Device 2. A sampled person, labeled, say, i is approached with a box containing a large number of cards of which a proportion $C (0 < C < 1)$ is marked 'yes' and proportions $q_1, \cdots, q_M (0 < q_j < 1, \sum_1^M q_j = 1 - C)$ are marked $x_1, \cdots, x_j, \cdots, x_M$ and on request is to draw a card at random and report a value z_j as a response such that

$$\begin{aligned} z_j &= y_i \text{ if a 'yes' marked card is drawn} \\ &= x_j \text{ if an '}x_j\text{'-marked card is drawn.} \end{aligned}$$

Then
$$E_R(z_j) = Cy_i + \sum_1^N q_j x_j, \text{ yielding}$$

$$r_i = \frac{1}{C}\left(z_j - \sum_{j=1}^M q_j x_j\right) \text{ for which}$$

$$E_R(r_i) = y_i \text{ and } V_i = V_R(r_i) = \alpha y_i^2 + \beta y_i + \psi$$

where
$$\alpha = \frac{1}{C} - 1, \beta = -\frac{2}{C}\sum_1^M q_j x_j \text{ and}$$

$$\psi = \frac{1}{C^2}\left[\sum_{j=1}^M q_j x_j^2 - \left(\sum_1^M q_j x_j\right)^2\right]$$

and
$$\nu_i = \frac{\alpha r_i^2 + \beta r_i + \psi}{1 + \alpha} \text{ for which } E_R(v_i) = V_i.$$

So, unbiased estimation of Y along with suitably unbiased variance estimation promptly follows.

Chaudhuri and Christofides (2013) have presented results for RR Device I and RR Device II; the measures of protection are as follows:
Device I First box cards numbered $a_1, \cdots, a_M,\ \mu_a = \frac{1}{M}\sum_1^M a_j \neq 0$; second box cards numbered $b_1, \cdots, b_T,\ \mu_b = \frac{1}{T}\sum_1^T b_K$.

RR from i is $z_i = a_j y_i + b_K,\ E_R(z_i) = y_i\mu_a + \mu_b$.

With $L_i =$ Prior Prob$[y_i = 1]$, the posterior is

$$L(y_i|z_i) = \frac{L_i P(z_i|y_i)}{P(z_i)} = \frac{L_i(\frac{1}{TM})}{(\frac{1}{TM})} = L_i.$$

So, privacy is protected unless $T = 1$ and $M = 1$ for which choice, with z_i the value y_i will be immediately revealed.
For the RR **Device II**

$$\begin{aligned} z_i &= 1 \text{ if true } y_i \text{ is given out with probability } C(0 < C < 1) \\ &= x_j,\ j = 1, \cdots, M \text{ with probability } q_j(0 < q_j < 1, \sum_1^M q_j = 1 - C) \end{aligned}$$

then, $\quad E_R(z_i) = Cy_i + \sum_1^M q_j x_j.$

Then, $$L(y_i|z_i) = \frac{L(y_i)C}{L(y_i)C + (1 - L(y_i)\sum_1^M q_j} = \frac{L_i}{L_i\left(2 - \frac{1}{C}\right) + \left(\frac{1}{C} - 1\right)} \to L_i \text{ if } C \to \frac{1}{2}.$$

Chaudhuri, Christofides and Saha (2007) is also a relevant reference.

$M = 2$, for example, privacy is hardly compromised. Chaudhuri and Dihidar (2009) have a second device to cover quantitative response. In this device an investigator carries a box of cards, a proportion C $(0 < C < 1)$ marked blank and the remaining numbers $x_1, \cdots, x_M$ such that their respective proportions $q_j\,(j = 1, \cdots, M)$ are such that $0 < q_j < 1$ but $\sum_1^M q_j = (1 - C)$. Then, on request, from the ith person the forthcoming response is

$$\begin{aligned} z_i &= y_i \text{ if a blank is drawn} \\ &= x_j \text{ if an } x_j\text{-marked card is drawn.} \end{aligned}$$

Then, $$E_R(z_i) = Cy_i + \sum_{j=1}^{M} q_j x_j.$$

Letting $$r_i = \left(z_i - \sum_1^M q_j x_j\right) / C, i \in U,$$

$$E_R(r_i) = y_i \;\forall\; i \in U.$$

Also, $$V_R(z_i) = Cy_i^2 + \sum_1^M q_j x_j^2 - \left(Cy_i + \sum_1^M q_j x_j\right)^2.$$

So, $$V_i V_R(r_i) = \frac{1}{C^2} V_R(z_i) = \alpha y_i^2 + \beta y_i + \phi, \text{ say,}$$

with α, β, ϕ as known. Then,

$$v_i = \left(\alpha r_i^2 + \beta r_i + \phi\right) / (1 + \alpha) \quad \text{has} \quad E_R(v_i) = V_i.$$

Now paralleling the situations as in the earlier device we may get

$$\begin{aligned} L(y_i | z_i) &= \frac{L(y_i)\, C}{L(y_i)\, C + (1 - C)(1 - L(y_i))} \\ &= \frac{CL_i}{CL_i + (1 - C)(1 - L_i)} = \frac{CL_i}{L_i(2C - 1) + (1 - C)} \\ &= \frac{L_i}{L_i\left(2 - \frac{1}{C}\right) + \left(\frac{1}{C} - 1\right).} \end{aligned}$$

So, $L(y_i | z_i)$ matches L_i if $C = \frac{1}{2}$.

