STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. 11, March 2023, pp 368–389.
Published online in International Academic Press (www.IAPress.org)

On Kernel-Based Estimator of Odds Ratio Using Different Stratified Sampling Schemes

Abbas Eftekharian,[1] Hani Samawi[2] and Haresh Rochani[2]

[1]Department of Statistics, School of Sciences, University of Hormozgan, P.O. Box 3995, Bandar Abbas, Iran.
[2] Department of Biostatistics, Jiann-Ping Hsu College of Public Health, Georgia Southern University,
P.O. Box 8015, Statesboro, GA 30460, USA

Abstract The kernel-based estimator of Cochran Mantel-Haenszel odds ratio based on stratified simple and ranked set sampling is proposed. The expectation and variance of the estimator are analytically obtained. Using a simulation study, the estimator based on stratified ranked set sampling is more efficient than its counterpart based on stratified simple random sampling. Finally, the estimator's performance is investigated by using base deficit data.

Keywords Cochran Mantel-Haenszel odds ratio, Kernel estimation, Odds ratio, Stratified simple random sampling, Stratified ranked set sampling.

AMS 2010 subject classifications 62G30

DOI: 10.19139/soic-2310-5070-1425

1. Introduction

Odds ratio (OR) is widely used in medical, social, behavioral, and public health sciences. The OR is equally valid for retrospective, prospective, and cross-sectional sampling designs. The OR is the ratio of odds of an event occurring in one group to the other group's odds. Let the probabilities of an event in each of the groups be π_1 (first group) and π_2 (second group) respectively, and then the OR is

$$\frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)}. \tag{1}$$

OR of 1 indicates that the condition or event under study is equally likely to occur in both groups. The OR of greater than 1 means that the condition is more likely to happen in the first group. However, the OR less than 1 indicates that the condition is less likely to occur in the first group. The OR must be greater than or equal to zero if it is defined. It is undefined if $\pi_2(1-\pi_1)$ equals zero.

The OR estimation problem has been widely discussed in the literature by many authors. Most of the literature has developed under the simple random sampling (SRS) method by focusing on different estimation methods of the distribution functions of two populations and using different types of data. [1] presented some properties of OR based on the categorical data. Some of the recent contributions are by [19], [13], [39], [37], [18], [26].

In some situations, the actual measurement is expensive or time-consuming and/or challenging to obtain. Simultaneously, judgment ranking or visual inspection for some variable is more straightforward

*Correspondence to: Hani Samawi (Email: samawi.hani2@gmail.com).

ISSN 2310-5070 (online) ISSN 2311-004X (print)

than the exact measurement. Using judgment ranking, many researchers have used sampling methods such as ranked set sampling (RSS) to collect the data, which was first introduced by [21]. This method is widely discussed in biomedical and agriculture studies, especially for estimating population parameters such as the mean and distribution function. By performing the following steps, you can obtain an RSS as follows:

- Select r independent sets, each involving r independent and identically distributed (iid) sampling units from an underlying population.
- Rank each set by judgment based on the certain knowledge of the subject or by visual inspection, without actual measurement of the interesting variable.
- Measure the jth smallest one from the jth set $(j = 1, \cdots, r)$.

The aforementioned steps present one cycle of an RSS data set. By repeating the process for m times, an RSS data set is obtained with m and $n = mr$ cycles and size. If judgment rankings are done without error, then ranking is perfect, and otherwise, it is called imperfect ranking. Recently, some attempts are performed on the effect of imperfect ranking, as in [14] and [36].

Some authors have recently introduced new schemes based on RSS ideas such as [2], [23] and [31]. An extension of RSS is stratified RSS (SRSS), which is introduced by [30]. SRSS is a sampling scheme in which a population is separated into L mutually distinct strata, with set size r_h of the ranked set sample, which is quantified within the stratum h. In other words, this scheme can be considered as a set of L distinct ranked set samples. To explain SRSS, suppose that the population is divided into L mutually distinct strata and

$$\left(X^*_{h11}, X^*_{h12}, \cdots, X^*_{h1r_h}\right); \left(X^*_{h21}, X^*_{h22}, \cdots, X^*_{h2r_h}\right); \cdots; \left(X^*_{hr_h1}, X^*_{hr_h2}, \cdots, X^*_{hr_hr_h}\right)$$

be r_h independent random samples with size r_h selected from the hth stratum $(h = 1, 2, \cdots, L)$. Hereafter, assume that X_{hji} stands for the quantitative measurement of the unit X^*_{hji}. Let $(X_{h1i}, X_{h2i}, \cdots, X_{hr_hi}; i = 1, 2, \cdots, m_h)$ be the simple random sample with size $n_h = m_h r_h$ which is collected from the hth subpopulation characterized by $F_h(x)$ and $f_h(x)$ as the cumulative distribution function (CDF) and probability density function (PDF), respectively. We should highlight that the CDF of the underlying population can be written as $F(x) = \sum_{h=1}^{L} W_h F_h(x)$, where W_h is the weight of the hth stratum and predetermined from the sampling scheme. To performing RSS scheme in the population with h subpopulation, we can consider the SRSS scheme as a two-phase sampling scheme. First, rank the units in each sample with respect to the variable of interest without actual measurement. In the hth stratum, the judgment ordered sample corresponding to the jth sampling units is denoted by $(X^*_{hj[1]i}, X^*_{hj[2]i}, \cdots, X^*_{hj[r_h]i}; i = 1, 2, \cdots, m_h)$. Second, the ranked set sample with size $n_h = m_h r_h$ in the hth stratum is given by $(X_{h1[1]i}, X_{h2[2]i}, \cdots, X_{hr_h[r_h]i}; i = 1, 2, \cdots, m_h)$. Also, suppose that $N_1, N_2, \cdots, N_L$ are the number of sampling units within each stratum such that $N = \sum_{h=}^{L} N_h$ represents the total underlying population size. In addition, let $n_1, n_2, \cdots, n_L$ denote the number of sampling units measured within each stratum such that $n = \sum_{h=1}^{L} n_h$ represents the total sample size. It should be noted that the notation $[\cdot]$ in subscript is used for imperfect ranked set sample.

Recently, some papers have been published based on the SRSS scheme; [33] presented a ratio estimation for the ratio of means of two dependent variables. Modified ratio estimators for the mean of the finite population have been introduced by [20] using auxiliary variable information. [28] published a kernel density estimation based on the SRSS with optimal allocation. Also, [29] used SRSS to improve distribution and quality estimations' performance. Moreover, [12] proposed a kernel-based estimator of the quantile based on SRSS with optimal allocation.

Unlike the SRS case, a few numbers of papers addressed the OR problem using RSS scheme. The evaluation and the estimation of OR were investigated in the literature when the sampling scheme is moving extreme RSS by [32] and [15]. They used the moving extreme RSS empirical distribution function to estimate OR. By the definition of the OR, it is clear that one can estimate the parameter using the CDF estimation, for example, by an empirical distribution function. Although the empirical distribution function is the most well-known non-parametric estimation of the CDF, it is a step function. The empirical estimator's deficiency share is that it can not show the corresponding continuous parameter's smoothness being estimated. Furthermore, such an estimator may represent much bias near the boundaries, which implies that it cannot estimate beyond the extreme observations.

On the other hand, in the most practical situations, such as lifetime and biomedical, the underlying population is continuous, and the OR is a continuous parameter. We want to find a smoother estimator than the empirical estimator, for example, a kernel estimator. It is obvious that the kernel estimator is a continuous and smooth function, and on the one hand, it is mentioned by many authors that smooth estimators have better performance than that of empirical counterparts, provided that a proper bandwidth is chosen. Therefore, these motivations encourage us to use a smooth estimator such as the kernel estimator for estimating the OR. The kernel estimator is more efficient than the empirical in the literature (see [4]) and is available on SAS, R, and other software.

Some authors have recently placed their attention on the kernel estimator for evaluating interesting parameters using the RSS scheme. [8] used the kernel function to estimate the log OR for sparse data. A rank-based kernel estimator for the area under the ROC curve was proposed by [38]. [11] presented a kernel-based estimator for odds using the RSS method, and they show that the estimator has better performance than the empirical counterpart.

The rest of the paper is organized as follows. In Section 2, some preliminaries are presented. The kernel estimator of the OR based on the SRS in the hth stratum is investigated in Section 3. In Section 4, the OR's kernel estimator using the RSS scheme in the hth stratum is considered. The OR estimations on the basis of SRSS and SRSS are studied in Section 5. In Section 6, simulation and numerical computations studies are stated, while illustrations using real data of our proposed estimators are presented in Section 7. In Section 8, some discussions and conclusions are stated.

