



Estimation of a Finite Population Proportion in Light of Randomized Reporting

Pulakesh Maiti¹, Jyotirmoy Sarkar², and Bikas K. Sinha^{3,*}

¹Economic Research Unit, Indian Statistical Institute, 203, Barrackpore Trunk Road, Kolkata 700108, India

²Mathematical Sciences, Indiana University-Purdue University Indianapolis, 402 N Blackford Street, Indianapolis, Indiana 46202, USA

³Retired Professor of Statistics, Indian Statistical Institute, 203, Barrackpore Trunk Road, Kolkata 700108, India

ARTICLE INFO

Article History

Received 01 Jan 2017

Accepted 18 Jul 2017

Keywords

Sampling design
SRSWOR

Investigator intervention
Supervisor intervention
Study design
Probability model
Randomization distribution.

2000 Mathematics Subject Classifications

62D05, 94A20

ABSTRACT

In large scale surveys, it is customary to accept *unaltered* the responses provided by the respondents, leaving no provision for investigators to pass on any circumstantial judgment on the responses. This restrictive practice sometimes vitiates the estimation of a parameter. Here we consider a real life scenario in which the data gathering process is compounded by possible interventions by investigators/supervisors who may provide thoughtful judgments on the quality/category of the response given by a respondent. In this modified scenario, we address the problem of estimating the unknown population proportion P . In the context of an illustrative example, we develop a possible randomization theory and computational formulas to estimate P under intervention effects.

© 2018 The Authors. Published by Atlantis Press SARL.

This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

1. THE SURVEY DESIGN AND AN INTERVENTION-BASED PROBABILITY MODEL

We consider a simple set-up of an SRSWOR (N, n) sampling design in conjunction with equal-sized clustered population units. The problem is to estimate P , the unknown population proportion of a qualitative feature ‘A’ in terms of its incidence [Yes/No] on each individual in the population. Naturally, the sample incidence proportion p suffices for the task. However, in our study design we encounter a different situation: The investigators, and also the supervisors, are likely to ‘intervene and distort’ the responses received from the sampled units. How should we estimate P in this modified situation of randomized reporting?

Revisiting the example described in Maiti and Sinha [1], which studies a quantitative variable, in this paper we allow investigators and/or supervisors to affect the response profiles of a qualitative variable. In other words, we keep a provision for wise and thoughtful circumstantial intervention effects on the quality/category of responses given by the respondents, before the data are finally received by the data collection agency.

Let the population consist of N Clustered Respondent Units (CRUs), each of which is composed of a fixed number m of individual respondents. These clusters are already formed before the sampling operation takes place. The investigators/supervisors are supposed to provide information on each of the n sampled CRUs in the form of subtotals of the original responses from all m individuals within each sampled CRU. The i^{th} sampled CRU will be denoted by C_i , for $i = 1, 2, \dots, n$. Let $L_{[i]}$ denote the number of schedule-based subtotals collected on C_i by various selected combinations of investigators $j = 1, 2, \dots, J$ and supervisors $k = 1, 2, \dots, K$, specified by the study design. That is,

$$L_{[i]} = \sum_k \sum_j I[i; (j, k)],$$

*Corresponding author. Email: bikassinha1946@gmail.com

Table 1 | Incidence matrix of investigator and supervisor assigned to the 7 selected CRUs.

Supervisor	k	1	1	1	1,2	2	2	2
Investigator	j	1	2	3	4	5	6	7
CRU	C1	1	1	0	1	0	0	0
	C2	0	1	1	0	1	0	0
	C3	0	0	1	1	0	1	0
	C4	0	0	0	1	1	0	1
	C5	1	0	0	0	1	1	0
	C6	0	1	0	0	0	1	1
	C7	1	0	1	0	0	0	1

where $I[i; (j, k)] = 1$ if Investigator j and Supervisor k have both worked on a schedule assigned to CRU C_i ; otherwise, $I[i; (j, k)] = 0$. Of course, $L_{[i]} \geq 1$ for each responding C_i . Whenever $I[i; (j, k)] = 1$, we will denote by $F_{[i; (j, k)]}$ the underlying response (the subtotal) on the qualitative feature ‘A’ for C_i , reported by the (j, k) -combination of investigator and supervisor.

Note that the same CRU may be surveyed by multiple investigators and/or multiple supervisors; that is, $L_{[i]}$ may exceed 1. This happens, for example, when several investigators visit the same CRU to collect information on different parts of a large-scale survey, but there is smaller set of questions (including the question on the presence or absence of the qualitative feature ‘A’) that are included on each investigator’s schedule. In case there is no intervention effect of the investigators/supervisors, all $L_{[i]}$ observations on CRU C_i will be identical for each sampled CRU. On the other hand, if such interventions are anticipated, these observations can potentially differ. Also, the supervisor must be blinded from the CRU label, so that he/she can genuinely assess the subtotals provided by different investigators on the same CRU and possibly provide independent thoughtful modifications on each of them.

