency to provide untruthful answers decrease. Warner's RRT achieves this by utilizing a randomization device. However, the penalty for using the randomization device is the inflated theoretical variance. We showed that, although the theoretical variance of the Warner's estimator is larger than the one from the DQT, when one takes into account the cheating proportion among the participants, the empirical variance from the Warner's RRT can be obtained smaller, especially for large samples. As an example, a study that was performed at LSUHSC where a variation of the Warner's RRT was implemented in the National HIV Behavioral Surveillance System (NHBS) in collaboration with the Louisiana Office of Public Health STD/HIV Program found smaller variance of the proportion estimate for the question "If there was a violent crime in your neighborhood and you knew who did it, would you report the person to authorities?[1]"

References

1. Leonardi C. (2014) A Two-stage Randomized Response Technique For Surveying Sensitive Topics, Unpublished PhD Dissertation, LSUHSC, New Orleans, LA.
2. Oral E. (2019) Surveying Sensitive Topics with Indirect Questioning, in: Statistical Methodologies. IntechOpen, doi: 10.5772/intechopen.84524.
3. Warner S.L. (1965) Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *JASA* 60, 63–69.