2. Preliminaries

In this paper, we will focus on continuous biomarkers to simplify the interpretation of the results. Let X and Y denote biomarker results of subjects from the healthy and diseased populations, respectively. Let x_0 be the cut-off point. Suppose $F_X(x_0) = P(X \leq x_0)$ and $F_Y(x_0) = P(Y \leq x_0)$ are the underlying CDFs of two absolutely continuous random variables X and Y, respectively. Let the positive test result to be $D = \{y : Y > x_0\}$ for the diseased and $H = \{x : X > x_0\}$ for the healthy, respectively. The odds of the positive test in the disease population is given by

$$O_Y(x_0) = \frac{P(D)}{1-P(D)} = \frac{1-F_Y(x_0)}{F_Y(x_0)} = \frac{\pi_1}{1-\pi_1},$$

and the odds of the positive test in the health population is given by

$$O_X(x_0) = \frac{P(H)}{1-P(H)} = \frac{1-F_X(x_0)}{F_X(x_0)} = \frac{\pi_2}{1-\pi_2}.$$

Therefore, the odds ratio is given by

$$\theta(x_0) = \frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)} = \frac{O_Y(x_0)}{O_X(x_0)}.$$

Note that $O_Y(x_0) \geq 0$ and $O_X(x_0) > 0$. Throughout the paper, we will assume that $0 < F_Y(x_0) < 1$ and $0 < F_X(x_0) < 1$ so that $O_Y(x_0)$ and $O_X(x_0)$ are both finite.

In the healthy and the diseased population naturally stratified by some covariates, such as age groups and gender, the biomarker results are homogeneous within each stratum (sub-group) and heterogeneous between strata (subgroups). Let the positive test result within stratum to be $D_h = \{y_h : Y_h > x_0\}$ for the diseased and $H_h = \{x_h : X_h > x_0\}$ for the healthy, respectively, where $h = 1, 2, \cdots, L$. The odds of a positive test in the disease population within the hth stratum is given by

$$O_{Y_h}(x_0) = \frac{P(D_h)}{1 - P(D_h)} = \frac{1 - F_{Y_h}(x_0)}{F_{Y_h}(x_0)} = \frac{\pi_{h1}}{1 - \pi_{h1}}, \tag{2}$$

and the odds of a positive test in health population within the hth stratum is given by

$$O_{X_h}(x_0) = \frac{P(H_h)}{1 - P(H_h)} = \frac{1 - F_{X_h}(x_0)}{F_{X_h}(x_0)} = \frac{\pi_{h2}}{1 - \pi_{h2}}. \tag{3}$$

Therefore, the common odds ratio is given by

$$\theta(x_0) = \sum_{h=1}^{L} W_h \theta_h(x_0),$$

where $W_h = N_h/N$ and N_h is the hth stratum size and

$$\theta_h(x_0) = \frac{O_{Y_h}(x_0)}{O_{X_h}(x_0)}.$$

The other way of looking at the common odds ratio is the Cochran Mantel-Haenszel approach as follows:

$$\theta(x_0) = \frac{\sum_{h=1}^{L} \gamma_h \, \theta_h(x_0)}{\sum_{h=1}^{L} \gamma_h}, \tag{4}$$

where $\gamma_h = N_h F_{Y_h}(x_0)(1 - F_{X_h}(x_0))$.

The OR of the underlying population, $\theta(x_0)$, is an unknown parameter, and to estimate OR, we need to estimate the CDFs in both groups. In the next two sections, we introduce the kernel estimator of OR, based on two sampling methods, SRS and RSS.

3. Kernel estimation of OR based on SRS in the hth stratum

Kernel estimation was first introduced by [27] for estimating density function. It has been widely applied as a smoothing method to estimate other quantities such as CDF in the last three decades. The first contributions based on SRS were presented by [24], [4] and [3]. [4] illustrated that the kernel estimator of CDF with bounded support is more efficient than the unbounded backing or empirical distribution function. This result has been confirmed again in the literature by other authors. In the present paper, we consider the kernel function with bounded support.

Let $Y_{h1}, \cdots, Y_{hn_{Y_h}}$ and $X_{h1}, \cdots, X_{hn_{X_h}}$ be simple random samples of sizes n_{Y_h} and n_{X_h} for the hth stratum in both groups with CDFs $F_{Y_h}(\cdot)$ and $F_{X_h}(\cdot)$, respectively. Accordingly, the kernel estimators of $F_{Y_h}(\cdot)$ and $F_{X_h}(\cdot)$ are as follows:

$$\hat{F}_{Y_h}^{SRS}(x_0) = \frac{1}{n_{Y_h}} \sum_{i=1}^{n_{Y_h}} K_{Y_h}\left(\frac{x_0 - Y_{hi}}{d_{Y_h}}\right), \quad \hat{F}_{X_h}^{SRS}(x_0) = \frac{1}{n_{X_h}} \sum_{i=1}^{n_{X_h}} K_{X_h}\left(\frac{x_0 - X_{hi}}{d_{X_h}}\right), \tag{5}$$

where d_{Y_h} and d_{X_h} are the bandwidths and

$$K_s(u) = \begin{cases} 0, & u \leq -a, \\ \int_{-a}^{u} k_s(t)dt, & |u| < a, \\ 1, & u \geq a, \end{cases}$$

$k_s(t)$ is the kernel function with bounded support, that is, $k_s(t) = 0$, if $|t| > 0$, for some positive a and $s = Y_h, X_h$. We consider the following assumptions throughout the paper:
(A1) The underlying population has the CDFs as $F_{Y_h}(\cdot)$ and $F_{X_h}(\cdot)$, which are Hölder continuous with a square-integrable second derivative for any $h = 1, \cdots, L$.
(A2) $K_s(\cdot)$ is an absolutely continuous function, such that $\lim_{x\to -a} K_s(x) = 0$ and $\lim_{x\to a} K_s(x) = 1$.
(A3) The kernel function $k_s(x)$ satisfies the following conditions for any x

$$k_s(x) = k_s(-x), \quad \int_{-a}^{a} k_s(x)dx = 1, \quad \int_{-a}^{a} x^2 k_s(x)dt \neq 0.$$

Since d_ss are smoothing parameters in the kernel estimation problem, the choice of optimal bandwidths is more important for the convergence of $\hat{F}_s^{SRS}$s to F_{Y_h}s. The most common method used in the literature to find the optimal bandwidth is based on the mean integrated squared error (MISE) criterion that is defined as follows:

$$\begin{aligned}\text{MISE}(\hat{F}_s^{SRS}) &= \int \text{MSE}(\hat{F}_s^{SRS}(x))dx \\ &= \int E\left[\hat{F}_s^{SRS}(x) - F_s(x)\right]^2 dx = E\int \left[\hat{F}_s^{SRS}(x) - F_s(x)\right]^2 dx, \end{aligned} \tag{6}$$

where MSE is mean squared error and for $s = Y_h, X_h$. There are several asymptotic technics to find the optimal bandwidth. The most popular technics are plug-in and cross-validation. [3] showed that the estimator obtained based on the plug-in method has better performance than using cross-validation to estimate the CDF. A multistage plug-in method to find the optimal bandwidth was proposed by [25] using an iterative algorithm. The MSE and MISE of the kernel-based estimator of CDF obtained by [4], [16] and [3] are confirmed by [25]. As in [16], by assuming large subpopulations, $d_s \to 0$ and $n_s d_s \to \infty$ as $n_s \to \infty$ and under the assumptions (A1)-(A3), the bias and variance of $\hat{F}_s^{SRS}(x)$ can be obtained as

$$\text{Bias}(\hat{F}_s^{SRS}(x)) = \frac{d_s^2}{2} f_s'(x) \int_{-a}^{a} u^2 k_s(u)du + o_s(d_s^4), \tag{7}$$

$$\text{Var}(\hat{F}_s^{SRS}(x)) = \frac{1}{n_s} F_s(x)\bar{F}_s(x) - \frac{d_s f_s(x)}{n_s}\left\{a - \int_{-a}^{a} K_s^2(u)du\right\} + o_s(d_s^4), \tag{8}$$

where $\bar{F}_s(\cdot) = 1 - F_s(\cdot)$ is the survival function.

Therefore, the MSE of $\hat{F}_s^{SRS}(x)$ as given by (4)

$$\begin{aligned}\text{MSE}\left\{\hat{F}_s^{SRS}(x)\right\} =& \frac{1}{n_s} F_s(x)\bar{F}_s(x) - \frac{d_s f_s(x)}{n_s}\left\{a - \int_{-a}^{a} K_s^2(u)du\right\} + \frac{d_s^4}{4}\left\{f_s'(x)\int_{-a}^{a} u^2 k_s(u)du\right\}^2 \\ &+ o_s(d_s^4). \end{aligned} \tag{9}$$

Moreover, using (9), the MISE of $\hat{F}_s^{SRS}$ is given by

$$\text{MISE}(\hat{F}_s^{SRS}) = \int E\left[\hat{F}_s^{SRS}(x) - F_s(x)\right]^2 dx = \int \left\{\left[\text{Bias}(\hat{F}_s^{SRS}(x))\right]^2 + \text{Var}(\hat{F}_s^{SRS}(x))\right\} dx.$$

As mentioned in [16] and [25], by assuming $d_s \to 0$ and $n_s d_s \to \infty$ as $n_s \to \infty$ and under the assumptions (A1)-(A3), the asymptotic optimal bandwidth is as

$$d_s^{SRS} = \left[\frac{a - \int_{-a}^{a} K_s^2(u)du}{\left\{\int_{-a}^{a} u^2 k_s(u)du\right\}^2 R(f_s')}\right]^{1/3} n_s^{-\frac{1}{3}}, \tag{10}$$

where $R(f) := \int_{-\infty}^{\infty} (f(x))^2 \, dx$.