As in Maiti and Sinha [1], let us consider the following simple example of a study design: The population consists of $N = 70$ CRUs, each composed of $m = 10$ individuals. A sample of $n = 7$ CRUs is taken using a fixed size sampling design [say, for example, SRSWOR(70, 7)]. Let there be $J = 7$ investigators and $K = 2$ supervisors engaged in the process of data collection. We name the selected CRUs as C_1, \dots, C_7 . Next, we assign investigators (considered as treatments) to gather information from some selected CRUs (considered as blocks) by utilizing a symmetric BIBD(7, 7, 3, 3, 1), developed from the initial set (1, 2, 4) following Bose’s technique. See Raghavarao [2]. Additionally, Investigators 1–3 are supervised by Supervisor 1, Investigators 5–7 are supervised by Supervisor 2, and Investigator 4 is supervised by both Supervisors 1 and 2. Table 1 gives the complete details of the study design.

In essence, the allocation design of Table 1 specifies that the first CRU C_1 will be approached by Investigators 1 and 2, and their response profiles will be checked by Supervisor 1. Further, the same CRU C_1 will also be approached by Investigator 4 and the response profile will be checked by both Supervisors 1 and 2. Similar explanations apply to the other selected CRUs as well. Since the CRU sizes are all equal ($m = 10$), we will ignore the CRU size effect and treat each one as a singleton.

Let us denote by $F_{[1]}, F_{[2]}, \dots, F_{[7]}$ the ‘data’ on the selected CRUs accrued from the field. Without any intervention effect from the investigators/supervisors, we would have regarded such data as ‘error-free,’ and so standard estimation techniques could be routinely used. Note that in such an ‘error-free’ scenario, there is no difference among $F_{[1;(1,1)]}, F_{[1;(2,1)]}, F_{[1;(4,1)]}$ and $F_{[1;(4,2)]}$, for example. However, we want to examine the possibility of intervention effect(s) from one or the other group or possibly from both groups. So we adopt an intervention probability distribution which is fully described in Section 2.

2. RANDOMIZATION DISTRIBUTIONS

To fix ideas, as mentioned earlier, let each CRU be composed of $m = 10$ households (HHs). The response on each selected CRU is taken to be the sum of the responses of the 10 HHs within the CRU. Since we are dealing with a binary response to the qualitative feature ‘A’, each such subtotal $F_{[(i;j,k)]}$ will take on a count value in the range $\{0, 1, \dots, 10\}$. Therefore, an intervention would amount to either keeping the response count intact or ‘twisting’ the response count. This will apply at the investigator level as well as at the supervisor level. More specifically, an investigator takes the true score T from the field and after intervention produces a revised score R ; and a supervisor takes this revised score R and after intervention produces a final declared score D . Also, we assume that for each supervisor, the modification (through intervention, if any) will be independent of whichever investigator forwards the response to him/her; that is, there is no interaction effect between investigator and supervisor.

For simplicity, for each of Investigators 1–4, we assume the same randomization probability distribution given by the stochastic diagram in Fig. 1(a), where the label of each node represents the true response T from the field, and the randomized revised score R is the label of either that node itself (with probability written inside that node), or of an adjacent node (with probability written on the outgoing arc). For Investigators 5–7, we assume a different randomization probability distribution as depicted in Fig. 1(b). For Supervisor 1, the randomization distribution is shown in Fig. 1(c), where the node label represents the revised score R given to her by the investigator, and the final declared score D is the label of either that node (with probability written inside that node), or of an adjacent node (with probability written on the outgoing arc); while for Supervisor 2, the randomization distribution is the same as in Fig. 1(a).