Taking $C = \frac{1}{2}$, the privacy is fully protected. So, C is to be appropriately fixed to keep $V_R(r_i)$ under control and $L(y_i | z_i)$ kept as close to L_i as practicable.

11. Question:

Suppose y is a real-valued variable relating to a stigmatizing characteristic like expenses on treatment of AIDS, gain or loss last month through

gambling, money earned or spent in dubious means with y_i as values relevant to an ith person $i = 1, \cdots, N$. To gather a response from a sampled person suppose an investigator approaches with a box of cards marked either (i) genuine with $C\,(0 < C < 1)$ as their proportion or (ii) marked $x_i, \ldots . x_j \ldots . x_M$ with respective proportions $q_1, \ldots q_j \ldots q_M$ $\left(0 < q_j < 1, j = 1, \cdots, M \text{ such that } C + \sum_{j=1}^{M} q_j = 1\right)$. The device produces the RR from the ith person as

$$
\begin{aligned}
z_i &= y_i \text{ if "genuine" card appears} \\
&= x_j \text{ if } x_j \text{ "marked" card appears}
\end{aligned}
$$

Then, $$E_R\left(z_i\right) = C y_i + \sum_{j=1}^{M} q_j x_j$$

So, $$r_i = \frac{1}{C}\left(z_i - \sum_{1}^{M} q_j x_j\right) \text{ satisfies } E_R\left(r_i\right) = y_i \text{ and}$$

$$
\begin{aligned}
V_i = V_R\left(r_i\right) &= \frac{1}{C^2}\, V_R\left(z_i\right) \\
&= \frac{1}{C^2}\left[E_R\left(z_i^2\right) - E_R^2\left(z_i\right)\right] \\
&= \frac{1}{C^2}\left[\left(C\, y_i^2 + \sum_{j=1}^{M} q_j x_j^2\right) - \left(C^2 y_i^2 + \left(\sum_{1}^{M} q_j x_j\right)^2\right.\right. \\
&\qquad \left.\left. + 2C y_i \left(\sum_{1}^{M} q_j x_j\right)\right)\right] \\
&= \frac{1}{C^2}\left[C\,(1 - C)\, y_i^2 - 2C y_i \left(\sum_{1}^{M} q_j x_j\right)\right. \\
&\qquad \left. + \sum q_j x_j^2 - \left(\sum q_j x_j\right)^2\right] \\
&= \alpha\, y_i^2 + \beta\, y_i + \theta, \text{ say, with } \alpha, \beta \text{ and } \theta
\end{aligned}
$$

as known quantities.

Solution:

An unbiased estimator for V_i is

$$v_i = \frac{\alpha\, r_i^2 + \beta\, r_i + \theta}{(1 + \alpha)}.$$

For a proof Chaudhuri (1992) may be seen.

12. Question:

With the setup in Question 11 find an RR-based estimator for y_i with a variance less than V_i.

Solution:

Let from the ith sample person the RR be

$$\begin{aligned} z_i &= y_i \text{ if "genuine" card comes} \\ &= x_j + f y_i \text{ if } x_j \text{ "marked" card comes.} \end{aligned}$$

Then,

$$\begin{aligned} E_R(z_i) &= C y_i + \sum_1^M (x_j + f y_i) \\ &= [C + f(1 - C)] y_i + \sum q_j x_j = A y_i + B, \text{ say.} \end{aligned}$$

Then,
$$\begin{aligned} & r_i = (z_i - B)/A \quad \text{and} \quad E_R(r_i) = y_i, \\ & A = C + f(1 - C) \\ & E_R(z_i^2) = C y_i^2 + \sum_j q_j (x_j + f y_i)^2. \end{aligned}$$

$$\begin{aligned} \text{So, } V_R(z_i) &= C y_i^2 + f^2 y_i^2 \sum q_j \\ &\quad + 2 f y_i \sum q_j x_j - (A y_i + B)^2 + \sum q_j x_j^2 \\ &= T y_i^2 + F y_i + G, \end{aligned}$$

writing
$$\begin{aligned} & T = C + f^2(1 - C) - A^2, \\ & F = 2f \sum q_j x_j - 2AB, \\ & G = \sum q_j x_j^2 - B^2. \end{aligned}$$

Now compare $\alpha y_i^2 + \beta y_i + \theta$ versus $T y_i^2 + F y_i + G$ to check their comparative values given $\alpha, \beta, \theta, T, F, G$.