Using (2) and (3) the SRS kernel-based estimators for the odds of positive tests in disease and health populations within the hth stratum are given by

$$\hat{O}_{Y_h}^{SRS}(x_0) = \frac{1 - \hat{F}_{Y_h}^{SRS}(x_0)}{\hat{F}_{Y_h}^{SRS}(x_0)}, \quad \hat{O}_{X_h}^{SRS}(x_0) = \frac{1 - \hat{F}_{X_h}^{SRS}(x_0)}{\hat{F}_{X_h}^{SRS}(x_0)} = \frac{1}{\hat{F}_{X_h}^{SRS}(x_0)} - 1. \tag{11}$$

Then, the $\mathrm{E}\left[\hat{O}_s^{SRS}(x_0)\right]$ can be obtained by Taylor expansion of $1/\hat{F}_s^{SRS}(x_0)$ as given by ([22, p.181])

$$\begin{aligned}\mathrm{E}\left[\hat{O}_s^{SRS}(x_0)\right] &= \mathrm{E}\left[\frac{1}{\hat{F}_s^{SRS}(x_0)}\right] - 1 \approx \frac{1}{\mathrm{E}\left[\hat{F}_s^{SRS}(x_0)\right]} + \frac{\mathrm{Var}\left(\hat{F}_s^{SRS}(x_0)\right)}{\left(\mathrm{E}\left[\hat{F}_s^{SRS}(x_0)\right]\right)^3} - 1 \\ &= \frac{1 - F_s(x_0) - \frac{d_s^2}{2} f_s'(x) \int_{-a}^{a} u^2 k_s(u)du}{F_s(x_0) + \frac{d_s^2}{2} f_s'(x) \int_{-a}^{a} u^2 k_s(u)du} + \frac{\frac{1}{n_s} F_s(x)\bar{F}_s(x) - \frac{d_s f_s(x)}{n_s}\left\{a - \int_{-a}^{a} K_s^2(u)du\right\}}{\left(F_s(x_0) + \frac{d_s^2}{2} f_s'(x) \int_{-a}^{a} u^2 k_s(u)du\right)^3} \\ &\quad + o_s(d_s^4). \end{aligned} \tag{12}$$

With the same argument it can be shown that

$$\begin{aligned}\mathrm{Var}\left[\hat{O}_s^{SRS}(x_0)\right] = \mathrm{Var}\left[\frac{1}{\hat{F}_s^{SRS}(x_0)}\right] &\approx \frac{\mathrm{Var}\left(\hat{F}_s^{SRS}(x_0)\right)}{\left(\mathrm{E}\left[\hat{F}_s^{SRS}(x_0)\right]\right)^4} \\ &= \frac{\frac{1}{n_s} F_s(x)\bar{F}_s(x) - \frac{d_s f_s(x)}{n_s}\left\{a - \int_{-a}^{a} K_s^2(u)du\right\}}{\left(F_s(x_0) + \frac{d_s^2}{2} f_s'(x) \int_{-a}^{a} u^2 k_s(u)du\right)^4} + o_s(d_s^4). \end{aligned} \tag{13}$$

Hence, the MSE of $\hat{O}_s^{SRS}(x_0)$ obtained using (12) and (13). From (11) the SRS kernel-based estimator for the OR of the positive tests in disease to the positive tests in health populations within the hth stratum is given by

$$\hat{\theta}_h^{SRS}(x_0) = \frac{\hat{O}_{Y_h}^{SRS}(x_0)}{\hat{O}_{X_h}^{SRS}(x_0)}. \tag{14}$$

Now, using the presented approximations in [22, p.181] for the expectation and variance of the ratio of two random variables we get

$$\mathrm{E}\left[\hat{\theta}_h^{SRS}(x_0)\right] \approx \frac{\mathrm{E}\left[\hat{O}_{Y_h}^{SRS}(x_0)\right]}{\mathrm{E}\left[\hat{O}_{X_h}^{SRS}(x_0)\right]} + \frac{\mathrm{E}\left[\hat{O}_{Y_h}^{SRS}(x_0)\right]}{\left(\mathrm{E}\left[\hat{O}_{X_h}^{SRS}(x_0)\right]\right)^3} \mathrm{Var}\left(\hat{O}_{X_h}^{SRS}(x_0)\right), \tag{15}$$

$$\mathrm{Var}\left[\hat{\theta}_h^{SRS}(x_0)\right] \approx \left(\frac{\mathrm{E}\left[\hat{O}_{Y_h}^{SRS}(x_0)\right]}{\mathrm{E}\left[\hat{O}_{X_h}^{SRS}(x_0)\right]}\right)^2 \left\{\frac{\mathrm{Var}\left(\hat{O}_{Y_h}^{SRS}(x_0)\right)}{\left(\mathrm{E}\left[\hat{O}_{Y_h}^{SRS}(x_0)\right]\right)^2} + \frac{\mathrm{Var}\left(\hat{O}_{X_h}^{SRS}(x_0)\right)}{\left(\mathrm{E}\left[\hat{O}_{X_h}^{SRS}(x_0)\right]\right)^2}\right\}, \tag{16}$$

where $\mathrm{E}\left[\hat{O}_s^{SRS}(x_0)\right]$ and $\mathrm{Var}\left(\hat{O}_s^{SRS}(x_0)\right)$ are as in (12) and (13), for $s = Y_h, X_h$. Therefore, the MSE of $\hat{\theta}_h^{SRS}(x_0)$ can be derived by using (15) and (16).

4. Kernel estimation of OR using RSS scheme for the $\boldsymbol{h}$th stratum

The first attempt to find an estimate of CDF using the RSS scheme returns to [34]. They introduced the RSS empirical distribution function and investigated some properties of this estimator. Using the RSS scheme, most authors such as [7], [5] and [17] have been focused on kernel estimation of PDF. [11] introduced the RSS kernel-based of CDF and compared some properties of this estimator with SRS and empirical estimator.

Let $(Y_{h1[1]i}, Y_{h2[2]i}, \cdots, Y_{hr_{Y_h}[r_{Y_h}]i}\,; i = 1, 2, \cdots, m_{Y_h})$ and $(X_{h1[1]i}, X_{h2[2]i}, \cdots, X_{hr_{X_h}[r_{X_h}]i}\,; i = 1, 2, \cdots, m_{X_h})$ be ranked set samples of sizes $n_{Y_h} = m_{Y_h} r_{Y_h}$ and $n_{X_h} = m_{X_h} r_{X_h}$ from the hth stratum with CDFs $F_{Y_h}(\cdot)$ and $F_{X_h}(\cdot)$, respectively. The kernel-based estimators of the CDFs for the hth stratum can be defined as follows

$$\hat{F}^{RSS}_{Y_h}(x_0) = \frac{1}{m_{Y_h} r_{Y_h}} \sum_{i=1}^{m_{Y_h}} \sum_{j=1}^{r_{Y_h}} K_{Y_h}\left(\frac{x_0 - Y_{hj[j]i}}{d_{Y_h}}\right),$$

$$\hat{F}^{RSS}_{X_h}(x_0) = \frac{1}{m_{X_h} r_{X_h}} \sum_{i=1}^{m_{X_h}} \sum_{j=1}^{r_{X_h}} K_{X_h}\left(\frac{x_0 - X_{hj[j]i}}{d_{X_h}}\right). \tag{17}$$

Consider (A1)-(A3) and assume that $d_s \to 0$ and $n_s d_s \to \infty$ as $n_s \to \infty$. Under these assumptions, as in [11], the bias and variance of the $\hat{F}^{RSS}_s(x_0)$ within stratum h is as

$$\text{Bias}(\hat{F}^{RSS}_s(x_0)) = \frac{d_s^2}{2} f_s'(x_0) \int_{-a}^{a} u^2 k_s(u)du + o_s(d_s^4),$$

$$\text{Var}(\hat{F}^{RSS}_s(x_0)) = \frac{1}{m_s r_s^2} \sum_{j=1}^{r_s} F_{[j]s}(x_0)\bar{F}_{[j]s}(x_0) - \frac{d_s f_s(x_0)}{n_s}\left\{a - \int_{-a}^{a} K_s^2(u)du\right\} + o_s(d_s^4).$$

Therefore, it can be concluded that

$$\begin{aligned} \text{MSE}\left\{\hat{F}^{RSS}_s(x)\right\} =& \frac{1}{m_s r_s^2} \sum_{j=1}^{r_s} F_{[j]s}(x)\bar{F}_{[j]s}(x) - \frac{d_s f_s(x)}{n_s}\left\{a - \int_{-a}^{a} K_s^2(u)du\right\} \\ &+ \frac{d_s^4}{4}\left\{f_s'(x)\int_{-a}^{a} u^2 k_s(u)du\right\}^2 + o_s(d_s^4). \end{aligned} \tag{18}$$

By comparing (9) with (18), we observed that the MSE's are the same except in the first terms of the right-hand side of equalities. As in [11], under the above assumptions and for fixed $n_s = m_s r_s$, $\text{MSE}(\hat{F}^{RSS}_s(x_0)) \le \text{MSE}(\hat{F}^{SRS}_s(x_0))$. As in [11], the optimal bandwidth based on RSS scheme for the hth stratum is the same as in (10) and $\text{MISE}(\hat{F}^{RSS}_h(x)) \le \text{MISE}(\hat{F}^{SRS}_h(x))$.