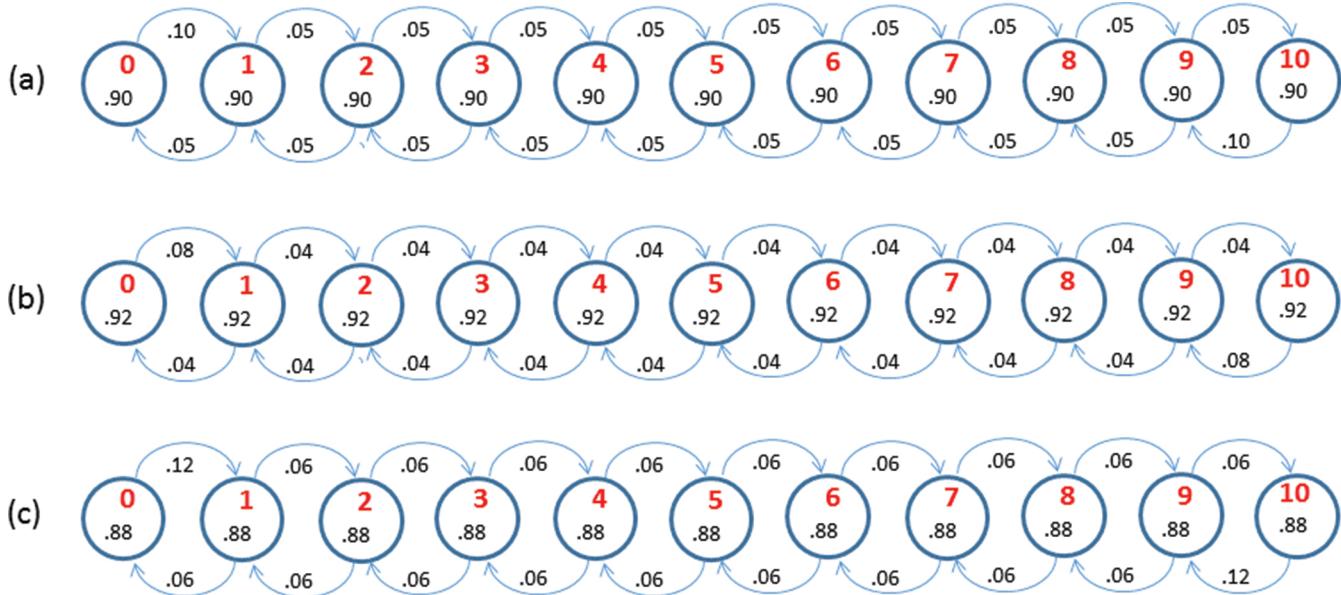


Figure 1 | Stochastic diagrams of randomization probability distributions for (a) Investigators 1–4, and Supervisor 2; (b) Investigators 5–7; and (c) Supervisor 1.

Instead of Fig. 1, equivalently the randomization probability distribution is also given by an 11×11 stochastic matrix $Q(q)$, whose rows and columns are labeled as $0, 1, \dots, 10$, and where q denotes the retention probability at each node. That is,

$$Q(q) = \begin{bmatrix} q & 1-q & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 \\ \frac{1-q}{2} & q & \frac{1-q}{2} & 0 & \dots & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1-q}{2} & q & \frac{1-q}{2} & \dots & 0 & 0 & 0 & 0 & 0 \\ \dots & \dots \\ \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & \frac{1-q}{2} & q & \frac{1-q}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & \frac{1-q}{2} & q & \frac{1-q}{2} & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1-q & q & 0 \end{bmatrix}.$$

In particular, Fig. 1(a)–(c) are equivalent to matrices $Q(.9)$, $Q(.92)$, $Q(.88)$, respectively.

3. DATA GATHERING MECHANISM

The statistics software R has been used throughout this paper for data generation, computations, data analysis and simulation.

We have denoted by P the proportion of HHs in the whole population possessing the qualitative feature ‘A’. Our purpose is to estimate this parameter based on the data declared by the supervisors. In case of no-intervention effect, we would have gathered ‘true scores’ from the $n = 7$ selected CRUs, each comprising of $m = 10$ HHs. Assuming independent behavior of the HHs with respect to possession of the attribute ‘A’, each true score obtained from a selected CRU behaves like an observation from a binomial $(10, P)$ distribution. Therefore, the sample mean of 7 scores divided by 10 (or equivalently, the sum of 7 scores divided by 70) would provide a valid estimate for P . How should estimation proceed under the intervention effect(s)?

Let us explain the entire data generating process: We set $P = 0.2$, and generate 70 Bernoulli observations, in sets of 10, shown in Table 2. These are the true responses of the HHs. The i^{th} row sum is the true subtotal $F_{[i]}$ from the field corresponding to CRU C_i .

Table 2 | True responses of 70 households in sets of 10, and true scores (subtotals) of the 7 selected CRUs.

Households	1	2	3	4	5	6	7	8	9	10	True score
CRU	C1	0	1	1	0	1	0	0	0	0	3
	C2	0	0	0	0	1	0	0	0	1	2
	C3	0	0	0	0	1	1	0	0	0	2
	C4	0	0	0	0	0	0	0	0	0	0
	C5	0	0	0	1	0	0	0	0	0	1
	C6	0	0	0	0	0	1	0	0	1	2
	C7	0	0	0	1	0	1	0	0	0	2
Sum											12

Table 3 | Calculating the randomization distribution of the declared score of CRU 1 after interventions by Investigator 1 and Supervisor 1.

CRU 1 has TR = 3	1	2	3	4	5	1	2	3	4	5		
j = 1 reports	w.p.	Cond. prob. of RR by k = 1					Joint prob. of RR by (j = 1, k = 1)					
2	.05	.06	.88	.06	0	0	.003	.044	.003	0	0	
3	.90	0	.06	.88	.06	0	0	.054	.792	.054	0	
4	.05	0	0	.06	.88	.06	0	0	.003	.044	.003	
Column sum						.003	.098	.798	.098	.003	Distribution of declared score	

Under a direct response scenario without any distortion from any external source, the investigator-cum-supervisor allocation design has no role whatsoever on the estimation of P . From the raw data of Table 2, we obtain:

$$\hat{P}_1 = \frac{1}{mn} \sum_i F_{[i]} = \frac{12}{70} = 0.1714. \quad (1)$$

Also, routine computation yields a standard error (SE) of the estimate as

$$SE_1 = \sqrt{\left(\frac{1}{n} - \frac{1}{N}\right) \frac{\hat{P}_1(1 - \hat{P}_1)}{m}} = 0.0427. \quad (2)$$

On the other hand, if there is/are intervention(s) by investigator/supervisor, in order to explain the effect of the distortion mechanism on the estimation of P , let us carefully review the generation of ‘declared’ scores according to the allocation design (of investigators and supervisors to CRUs) given in Section 1, and the randomization probability distributions adopted by them described in Section 2.

For the CRU C_1 , the investigator-supervisor combinations are: $(j, k) = (1, 1), (2, 1), (4, 1), (4, 2)$. Let us explain how the declared score distribution arises out of the $(j, k) = (1, 1)$ combination, given that the true score is $T = 3$: First, construct a matrix for which the rows represent the randomized revised values {2, 3, 4} by Investigator 1, and the columns represent the randomized declared values {1, 2, 3, 4, 5} by Supervisor 1. The (u, v) -th element of the matrix is the joint probability obtained by multiplying the randomization probability that Investigator 1 reports a value u given $T = 3$, and the conditional randomization probability that Supervisor 1 reports a value v given that Investigator 1 reported u . Then the randomization distribution of ‘declared’ score is simply the column sum of the matrix of joint distribution. The details of the calculations are shown in Table 3.

The calculations in Table 3 show that the declared score can take on values (1, 2, 3, 4, 5) with probabilities (.003, .098, .798, .098, .003).

Alternatively, a more efficient algorithm to construct this randomization distribution of the declared score (ranging from 0 to 10) is to compute the product matrix $Q(.90) Q(.88)$, and then simply take the row labeled 3 (since $T = 3$), which is

$$(0.0000, 0.0030, 0.0980, 0.7980, 0.0980, 0.0030, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000).$$

We generate a random number according to this randomization distribution, and obtain a declared score of ‘3’.

For the CRU C_1 , with true score $T = 3$, we must construct the declared score distributions corresponding to three more investigator-supervisor combinations. For $(j, k) = (2, 1), (4, 1)$ combinations, the same conditional distribution holds as for the $(j, k) = (1, 1)$ combination just shown above; and we generate declared scores ‘3’ and ‘3’ in these two cases. Lastly, for the $(j, k) = (4, 2)$ combination, we have a different declared score distribution given by the row labeled 3 of the product matrix $Q(.90) Q(.90)$, which is

$$(0.0000, 0.0025, 0.0900, 0.8150, 0.0900, 0.0025, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000).$$

Table 4 | The 24 declared scores produced by investigator-supervisor combinations.

Supervisor	k	1	1	1	1,2	2	2	2	True
Investigator	j	1	2	3	4	5	6	7	Score
CRU	C1	3	3	-	3,2	-	-	-	3
	C2	-	3	2	-	2	-	-	2
	C3	-	-	2	2,3	-	2	-	2
	C4	-	-	-	1,0	1	-	0	0
	C5	1	-	-	-	1	1	-	1
	C6	-	2	-	-	-	1	1	1
	C7	3	-	3	-	-	-	2	2

We generate a random number according to this randomization distribution and obtain a declared score of '2'. Thus, for CRU C_1 , while the true score is 3, the intervention-based scores are (3, 3, 3, 2) corresponding to the four investigator-supervisor combinations reporting on CRU C_1 .

Next, for the CRU C_2 , the true score is $T = 2$. According to the design plan, the possible intervention pairs are $(j, k) = (2, 1), (3, 1), (5, 2)$. The declared score distributions arising out of $(j, k) = (2, 1), (3, 1)$ pairs are exactly the same; namely,

$$(0.0030, .0980, .7980, .0980, .0030, .0000, .0000, .0000, .0000, .0000).$$

We draw random numbers from this randomization distribution and derive a score of '3' for $(j, k) = (2, 1)$, and '2' for $(j, k) = (3, 1)$. Lastly, for $(j, k) = (5, 2)$, the declared score distribution is given by

$$(0.0016, .0736, .8496, .0736, .0016, .0000, .0000, .0000, .0000, .0000).$$

We generate a random number according to this randomization distribution, and obtain a declared score of '2'. Thus, for CRU C_2 , while the actual score is 2, the three intervention-based scores are (3, 2, 2).

In this manner we obtain the randomization distributions and obtain declared scores for the remaining CRUs for each investigator-supervisor combination mentioned in the study design. These are summarized in Table 4.

4. DATA ANALYSIS

We have explained the generation of declared scores in all 24 cases of investigator-supervisor interventions based on the adopted study design. The frequency $(f)(x)[x, f] = 0, 2, 1, 5, 2, 10, 3, 7$. Computation yields a mean score of $\sum fx / \sum f = 46/24 = 1.9167$, which provides an estimate of P as $\hat{P}_2 = 0.19167$, which is very close to the true value of $P = 0.20$. The associated SE, if computed by the usual formula assuming that all the 24 estimates are independent, becomes $SE_2 = 0.0189$. One may be tempted to conclude that estimator \hat{P}_2 is quite sound, and we are able to provide an estimate of P with utmost accuracy based on this intervention-based study design. However, there is a sharp criticism regarding the validity of this approach. The 'scores' arising out of 'within CRU intervention sessions' are heavily dependent; and hence simple averaging of 24 scores is not a valid estimate.