According to (17), the RSS kernel estimations of the odds of positive tests in disease and health populations within the hth stratum are obtained by

$$\hat{O}^{RSS}_{Y_h}(x_0) = \frac{1 - \hat{F}^{RSS}_{Y_h}(x_0)}{\hat{F}^{RSS}_{Y_h}(x_0)}, \quad \hat{O}^{RSS}_{X_h}(x_0) = \frac{1 - \hat{F}^{RSS}_{X_h}(x_0)}{\hat{F}^{RSS}_{X_h}(x_0)}. \tag{19}$$

Using the Taylor expansion, we get

$$\begin{aligned} \text{E}\left[\hat{O}^{RSS}_s(x_0)\right] \approx& \frac{1 - F_s(x_0) - \frac{d_s^2}{2} f_s'(x)\int_{-a}^{a} u^2 k_s(u)du}{F_s(x_0) + \frac{d_s^2}{2} f_s'(x)\int_{-a}^{a} u^2 k_s(u)du} \\ &+ \frac{\frac{1}{m_s r_s^2}\sum_{j=1}^{r_s} F_{[j]s}(x_0)\bar{F}_{[j]s}(x_0) - \frac{d_s f_s(x_0)}{n_s}\left\{a - \int_{-a}^{a} K_s^2(u)du\right\}}{\left(F_s(x_0) + \frac{d_s^2}{2} f_s'(x)\int_{-a}^{a} u^2 k_s(u)du\right)^3} + o_s(d_s^4), \end{aligned} \tag{20}$$

and

$$\mathrm{Var}\left[\hat{O}_s^{RSS}(x_0)\right] \approx \frac{\frac{1}{m_s r_s^2}\sum_{j=1}^{r_s} F_{[j]s}(x_0)\bar{F}_{[j]s}(x_0) - \frac{d_s f_s(x_0)}{n_s}\left\{a - \int_{-a}^{a} K_s^2(u)du\right\}}{\left(F_s(x_0) + \frac{d_s^2}{2} f_s'(x)\int_{-a}^{a} u^2 k_s(u)du\right)^4} + o_s(d_s^4). \tag{21}$$

Also, the MSE of $\hat{O}_s^{RSS}(x_0)$ can be derived by equations (20) and (21).

Corollary 1
By comparing equations (12) and (13) with (20) and (21), we have

$$\mathrm{MSE}\left[\hat{O}_s^{RSS}(x_0)\right] \leq \mathrm{MSE}\left[\hat{O}_s^{SRS}(x_0)\right].$$

Proof
From $\frac{1}{m_s r_s^2}\sum_{j=1}^{r_s} F_{[j]s}(x_0)\bar{F}_{[j]s}(x_0) \leq \frac{1}{n_s}F_s(x_0)\bar{F}_s(x_0)$ (see, [11]), it may be concluded that $\mathrm{MSE}\left[\hat{O}_s^{RSS}(x_0)\right] \leq \mathrm{MSE}\left[\hat{O}_s^{SRS}(x_0)\right]$. □ □

Now, based on equation (19), the RSS kernel-based estimator of OR within stratum h is defined as

$$\hat{\theta}_h^{RSS}(x_0) = \frac{\hat{O}_{Y_h}^{RSS}(x_0)}{\hat{O}_{X_h}^{RSS}(x_0)}, \tag{22}$$

with expectation and variance

$$\mathrm{E}\left[\hat{\theta}_h^{RSS}(x_0)\right] \approx \frac{\mathrm{E}\left[\hat{O}_{Y_h}^{RSS}(x_0)\right]}{\mathrm{E}\left[\hat{O}_{X_h}^{RSS}(x_0)\right]} + \frac{\mathrm{E}\left[\hat{O}_{Y_h}^{RSS}(x_0)\right]}{\left(\mathrm{E}\left[\hat{O}_{X_h}^{RSS}(x_0)\right]\right)^3}\mathrm{Var}\left(\hat{O}_{X_h}^{RSS}(x_0)\right), \tag{23}$$

$$\mathrm{Var}\left[\hat{\theta}_h^{RSS}(x_0)\right] \approx \left(\frac{\mathrm{E}\left[\hat{O}_{Y_h}^{RSS}(x_0)\right]}{\mathrm{E}\left[\hat{O}_{X_h}^{RSS}(x_0)\right]}\right)^2 \left\{\frac{\mathrm{Var}\left(\hat{O}_{Y_h}^{RSS}(x_0)\right)}{\left(\mathrm{E}\left[\hat{O}_{Y_h}^{RSS}(x_0)\right]\right)^2} + \frac{\mathrm{Var}\left(\hat{O}_{X_h}^{RSS}(x_0)\right)}{\left(\mathrm{E}\left[\hat{O}_{X_h}^{RSS}(x_0)\right]\right)^2}\right\}, \tag{24}$$

where $\mathrm{E}\left[\hat{O}_s^{RSS}(x_0)\right]$ and $\mathrm{Var}\left(\hat{O}_s^{RSS}(x_0)\right)$ are given in Equations (20) and (21) for $s = Y_h, X_h$. Hence, the MSE of $\hat{\theta}_h^{RSS}(x_0)$ derives easily using (23) and (24).

5. The OR estimation using stratified sampling

5.1. The OR estimation using SSRS

In order to get the estimator of OR, we need to investigate the convergence property of $\hat{F}_s^{SRS}(x)$ and $\hat{F}_s^{RSS}(x)$ to $F_s(x)$. In this section, we present the common OR estimator and investigate some asymptotic properties. Note that we assume large subpopulations, so the finite correction fraction is negligible. First, consider $\hat{F}_s^{SRS}(x)$ and that $F_s : \mathbb{R} \to [0,1]$ is continuous at x. [24] verified under assumptions (A1)-(A3) that $\mathrm{MSE}(\hat{F}_s^{SRS}(x)) \to 0$ as $n_s \to \infty$. So, the $\hat{F}_s^{SRS}(x)$ is a consistent estimator and it implies $\hat{F}_s^{SRS}(x) \xrightarrow{P} F_s(x)$. On the other hand, $g(x) = \frac{1-x}{x}$ is a continuous function for any $x \neq 0$. Therefore, $\hat{O}_s^{SRS}(x) \xrightarrow{P} O_s(x)$. Consequently, by assuming $\hat{O}_{X_h}^{SRS}(x) \neq 0$ and $O_{X_h}(x) \neq 0$, and using properties of convergence in probability, we have $\hat{\theta}_h^{SRS}(x) \xrightarrow{P} \theta_h(x)$.

Finally, the kernel-based Cochran Mantel-Haenszel estimator for the underlying population common OR based on the SSRS is given by

$$\hat{\theta}^{SSRS}(x_0) = \frac{\sum_{h=1}^{L} \hat{\gamma}_h^{SRS}\, \hat{\theta}_h^{SRS}(x_0)}{\sum_{h=1}^{L} \hat{\gamma}_h^{SRS}}, \tag{25}$$

where $\hat{\gamma}_h^{SRS} = n_h \hat{F}_{Y_h}^{SRS}(x_0)\left(1 - \hat{F}_{X_h}^{SRS}(x_0)\right)$, and $\hat{\theta}_h^{SRS}(x_0)$ as defined in equation (14), is the estimated OR for the hth stratum. By rewriting (25) we get

$$\hat{\theta}^{SSRS}(x_0) = \frac{\sum_{h=1}^{L} n_h \hat{F}_{X_h}^{SRS}(x_0)\left(1 - \hat{F}_{Y_h}^{SRS}(x_0)\right)}{\sum_{h=1}^{L} n_h \hat{F}_{Y_h}^{SRS}(x_0)\left(1 - \hat{F}_{X_h}^{SRS}(x_0)\right)}. \tag{26}$$