We may, however, adopt a two-stage estimation procedure. In the first stage, we estimate the 'true' scores for the CRUs in the presence of intervention mechanism. These are essentially simple means of all declared scores within each CRU. In the second stage, we compute the simple average of these first stage estimates, and hence obtain estimator \hat{P}_3 of P . That is,

$$\hat{P}_3 = \frac{1}{n} \sum_i \bar{F}_{[i]}; \quad \text{where} \quad \bar{F}_{[i]} = \frac{1}{L_{[i]}} \sum_j \sum_k F_{[i(j,k)]}. \quad (3)$$

Alternatively, the first-stage estimates of true scores for the CRUs may be calculated as the mean declared score across the investigators—that is, first we replace the two declared scores reported by Supervisors 1 and 2 based on the revised score produced by Investigator 4, by their mean, and we consider this mean as the single score produced by Investigator 4; then we take the mean of the scores across the three investigators reporting on each CRU. Then in the second stage we compute the simple average of these revised first-stage means to obtain estimator \hat{P}_4 of P . That is,

$$\hat{P}_4 = \frac{1}{7} \sum_i \bar{\bar{F}}_{[i]}; \quad \text{where} \quad \bar{\bar{F}}_{[i]} = \frac{1}{3} \sum_j F_{[i(j,:)];} \quad (4)$$

$$\text{with } F_{[i(4,:)]} = \{F_{[i(4,1)]} + F_{[i(4,2)]}\}/2 \quad \text{and} \quad F_{[i(j,:)]} = F_{[i(j,k)]} \quad \text{for } j \neq 4.$$

Applying these two two-stage approaches, we compute the first-stage means of all declared scores and the first-stage means of investigator scores as shown in Table 5.

Table 5 | Two ways to calculate the first stage means within each CRU.

Supervisor	k	1	1	1	1,2	2	2	2	True	Mean	Mean
Investigator	j	1	2	3	4	5	6	7	Score	All DS	By inv.
CRU	C1	3	3	-	3,2	-	-	-	3	2.75	2.83
	C2	-	3	2	-	2	-	-	2	2.33	2.33
	C3	-	-	2	2,3	-	2	-	2	2.25	2.17
	C4	-	-	-	1,0	1	-	0	0	0.50	0.50
	C5	1	-	-	-	1	1	-	1	1.00	1.00
	C6	-	2	-	-	-	1	1	1	2.00	2.00
	C7	3	-	3	-	-	-	2	2	2.67	2.67

In the second stage, we compute the simple average of these first stage estimates of true CRU scores to obtain $\hat{P}_3 = 0.1928$ and also $\hat{P}_4 = 0.1928$. Since the CRUs are selected according to SRSWOR sampling, the simple averaging of CRU-based estimates in the second stage produces a valid point estimate of P . In fact, for our particular data set, the two estimates happen to be exactly the same. The sampling distributions of \hat{P}_3 and \hat{P}_4 are comparable, though the former seem to have a smaller SE while the later seem to have a narrower central 95% coverage region, as we shall see shortly.

Actually, the computation of the associated SE for \hat{P}_3 or \hat{P}_4 does not seem to be easy. We may argue that since only the calculation of CRU subtotals have changed due to interventions, as a first approximation it may suffice to simply use the same SE formula for \hat{P}_1 , given in Eq. (2), which we use in case of no-intervention. That is, a first approximation of SE of \hat{P}_3 or \hat{P}_4 is about $\sqrt{(1/n - 1/N) \hat{P}_3 (1 - \hat{P}_3) / m} = .04473$. This crude computation of SE indicates that there is not much loss of precision due to investigator/supervisor intervention(s). A more refined SE formula can be derived through probability theory by incorporating the effects of all possible randomization distributions on the true score of a typical CRU. We leave the details to the reader.

In the absence of any analytical expressions for the SE of \hat{P}_3 or \hat{P}_4 , we will take recourse to a (parametric) bootstrap computation. See Efron and Tibshirani [3] or Davison and Hinkley [4], for a justification of the bootstrap method. The bootstrap set up is described as follows: Assume the true proportion of HHs exhibiting the qualitative feature 'A' is $P_0 = \hat{P}$. In each bootstrap iteration, we generate the true responses (the number of HHs exhibiting feature 'A' out of the 10 HHs in the cluster) from the 7 sampled CRUs using a binomial ($10, P_0$) distribution. Then for each investigator-supervisor combination (j, k) assigned to a CRU receiving true score T , we obtain the declared score according to the randomization distribution given by the row labeled T of the product matrix $Q(q_j) Q(q_k)$. Using these declared scores for all 24 investigator-supervisor combinations, we compute \hat{P}_3 as the simple average of first-stage means of declared scores, and we compute \hat{P}_4 as the simple average of three investigators' scores (assuming that Investigator 4's score is given by the average of declared scores by the two supervisors), within each sampled CRU. We repeat the process $M = 10^4$ times (to ensure that the mean of the sampling distribution is close to P_0 , correct to at least two decimal places). Then the SE of \hat{P}_3 or \hat{P}_4 is obtained as the standard deviation of the bootstrap distribution of \hat{P}_3 or \hat{P}_4 . Also, we construct a 95% confidence interval for P based on the 2.5-th and the 97.5-th percentiles of the bootstrap distributions. The results are summarized below

$$P_0 = .1928, \text{ mean } (\hat{P}_3) = .1947, \text{ SE } (\hat{P}_3) = .04708, 95\% \text{ CI of } P = (.1083, .2917).$$

$$P_0 = .1928, \text{ mean } (\hat{P}_4) = .1947, \text{ SE } (\hat{P}_4) = .04712, 95\% \text{ CI of } P = (.1095, .2905).$$

We only exhibit the bootstrap distribution of \hat{P}_4 in Fig. 2. The bootstrap distribution of \hat{P}_3 is very similar. These bootstrap distributions suggest that these estimators are slightly positively skewed. We can justify it as follows: A revised score R due to intervention by investigator j behaves as $R = T + Y$, where T is the true score from the binomial ($10, P_0$) distribution, and Y is a trinomial variable taking on values $-1, 0, 1$ with probabilities $(1 - q)/2, q, (1 - q)/2$ respectively when $1 \leq T \leq 9$, a binomial taking values 0, 1 with probabilities $q, 1 - q$ when $T = 0$, and a binomial taking values $-1, 0$ with probabilities $1 - q, q$ when $T = 10$. Hence, a revised score has a mean of

$$mP_0 + [(1 - P_0)^m - P_0^m] (1 - q_j).$$

Carrying on a similar reasoning, a declared score D due to interventions by investigator j and supervisor k has a mean of

$$\begin{aligned} & mP_0 + [(1 - P_0)^m - P_0^m] [1 - q_j + q_j (1 - q_k)] \\ & + m [(1 - P_0)^{m-1} P_0 - P_0^{m-1} (1 - P_0)] (1 - q_j) (1 - q_k) / 4, \end{aligned} \quad (5)$$

which exceeds mP_0 , resulting in a positive skewness. If we evaluate Eq. (5) for all investigator-supervisor combinations and then average them according to the two-stage procedure of calculating \hat{P}_3 and \hat{P}_4 , then the means of the declared scores are seen to be 1.9493 and

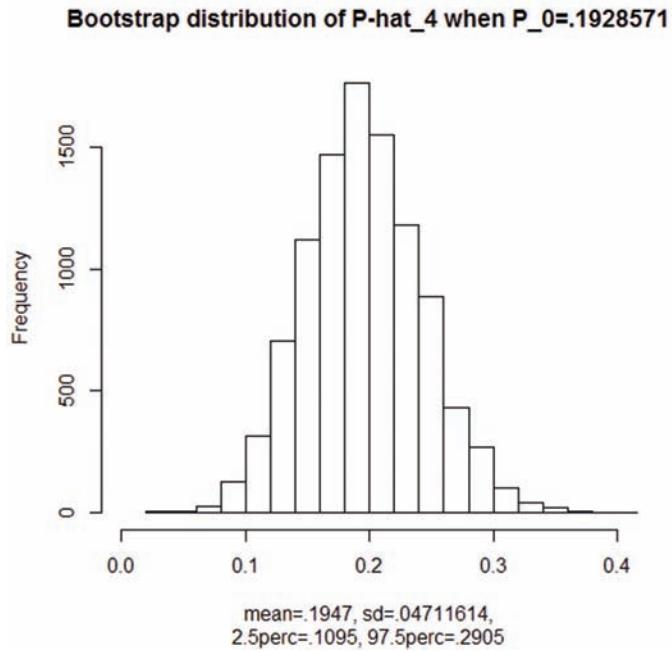


Figure 2 | Bootstrap distribution of estimator iterations.

1.9517 respectively. Thus, the bias of estimator \hat{P}_3 in estimating P is about .0021, and that of estimator \hat{P}_4 is about .0023. Indeed, the bootstrap distributions of \hat{P}_3 and \hat{P}_4 have the same mean $0.1947 = .1928 + .0019$. Also, from the bootstrap distributions we estimate that the investigator-supervisor interventions result in only about 5.2% increase in the SE of \hat{P}_3 or about 5.3% increase in the SE of \hat{P}_4 , compared against the SE of 0.04473 under the no-intervention scenario. While \hat{P}_3 seem to have a slightly smaller SE than \hat{P}_4 , a 95% confidence interval for P based on the bootstrap distribution of \hat{P}_4 is slightly narrower than that based on \hat{P}_3 . For all practical purposes estimators \hat{P}_3 and \hat{P}_4 have comparable behaviors.

5. POSTERIOR ANALYSIS

We are advocating for allowing wise and thoughtful interventions by investigators and/or supervisors. Even then, in this section we are interested in trying to recover the original scores received in the field, given the declared scores. This posterior analysis merits attention in its own right. Moreover, it demonstrates that the estimators \hat{P}_3 and \hat{P}_4 are closer to the true proportion P than the estimator \hat{P}_1 computed from ‘predicted’ true scores based on posterior analysis.