Under assumptions (A1)-(A3) and assume that $d_s \to 0$ and $n_s d_s \to \infty$ as $n_s \to \infty$ and by considering large population size, using the approximations for the expectation and variance of the ratio of two random variables, it can be shown by using similar argument as in [22, p.181] that

$$\begin{aligned}
\mathrm{E}\left[\hat{\theta}^{SSRS}(x_0)\right] \approx{}& \frac{\sum_{h=1}^{L} n_h \mathrm{E}\left[\hat{F}_{X_h}^{SRS}(x_0)\right] \mathrm{E}\left[1 - \hat{F}_{Y_h}^{SRS}(x_0)\right]}{\sum_{h=1}^{L} n_h \mathrm{E}\left[\hat{F}_{Y_h}^{SRS}(x_0)\right] \mathrm{E}\left[1 - \hat{F}_{X_h}^{SRS}(x_0)\right]} \\
&- \frac{1}{\left(\sum_{h=1}^{L} n_h \mathrm{E}\left[\hat{F}_{Y_h}^{SRS}(x_0)\right] \mathrm{E}\left[1 - \hat{F}_{X_h}^{SRS}(x_0)\right]\right)^2} \left[\sum_{h=1}^{L} n_h^2 \left\{ \mathrm{Var}\left(\hat{F}_{X_h}^{SRS}(x_0)\right) \left(\mathrm{E}^2\left[\hat{F}_{Y_h}^{SRS}(x_0)\right]\right.\right.\right. \\
&\left. - \mathrm{E}\left[\hat{F}_{Y_h}^{SRS}(x_0)\right]\right) + \mathrm{Var}\left(\hat{F}_{Y_h}^{SRS}(x_0)\right) \left(\mathrm{E}^2\left[\hat{F}_{X_h}^{SRS}(x_0)\right] - \mathrm{E}\left[\hat{F}_{X_h}^{SRS}(x_0)\right]\right) \\
&\left.\left. + \mathrm{Var}\left(\hat{F}_{Y_h}^{SRS}(x_0)\right) \mathrm{Var}\left(\hat{F}_{X_h}^{SRS}(x_0)\right)\right\}\right] + \frac{\sum_{h=1}^{L} n_h \mathrm{E}\left[\hat{F}_{X_h}^{SRS}(x_0)\right] \mathrm{E}\left[1 - \hat{F}_{Y_h}^{SRS}(x_0)\right]}{\left(\sum_{h=1}^{L} n_h \mathrm{E}\left[\hat{F}_{Y_h}^{SRS}(x_0)\right] \mathrm{E}\left[1 - \hat{F}_{X_h}^{SRS}(x_0)\right]\right)^3} \\
&\times \sum_{h=1}^{L} n_h^2 \left\{ \mathrm{Var}\left(\hat{F}_{X_h}^{SRS}(x_0)\right) \mathrm{E}^2\left[\hat{F}_{Y_h}^{SRS}(x_0)\right] + \mathrm{Var}\left(\hat{F}_{Y_h}^{SRS}(x_0)\right) \mathrm{E}^2\left[1 - \hat{F}_{X_h}^{SRS}(x_0)\right]\right. \\
&\left. + \mathrm{Var}\left(\hat{F}_{Y_h}^{SRS}(x_0)\right) \mathrm{Var}\left(\hat{F}_{X_h}^{SRS}(x_0)\right)\right\},
\end{aligned} \tag{27}$$

and

$$\begin{aligned}\operatorname{Var}\left(\hat{\theta}^{SSRS}(x_0)\right) \approx & \left\{\frac{\sum_{h=1}^{L} n_h \mathrm{E}\left[\hat{F}_{X_h}^{SRS}(x_0)\right] \mathrm{E}\left[1-\hat{F}_{Y_h}^{SRS}(x_0)\right]}{\sum_{h=1}^{L} n_h \mathrm{E}\left[\hat{F}_{Y_h}^{SRS}(x_0)\right] \mathrm{E}\left[1-\hat{F}_{X_h}^{SRS}(x_0)\right]}\right\}^2 \\
& \times \Bigg\{ \frac{1}{\left(\sum_{h=1}^{L} n_h \mathrm{E}\left[\hat{F}_{Y_h}^{SRS}(x_0)\right] \mathrm{E}\left[1-\hat{F}_{X_h}^{SRS}(x_0)\right]\right)^2} \sum_{h=1}^{L} n_h^2 \Big[\operatorname{Var}\left(\hat{F}_{X_h}^{SRS}(x_0)\right) \mathrm{E}^2\left[\hat{F}_{Y_h}^{SRS}(x_0)\right] \\
& + \operatorname{Var}\left(\hat{F}_{Y_h}^{SRS}(x_0)\right) \mathrm{E}^2\left[1-\hat{F}_{X_h}^{SRS}(x_0)\right] + \operatorname{Var}\left(\hat{F}_{Y_h}^{SRS}(x_0)\right) \operatorname{Var}\left(\hat{F}_{X_h}^{SRS}(x_0)\right) \Big] \\
& + \frac{1}{\left(\sum_{h=1}^{L} n_h \mathrm{E}\left[\hat{F}_{X_h}^{SRS}(x_0)\right] \mathrm{E}\left[1-\hat{F}_{Y_h}^{SRS}(x_0)\right]\right)^2} \sum_{h=1}^{L} n_h^2 \Big[\operatorname{Var}\left(\hat{F}_{Y_h}^{SRS}(x_0)\right) \mathrm{E}^2\left[\hat{F}_{X_h}^{SRS}(x_0)\right] \\
& + \operatorname{Var}\left(\hat{F}_{X_h}^{SRS}(x_0)\right) \mathrm{E}^2\left[1-\hat{F}_{Y_h}^{SRS}(x_0)\right] + \operatorname{Var}\left(\hat{F}_{Y_h}^{SRS}(x_0)\right) \operatorname{Var}\left(\hat{F}_{X_h}^{SRS}(x_0)\right) \Big] \\
& - 2\Bigg\{ \left(\sum_{h=1}^{L} n_h \mathrm{E}\left[\hat{F}_{Y_h}^{SRS}(x_0)\right] \mathrm{E}\left[1-\hat{F}_{X_h}^{SRS}(x_0)\right]\right) \Bigg(\sum_{h=1}^{L} n_h \mathrm{E}\left[\hat{F}_{X_h}^{SRS}(x_0)\right] \\
& \times \mathrm{E}\left[1-\hat{F}_{Y_h}^{SRS}(x_0)\right] \Bigg) \Bigg\}^{-1} \Bigg[\sum_{h=1}^{L} n_h^2 \Bigg\{ \operatorname{Var}\left(\hat{F}_{X_h}^{SRS}(x_0)\right) \Bigg(\mathrm{E}^2\left[\hat{F}_{Y_h}^{SRS}(x_0)\right] \\
& - \mathrm{E}\left[\hat{F}_{Y_h}^{SRS}(x_0)\right] \Bigg) + \operatorname{Var}\left(\hat{F}_{Y_h}^{SRS}(x_0)\right) \left(\mathrm{E}^2\left[\hat{F}_{X_h}^{SRS}(x_0)\right] - \mathrm{E}\left[\hat{F}_{X_h}^{SRS}(x_0)\right] \right) \\
& + \operatorname{Var}\left(\hat{F}_{Y_h}^{SRS}(x_0)\right) \operatorname{Var}\left(\hat{F}_{X_h}^{SRS}(x_0)\right) \Bigg\} \Bigg] \Bigg\}. \qquad (28)\end{aligned}$$

For more details see Appendix.

The MSE of $\hat{\theta}^{SSRS}(x_0)$ can be easily derived using equations (27) and (28).

5.2. The OR estimation using SRSS

In this section, we obtain the Cochran Mantel-Haenszel estimator based on the SRSS kernel-based. For this purpose, we first evaluate the convergency in probability of $\hat{F}_s^{RSS}(x)$ to $F_s(x)$.

Remark 1
[24] showed for any kernel function $k_s(\cdot)$ and under assumptions (A1)-(A3) that MSE($\hat{F}_s^{SRS}(x)$) $\to 0$ as $n_s \to \infty$. [4] verified the MSE of $\hat{F}_s^{SRS}(x)$ with a kernel function with bounded support as $[-a, a]$ is less than that of with unbounded support. So, in this case, we have $\hat{F}_s^{SRS}(x) \xrightarrow{P} F_s(x)$, as well.

Using Remark 1, we can present the following theorem.

Theorem 1
Let $\hat{F}_s^{RSS}(x)$ be the RSS kernel-based estimator of $F_s(x)$ as defined in equation (17) and suppose that $d_s \to 0$ and $n_s d_s \to \infty$ as $n_s \to \infty$. Then $\hat{F}_s^{RSS}(x) \xrightarrow{P} F_s(x)$.

Proof
It is enough to show that MSE($\hat{F}_s^{RSS}(x)$) $\to 0$. To do this, note that [11] illustrated that MSE($\hat{F}_s^{RSS}(x)$) $\leq$ MSE($\hat{F}_s^{SRS}(x)$). Therefore, using Remark 1 we can conclude that MSE($\hat{F}_s^{RSS}(x)$) $\to 0$. □ □

Since $\hat{F}_s^{RSS}(x) \xrightarrow{P} F_s(x)$, it implies that $\hat{O}_s^{RSS}(x) \xrightarrow{P} O_s(x)$ for $\hat{F}_s^{RSS}(x) \neq 0$ and $F_s(x) \neq 0$. Hence, by assuming $\hat{O}_{X_h}^{RSS}(x) \neq 0$ and $O_{X_h}(x) \neq 0$ it deduces that $\hat{\theta}_s^{RSS}(x) \xrightarrow{P} \theta_s(x)$.