Once we have the scores declared by different investigator-supervisor combinations, we can work out the posterior probability distribution of the true score, using Bayes’ formula. We obtain the posterior probability distribution of the true score T corresponding to a declared score D , reported by investigator-supervisor combination (j, k) , by implementing the following three steps: (1) We choose the column labeled D of the product matrix $Q(q_j) Q(q_k)$. Call this vector l . (2) We multiply this column vector l coordinate-wise by the prior probability vector π_0 , given by the binomial $(10, P_0)$ distribution. (3) We standardize the product vector by dividing each element by the sum of all the elements, to obtain the posterior probability vector π_1 . For example, for CRU C_1 , based on interventions by Investigator 1 and Supervisor 1, the declared score is ‘3’. Following the above three steps (and writing $Q(.90) Q(.88) \delta_4 = l_4$) we obtain

$$\begin{aligned}\pi_0 &= (.1174, .2804, .3015, .1921, .0803, .0230, .0046, .0006, .0001, .0000, .0000); \\ l_4 &= (.0000, .0030, .0980, .7980, .0980, .0030, .0000, .0000, .0000, .0000, .0000); \\ \pi_0 * l_4 &= (.0000, .0008, .0296, .1533, .0079, .0001, .0000, .0000, .0000, .0000, .0000); \\ \pi_1 &= (.0000, .0044, .1542, .8000, .0411, .0004, .0000, .0000, .0000, .0000, .0000).\end{aligned}$$

The same posterior distribution also applies to interventions by (Investigator 2, Supervisor 1) and (Investigator 4, Supervisor 1) combinations. For intervention by (Investigator 4, Supervisor 2), with a declared score of ‘2’, (and writing $Q(.90) Q(.90) \delta_3 = l_3$) we have

$$\begin{aligned}\pi_0 &= (.1174, .2804, .3015, .1921, .0803, .0230, .0046, .0006, .0001, .0000, .0000); \\ l_3 &= (.0050, .0900, .8150, .0900, .0025, .0000, .0000, .0000, .0000, .0000, .0000); \\ \pi_0 * l_3 &= (.0006, .0252, .2457, .0173, .0002, .0000, .0000, .0000, .0000, .0000, .0000); \\ \pi_1 &= (.0020, .0873, .8501, .0598, .0007, .0000, .0000, .0000, .0000, .0000, .0000).\end{aligned}$$

Table 6 | Posterior distributions of the true scores (and their averages), given the declared scores.

CRU	True scores												Posterior Mode		
	j	k	DS	0	1	2	3	4	5	6	7	8	9	10	
1	1	1	3	0	.0044	.1542	.8000	.0411	.0004	0	0	0	0	0	3
	2	1	3	0	.0044	.1542	.8000	.0411	.0004	0	0	0	0	0	3
	4	1	3	0	.0044	.1542	.8000	.0411	.0004	0	0	0	0	0	3
	4	2	2	.0020	.0873	.8501	.0598	.0007	0	0	0	0	0	0	2
	Average		2.75	.0005	.0251	.3282	.6150	.0310	.0003	0	0	0	0	0	
2	2	1	3	0	.0044	.1542	.8000	.0411	.0004	0	0	0	0	0	3
	3	1	2	.0024	.0955	.8358	.0654	.0008	0	0	0	0	0	0	2
	5	2	2	.0016	.0792	.8643	.0543	.0006	0	0	0	0	0	0	2
	Average		2.33	.0013	.0597	.6181	.3066	.0142	.0001	0	0	0	0	0	
3	3	1	2	.0024	.0955	.8358	.0654	.0008	0	0	0	0	0	0	2
	4	1	2	.0024	.0955	.8358	.0654	.0008	0	0	0	0	0	0	2
	4	2	3	0	.0037	.1416	.8168	.0377	.0003	0	0	0	0	0	3
	6	2	2	.0016	.0792	.8643	.0543	.0006	0	0	0	0	0	0	2
	Average		2.25	.0016	.0685	.6694	.2505	.0100	.0001	0	0	0	0	0	
4	4	1	1	.0828	.8087	.1064	.0021	0	0	0	0	0	0	0	1
	4	2	0	.7863	.2075	.0062	0	0	0	0	0	0	0	0	0
	5	2	1	.0692	.8406	.0889	.0014	0	0	0	0	0	0	0	1
	7	2	0	.8054	.1897	.0050	0	0	0	0	0	0	0	0	0
	Average		0.50	.4359	.5116	.0516	.0009	0	0	0	0	0	0	0	
5	1	1	1	.0828	.8087	.1064	.0021	0	0	0	0	0	0	0	1
	5	2	1	.0692	.8406	.0889	.0014	0	0	0	0	0	0	0	1
	6	2	1	.0692	.8406	.0889	.0014	0	0	0	0	0	0	0	1
	Average		1.00	.0737	.8300	.0947	.0016	0	0	0	0	0	0	0	
6	2	1	2	.0024	.0955	.8358	.0654	.0008	0	0	0	0	0	0	2
	6	2	2	.0016	.0792	.8643	.0543	.0006	0	0	0	0	0	0	2
	7	2	2	.0016	.0792	.8643	.0543	.0006	0	0	0	0	0	0	2
	Average		2.00	.0019	.0846	.8548	.0580	.0007	0	0	0	0	0	0	
7	1	1	3	0	.0044	.1542	.8000	.0411	.0004	0	0	0	0	0	3
	3	1	3	0	.0044	.1542	.8000	.0411	.0004	0	0	0	0	0	3
	7	2	2	.0016	.0792	.8643	.0543	.0006	0	0	0	0	0	0	2
	Average		2.67	.0005	.0293	.3909	.5514	.0276	.0003	0	0	0	0	0	

In this manner we obtain the posterior distribution of the true score T for each declared score D after investigator-supervisor interventions, and tabulate them in Table 6. We also compute the within CRU average posterior distribution by taking a simple average of the posterior distributions corresponding to the three or four investigator-supervisor combinations assigned to each CRU. Thereafter, we take the mode of the average posterior distribution as the ‘predicted’ true score of each CRU.

One might wonder about the strong resemblance (more specifically, exact identity) between the ‘predicted’ value given by the posterior mode and the ‘true score’ for each CRU. A natural justification lies in the fact that the ‘randomization probability distributions’ are highly peaked (around the respective true scores) for each investigator and each supervisor. It may be remarked that the randomization theory works perfectly well with any other choice of the randomized distributions. For example, we could lower the retention probability q (from .9 down to .5) for the investigator and/or the supervisor. This would lead to an increased variation in reporting by the investigator and/or supervisor. Consequently, the strong resemblance between the posterior mode and the true score could be reduced to some extent.

Having obtained the predicted true score of each CRU given by the mode of the average posterior distribution within each CRU, we can go back to estimator \hat{P}_1 , but this time calculated based on the predicted true scores. Let us call it \hat{P}_5 . Since the predicted true scores are exactly the same as the original true score, we have $\hat{P}_5 = \hat{P}_1 = .1714$, and its SE is .04273. Note that the reduction in SE of \hat{P}_5 , compared to that of \hat{P}_3 or \hat{P}_4 , is not much. On the other hand, the point estimate \hat{P}_5 is further away from the true value $P = .2$. Thus, this posterior analysis strongly supports the proposal for allowing interventions by investigator/supervisor.

6. CONCLUDING REMARKS

We presented an illustrative example which exhibits the benefits of allowing wise and thoughtful interventions by investigators and/or supervisors. It is advisable that the investigators, who are directly in conversation/contact with the respondents, should apply their strong feelings and wise thoughts, to modify the responses received in the field. They should not merely write down the responses, for that may not be conducive to the case being studied. With their educated interventions, the quality of estimation will improve (measured in terms of

low bias), while the precision of estimation (measured by the reciprocal of the SE of the point estimate) may decline only by a small amount. The example we presented here extends in the following two directions to cover increasingly more realistic situations:

1. We can change the sampling design for selecting the CRUs from SRSWOR scheme to some other unequal probability sampling scheme. For example, using any fixed sample size sampling design ($N = 70, n = 7$) with positive first and second order inclusion probabilities of the CRUs, we could apply Horvitz-Thompson estimator based on one of the first-stage estimates of individual CRU-based proportions. See Table 5 of Section 4.
2. We can change the transition matrices used to model the intervention of the interviewers and supervisors from having the very simple form of $Q(q)$ into being arbitrary stochastic matrices. We can even allow a unique stochastic matrix to model the intervention of each investigator and each supervisor. The computations extend in a straight-forward manner under such generalized scheme of randomization.

Analytical derivations of the sampling distributions of estimators \hat{P}_3 and \hat{P}_4 in general, and their SEs in particular, remain open problems.

ACKNOWLEDGMENTS

The authors sincerely thank a referee for pointing out one computational error in an earlier version of the manuscript and for indicating several places where more detailed explanations were needed. The second author thanks the authorities at the Indian Statistical Institute, Kolkata, for hosting his visit during which time the research was concluded. The editor's encouragement for carrying out the needed revision is also highly appreciated.

REFERENCES

1. P. Maiti, B.K. Sinha. *J. Stat. Theory Appl.* 13. 3. 2014. 263–272.
2. D. Raghavarao, *Constructions and Combinatorial Problems in Design of Experiments*, Wiley, New York, 1971.
3. B. Efron, R. Tibshirani. *An Introduction to the Bootstrap: Monographs on Statistics & Applied Probability*. Chapman & Hall/CRC Press, Boca Raton, FL, 1993.
4. A.C. Davison, D.V. Hinkley, *Bootstrap Methods and Their Application*, Cambridge University Press, New York, 1997.
5. A.S. Hedayat, Bikas K. Sinha, *Design and Inference in Finite Population Sampling*, Wiley, New York, 1991.
6. R Core Team. *A Language and Environment for Statistical Computing*. Vienna, Austria 2013. R Foundation for Statistical Computing.