Now, the kernel-based Cochran Mantel-Haenszel estimator of OR based on the SRSS scheme is as

$$\hat{\theta}^{SRSS}(x_0) = \frac{\sum_{h=1}^{L} \hat{\gamma}_h^{RSS}\, \hat{\theta}_h^{RSS}(x_0)}{\sum_{h=1}^{L} \hat{\gamma}_h^{RSS}}, \tag{29}$$

where $\hat{\gamma}_h^{RSS} = n_h \hat{F}_{Y_h}^{RSS}(x_0)\left(1 - \hat{F}_{X_h}^{RSS}(x_0)\right)$, and $\hat{\theta}_h^{RSS}(x_0)$ is the estimated OR for the hth stratum as in (22). Equation (29) can be written as

$$\hat{\theta}^{SRSS}(x_0) = \frac{\sum_{h=1}^{L} n_h \hat{F}_{X_h}^{RSS}(x_0)\left(1 - \hat{F}_{Y_h}^{RSS}(x_0)\right)}{\sum_{h=1}^{L} n_h \hat{F}_{Y_h}^{RSS}(x_0)\left(1 - \hat{F}_{X_h}^{RSS}(x_0)\right)}. \tag{30}$$

Under assumptions (A1)-(A3) and assume that $d_s \to 0$ and $n_s d_s \to \infty$ as $n_s \to \infty$ and by using similar approximations, as above, the expectation and variance of $\hat{\theta}^{SRSS}(x_0)$ are given by

$$\begin{aligned}
\mathrm{E}\left[\hat{\theta}^{SRSS}(x_0)\right] \approx{}& \frac{\sum_{h=1}^{L} n_h \mathrm{E}\left[\hat{F}_{X_h}^{RSS}(x_0)\right] \mathrm{E}\left[1 - \hat{F}_{Y_h}^{RSS}(x_0)\right]}{\sum_{h=1}^{L} n_h \mathrm{E}\left[\hat{F}_{Y_h}^{RSS}(x_0)\right] \mathrm{E}\left[1 - \hat{F}_{X_h}^{RSS}(x_0)\right]} \\
&- \frac{1}{\left(\sum_{h=1}^{L} n_h \mathrm{E}\left[\hat{F}_{Y_h}^{RSS}(x_0)\right] \mathrm{E}\left[1 - \hat{F}_{X_h}^{RSS}(x_0)\right]\right)^2} \left[\sum_{h=1}^{L} n_h^2 \Big\{ \mathrm{Var}\left(\hat{F}_{X_h}^{RSS}(x_0)\right) \left(\mathrm{E}^2\left[\hat{F}_{Y_h}^{RSS}(x_0)\right]\right.\right. \\
&\left. - \mathrm{E}\left[\hat{F}_{Y_h}^{RSS}(x_0)\right]\right) + \mathrm{Var}\left(\hat{F}_{Y_h}^{RSS}(x_0)\right) \left(\mathrm{E}^2\left[\hat{F}_{X_h}^{RSS}(x_0)\right] - \mathrm{E}\left[\hat{F}_{X_h}^{RSS}(x_0)\right]\right) \\
&\left. + \mathrm{Var}\left(\hat{F}_{Y_h}^{RSS}(x_0)\right) \mathrm{Var}\left(\hat{F}_{X_h}^{RSS}(x_0)\right) \Big\}\right] + \frac{\sum_{h=1}^{L} n_h \mathrm{E}\left[\hat{F}_{X_h}^{RSS}(x_0)\right] \mathrm{E}\left[1 - \hat{F}_{Y_h}^{RSS}(x_0)\right]}{\left(\sum_{h=1}^{L} n_h \mathrm{E}\left[\hat{F}_{Y_h}^{RSS}(x_0)\right] \mathrm{E}\left[1 - \hat{F}_{X_h}^{RSS}(x_0)\right]\right)^3} \\
&\times \sum_{h=1}^{L} n_h^2 \Big\{ \mathrm{Var}\left(\hat{F}_{X_h}^{RSS}(x_0)\right) \mathrm{E}^2\left[\hat{F}_{Y_h}^{RSS}(x_0)\right] + \mathrm{Var}\left(\hat{F}_{Y_h}^{RSS}(x_0)\right) \mathrm{E}^2\left[1 - \hat{F}_{X_h}^{RSS}(x_0)\right] \\
&+ \mathrm{Var}\left(\hat{F}_{Y_h}^{RSS}(x_0)\right) \mathrm{Var}\left(\hat{F}_{X_h}^{RSS}(x_0)\right) \Big\},
\end{aligned} \tag{31}$$

and

$$\begin{aligned}
\operatorname{Var}\left(\hat{\theta}^{SRSS}(x_0)\right) \approx{} & \left\{\frac{\sum_{h=1}^{L} n_h \mathrm{E}\left[\hat{F}_{X_h}^{RSS}(x_0)\right] \mathrm{E}\left[1-\hat{F}_{Y_h}^{RSS}(x_0)\right]}{\sum_{h=1}^{L} n_h \mathrm{E}\left[\hat{F}_{Y_h}^{RSS}(x_0)\right] \mathrm{E}\left[1-\hat{F}_{X_h}^{RSS}(x_0)\right]}\right\}^2 \\
& \times \Bigg\{ \frac{1}{\left(\sum_{h=1}^{L} n_h \mathrm{E}\left[\hat{F}_{Y_h}^{RSS}(x_0)\right] \mathrm{E}\left[1-\hat{F}_{X_h}^{RSS}(x_0)\right]\right)^2} \sum_{h=1}^{L} n_h^2 \Big[\operatorname{Var}\left(\hat{F}_{X_h}^{RSS}(x_0)\right) \mathrm{E}^2\left[\hat{F}_{Y_h}^{RSS}(x_0)\right] \\
& + \operatorname{Var}\left(\hat{F}_{Y_h}^{RSS}(x_0)\right) \mathrm{E}^2\left[1-\hat{F}_{X_h}^{RSS}(x_0)\right] + \operatorname{Var}\left(\hat{F}_{Y_h}^{RSS}(x_0)\right) \operatorname{Var}\left(\hat{F}_{X_h}^{RSS}(x_0)\right)\Big] \\
& + \frac{1}{\left(\sum_{h=1}^{L} n_h \mathrm{E}\left[\hat{F}_{X_h}^{RSS}(x_0)\right] \mathrm{E}\left[1-\hat{F}_{Y_h}^{RSS}(x_0)\right]\right)^2} \sum_{h=1}^{L} n_h^2 \Big[\operatorname{Var}\left(\hat{F}_{Y_h}^{RSS}(x_0)\right) \mathrm{E}^2\left[\hat{F}_{X_h}^{RSS}(x_0)\right] \\
& + \operatorname{Var}\left(\hat{F}_{X_h}^{RSS}(x_0)\right) \mathrm{E}^2\left[1-\hat{F}_{Y_h}^{RSS}(x_0)\right] + \operatorname{Var}\left(\hat{F}_{Y_h}^{RSS}(x_0)\right) \operatorname{Var}\left(\hat{F}_{X_h}^{RSS}(x_0)\right)\Big] \\
& - 2\Bigg\{\left(\sum_{h=1}^{L} n_h \mathrm{E}\left[\hat{F}_{Y_h}^{RSS}(x_0)\right] \mathrm{E}\left[1-\hat{F}_{X_h}^{RSS}(x_0)\right]\right)\Bigg(\sum_{h=1}^{L} n_h \mathrm{E}\left[\hat{F}_{X_h}^{RSS}(x_0)\right] \\
& \times \mathrm{E}\left[1-\hat{F}_{Y_h}^{RSS}(x_0)\right]\Bigg)\Bigg\}^{-1} \Bigg[\sum_{h=1}^{L} n_h^2 \Big\{\operatorname{Var}\left(\hat{F}_{X_h}^{RSS}(x_0)\right) \Big(\mathrm{E}^2\left[\hat{F}_{Y_h}^{RSS}(x_0)\right] \\
& - \mathrm{E}\left[\hat{F}_{Y_h}^{RSS}(x_0)\right]\Big) + \operatorname{Var}\left(\hat{F}_{Y_h}^{RSS}(x_0)\right) \left(\mathrm{E}^2\left[\hat{F}_{X_h}^{RSS}(x_0)\right] - \mathrm{E}\left[\hat{F}_{X_h}^{RSS}(x_0)\right]\right) \\
& + \operatorname{Var}\left(\hat{F}_{Y_h}^{RSS}(x_0)\right) \operatorname{Var}\left(\hat{F}_{X_h}^{RSS}(x_0)\right)\Big\}\Bigg]\Bigg\}. \qquad (32)
\end{aligned}$$

Also, the MSE of $\hat{\theta}^{SRSS}(x_0)$ is given by

$$\operatorname{MSE}\left(\hat{\theta}^{SRSS}(x_0)\right) = \mathrm{E}\left\{\hat{\theta}^{SRSS}(x_0) - \theta(x_0)\right\}^2 = \operatorname{Var}\left(\hat{\theta}^{SRSS}(x_0)\right) + \operatorname{Bias}\left(\hat{\theta}^{SRSS}(x_0)\right)^2. \qquad (33)$$

Remark 2
Since $\hat{O}_s^{SRS}(x) \xrightarrow{P} O_s(x)$, so we can estimate $O(x_0) = \sum_{h=1}^{L} W_h O_h(x_0)$ by $\hat{O}^{SSRS}(x_0) = \sum_{h=1}^{L} W_h \hat{O}_h^{SRS}(x_0)$. In addition, according to $\hat{O}_s^{RSS}(x) \xrightarrow{P} O_s(x)$, we can define that $\hat{O}^{SRSS}(x_0) = \sum_{h=1}^{L} W_h \hat{O}_h^{RSS}(x_0)$ as another estimator of $O(x_0)$. Consequently, using Corollary 1 it can be seen that $\operatorname{MSE}\left[\hat{O}^{SRSS}(x_0)\right] \leq \operatorname{MSE}\left[\hat{O}^{SSRS}(x_0)\right]$, where

$$\operatorname{MSE}\left[\hat{O}^{SSRS}(x_0)\right] = \sum_{h=1}^{L} W_h^2 \operatorname{MSE}\left[\hat{O}_h^{SRS}(x_0)\right], \quad \text{and} \quad \operatorname{MSE}\left[\hat{O}^{SRSS}(x_0)\right] = \sum_{h=1}^{L} W_h^2 \operatorname{MSE}\left[\hat{O}_h^{RSS}(x_0)\right].$$

6. Simulation study

A simulation study is presented to compare the OR estimator's performance based on SRSS with the estimator using SSRS. Since the SRSS estimator may be based on imperfect rankings, we can use one of the models of the imperfect rankings suggested by [6], [14] and [36]. Some of the models, such as fractions of the random rankings model, concomitant model, and fraction inverse rankings, are widely used in literature. In the present paper, we have used the fraction of random rankings model. Based on this model, one can consider the CDF of the jth judgment order statistic as a convex combination of the

CDF of the real jth order statistic with probability λ and the CDF of the underlying population with probability $1-\lambda$, that is,

$$F_{[j]}(x) = \lambda F_{(j)}(x) + (1-\lambda)F(x), \qquad \lambda \in [0,1], \tag{34}$$

where $F_{(j)}(\cdot)$ is the cdf of the true jth order statistic for any $j = 1, 2, ..., r$. It should be highlighted that when $\lambda = 1$, the fraction of the random rankings model is transformed into perfect rankings.

Using (34), we are generated needed data based on the Normal and Gamma distributions as underlying distribution $F(\cdot)$. According to equation (4), we obtained the OR for $L = 3$ strata. For the hth stratum, we should have chosen $F_{Y_h}(\cdot)$ and $F_{X_h}(\cdot)$, for $h = 1, 2, 3$. Based on the Normal distribution, we have considered $N(0.25, 3), N(0, 1)$ as $F_{X_1}(\cdot), F_{Y_1}(\cdot)$, $N(0.75, 2), N(0.5, 2.25)$ for $F_{X_2}(\cdot), F_{Y_2}(\cdot)$ and $N(0.35, 1.5), N(1.5, 3.5)$ as $F_{X_3}(\cdot), F_{Y_3}(\cdot)$, respectively. Moreover, based on the Gamma distribution, we have selected $G(3.8, 1.7), G(2.5, 1.1)$ as $F_{X_1}(\cdot), F_{Y_1}(\cdot)$, $G(4, 1.5), G(1.9, 0.7)$ for $F_{X_2}(\cdot), F_{Y_2}(\cdot)$ and $G(2, 0.9), G(3.5, 1.45)$ as $F_{X_3}(\cdot), F_{Y_3}(\cdot)$, respectively.

To compare MSE $\left(\hat{\theta}^{SRSS}(x_0)\right)$ with MSE $\left(\hat{\theta}^{SSRS}(x_0)\right)$, a simulation with 5000 replications has been run using R software. The Epanechnikov's kernel is applied as kernel function (4). The results are presented in Table 1 for the Normal distribution and Table 2 for the Gamma distribution for different m_h and r_h. Furthermore, x_0 is taken based on some values near to the first, second and third quartiles of X_h and Y_h's underlying distributions. Also, the behavior of imperfect rankings by considering different probabilities of true rankings, (i.e. different values of $\lambda = 0.5, 0.7, 0.9, 1$) is studied and compered with SSRS.
From Tables 1 and 2, the following outcomes can be deduced:

- For fixed values of m_h, r_h and λ, all results deduced based on the Normal distribution have the same behavior in comparing Gamma distribution. This illustrates that symmetry of baseline distribution does not affect the behavior of $\hat{\theta}^{SRSS}(x_0)$.
- For a given r_h, the MSEs of both SRSS and SSRS estimators decrease when m_h increases.
- For fixed m_h, the MSEs of both $\hat{\theta}^{SRSS}(x_0)$ and $\hat{\theta}^{SSRS}(x_0)$ estimators decrease in r_h.
- For given values of m_h and r_h, the MSEs of OR SRSS estimator decrease when λ increases. In other words, when by decreasing the error of rankings, the MSEs of $\hat{\theta}^{SRSS}(x_0)$ decrease and the minimum value of MSE happens when $\lambda = 1$, i.e. in perfect rankings case.
- For fixed strata, the MSEs of both $\hat{\theta}^{SRSS}(x_0)$ and $\hat{\theta}^{SSRS}(x_0)$ estimators are decreasing by increasing sample size (see Figure 2).
- For fixed values of m_h, r_h, x_0 and λ, the kernel-based estimator of OR based on SRSS scheme is more efficient than that of the estimator based on the SSRS.

Tables 1 and 2 exhibit we can survey the behavior of the MSEs in the presence of λ parameter. For this purpose, Figures 1 and 2 are used to present the results obtained graphically. Figure 1 depicts the MSE plots of $\hat{\theta}^{SRSS}(x_0)$ versus $\hat{\theta}^{SSRS}(x_0)$. From Figure 1, it deduces that SRSS estimators are more efficient than SSRS estimator. Furthermore, Figure 2 exhibits the behavior of the MSEs with respect to variations of sample size. The Figure shows that the MSEs decrease when the sample size increases.

In order to achieve more accurate results, a comparative study of the proposed method with its counterpart based on empirical distribution function as the traditional method of estimation is performed. Moreover, the small sample scenario is considered for each stratum and in overall case. The results of the MSE's are exhibited in Table 3 for the Normal and Gamma distributions by considering different values of m_h, r_h and x_0. From Table 3, it can be seen that all the above-mentioned outcomes are also valid for the small samples. Furthermore, the kernel-based estimator is more efficient than the empirical estimator based on both the sampling methods SSRS and SRSS.

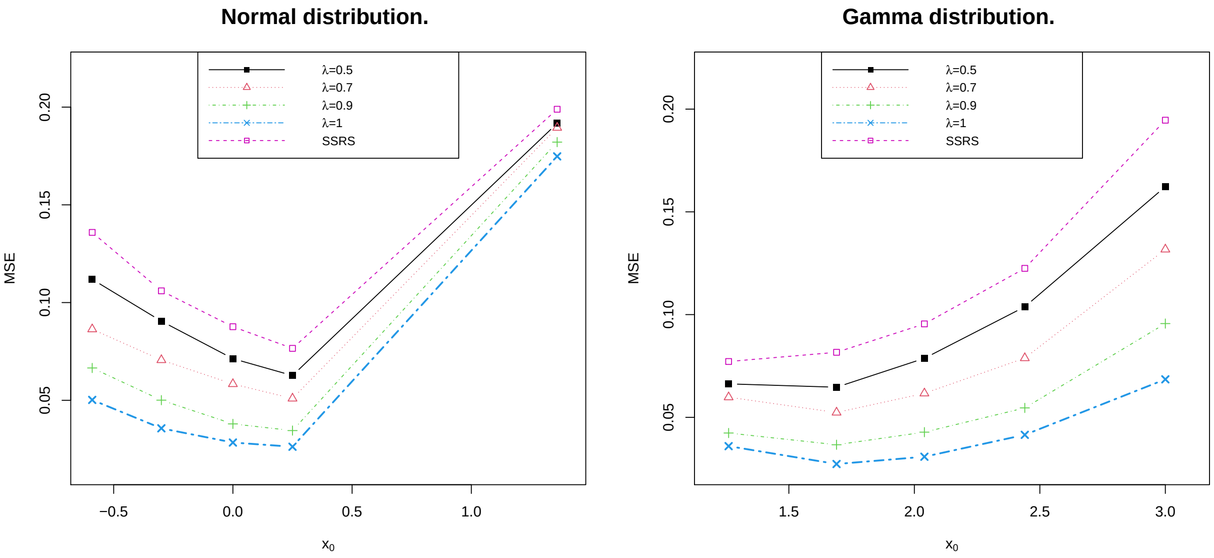

Figure 1. Plots of comparison SSRS with perfect and imperfect SRSS based on MSE for Normal and Gamma distributions.

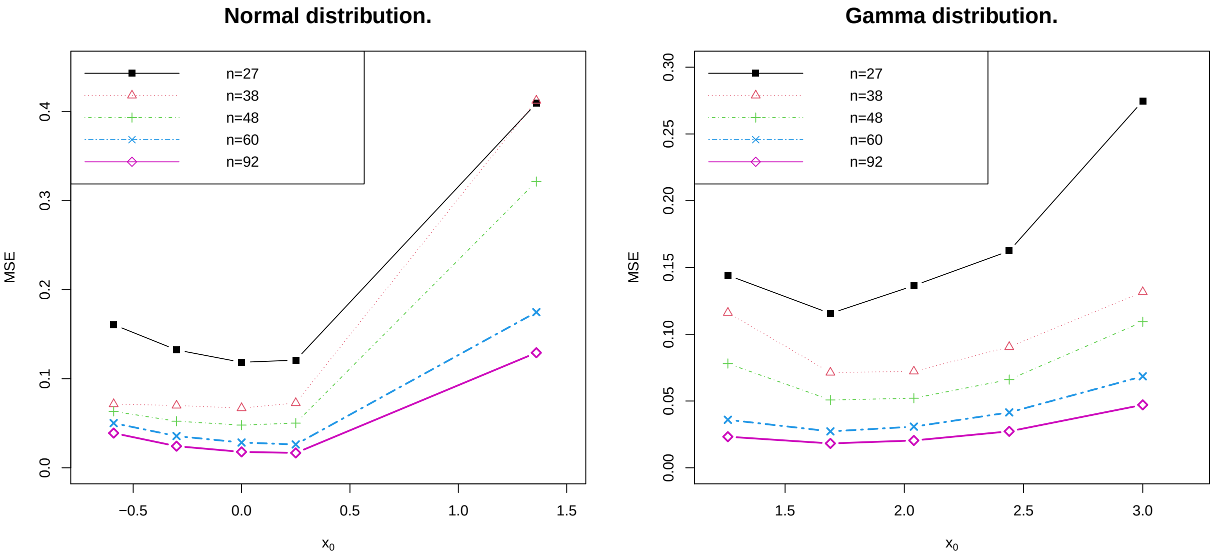

Figure 2. Plots of the behaviour of the MSEs for different values of sample size.

Table 1. The MSEs and bias estimated (in parentheses) for different values of r_h, m_h, x_0 and λ, based on Normal distribution.

				MSE $\left(\hat{\theta}^{SRSS}(x_0)\right)$				MSE $\left(\hat{\theta}^{SSRS}(x_0)\right)$
Distribution	m_h	r_h	x_0	$\lambda = 0.5$	$\lambda = 0.7$	$\lambda = 0.9$	$\lambda = 1$	
Normal	2,3,4	3,3,3	-0.59	0.25972 (-0.05740)	0.22050 (-0.05433)	0.17722 (-0.03339)	0.16076 (-0.01769)	0.29380 (-0.13846)
			-0.30	0.25357 (-0.07488)	0.20611 (-0.06961)	0.15228 (-0.04325)	0.13239 (-0.03297)	0.30080 (-0.09806)
			0.00	0.23190 (-0.03274)	0.18572 (-0.03455)	0.15580 (-0.02477)	0.11851 (-0.00471)	0.29225 (-0.07233)
			0.25	0.23899 (0.00777)	0.19572 (0.00482)	0.13812 (0.02265)	0.12092 (0.03881)	0.26313 (-0.02413)
			1.36	0.54084 (0.44637)	0.47698 (0.45943)	0.43640 (0.47408)	0.41377 (0.48360)	0.56759 (0.41633)
	2,3,4	3,4,5	-0.59	0.15753 (0.01313)	0.12383 (0.01313)	0.09504 (0.02751)	0.07157 (0.03296)	0.22247 (-0.21819)
			-0.30	0.14979 (-0.04584)	0.12328 (-0.04584)	0.08954 (-0.03157)	0.06998 (-0.01898)	0.19939 (-0.06094)
			0.00	0.15644 (-0.05604)	0.13198 (-0.05604)	0.08791 (-0.03239)	0.06707 (-0.02619)	0.20640 (-0.08207)
			0.25	0.16798 (-0.04744)	0.13123 (-0.04744)	0.09199 (-0.01286)	0.07281 (-0.00869)	0.20357 (-0.06098)
			1.36	0.52932 (0.26035)	0.47514 (0.26035)	0.43057 (0.29191)	0.41023 (0.30135)	0.54900 (0.47011)
	4,4,4	3,4,5	-0.59	0.12490 (-0.02987)	0.10838 (-0.01537)	0.07804 (-0.00704)	0.06338 (-0.00519)	0.16185 (-0.03779)
			-0.30	0.12118 (-0.03476)	0.09720 (-0.02637)	0.06910 (-0.01532)	0.05232 (-0.01110)	0.13682 (-0.04368)
			0.00	0.10642 (0.00774)	0.08882 (0.02080)	0.06363 (0.02282)	0.04799 (0.03306)	0.13138 (0.00583)
			0.25	0.10171 (0.06116)	0.08624 (0.07468)	0.06152 (0.08075)	0.05019 (0.08544)	0.13075 (0.04605)
			1.36	0.36303 (0.56322)	0.35104 (0.57765)	0.33044 (0.58495)	0.32154 (0.59338)	0.38319 (0.54785)
	4,4,4	6,5,4	-0.59	0.11190 (-0.10537)	0.08636 (-0.09373)	0.06653 (-0.08350)	0.05019 (-0.07429)	0.13589 (-0.11147)
			-0.30	0.09049 (-0.02218)	0.07060 (-0.02325)	0.05006 (-0.01128)	0.03566 (-0.01425)	0.10599 (-0.04349)
			0.00	0.07130 (0.07569)	0.05833 (0.08639)	0.03791 (0.08836)	0.02838 (0.09659)	0.08764 (0.07195)
			0.25	0.06266 (0.19370)	0.05092 (0.19350)	0.03450 (0.20359)	0.02623 (0.20825)	0.07657 (0.18381)
			1.36	0.19170 (0.91397)	0.18947 (0.91200)	0.18210 (0.92691)	0.17480 (0.93138)	0.19891 (0.90770)
	7,6,5	6,5,4	-0.59	0.07917 (-0.12835)	0.06628 (-0.12252)	0.04959 (-0.11177)	0.03899 (-0.11186)	0.09079 (-0.13910)
			-0.30	0.05769 (-0.02414)	0.04644 (-0.02008)	0.03349 (-0.01606)	0.02429 (-0.01028)	0.07255 (-0.03263)
			0.00	0.04535 (0.11029)	0.03575 (0.11044)	0.02466 (0.11755)	0.01786 (0.12196)	0.05600 (0.10793)
			0.25	0.03789 (0.23920)	0.03094 (0.23690)	0.02247 (0.24802)	0.01670 (0.24788)	0.04723 (0.23646)
			1.36	0.13754 (1.01955)	0.13255 (1.02784)	0.13185 (1.02922)	0.12925 (1.03086)	0.13863 (1.01767)

Table 2. The MSEs and bias estimated (in parentheses) for different values of r_h, m_h, x_0 and λ, based on Gamma distribution.

				MSE $\left(\hat{\theta}^{SRSS}(x_0)\right)$				MSE $\left(\hat{\theta}^{SSRS}(x_0)\right)$
Distribution	m_h	r_h	x_0	$\lambda = 0.5$	$\lambda = 0.7$	$\lambda = 0.9$	$\lambda = 1$	
Gamma	2,3,4	3,3,3	1.26	0.24431 (-0.13365)	0.21083 (-0.12085)	0.16694 (-0.10459)	0.14419 (-0.11129)	0.26904 (-0.13552)
			1.69	0.21114 (-0.14777)	0.17559 (-0.13118)	0.12978 (-0.12641)	0.11553 (-0.11407)	0.29210 (-0.14889)
			2.04	0.24158 (-0.14674)	0.20025 (-0.13369)	0.15750 (-0.11693)	0.13640 (-0.10798)	0.28844 (-0.15221)
			2.44	0.30847 (-0.15416)	0.26225 (-0.14693)	0.20338 (-0.11928)	0.16238 (-0.09436)	0.35534 (-0.16986)
			3.00	0.49225 (-0.13817)	0.39641 (-0.11945)	0.32968 (-0.08794)	0.27466 (-0.07430)	0.56845 (-0.16556)
	2,3,4	3,4,5	1.26	0.19076 (-0.18715)	0.17098 (-0.19475)	0.13324 (-0.16510)	0.11613 (-0.16360)	0.23040 (-0.18302)
			1.69	0.15325 (-0.18907)	0.12796 (-0.17575)	0.09448 (-0.16092)	0.07133 (-0.14971)	0.19634 (-0.19154)
			2.04	0.16426 (-0.17214)	0.14264 (-0.15601)	0.09152 (-0.13825)	0.07220 (-0.13639)	0.20262 (-0.19166)
			2.44	0.20550 (-0.13578)	0.15511 (-0.12975)	0.11019 (-0.10612)	0.09052 (-0.09797)	0.25049 (-0.14894)
			3.00	0.29915 (-0.09216)	0.23793 (-0.07430)	0.16955 (-0.05204)	0.13161 (-0.03946)	0.33164 (-0.10293)
	4,4,4	3,4,5	1.26	0.11749 (-0.18302)	0.10428 (-0.06322)	0.08683 (-0.05097)	0.07808 (-0.05512)	0.13882 (-0.08256)
			1.69	0.10025 (-0.19154)	0.08236 (-0.08903)	0.06243 (-0.08057)	0.05079 (-0.07550)	0.12518 (-0.10154)
			2.04	0.11639 (-0.19166)	0.09045 (-0.09361)	0.06722 (-0.08129)	0.05210 (-0.07853)	0.13958 (-0.09707)
			2.44	0.15283 (-0.14894)	0.11835 (-0.07826)	0.08282 (-0.05865)	0.06608 (-0.06000)	0.17210 (-0.09120)
			3.00	0.21107 (-0.10293)	0.16976 (-0.05476)	0.12914 (-0.03415)	0.10936 (-0.02639)	0.24922 (-0.07658)
	4,4,4	6,5,4	1.26	0.06628 (0.06946)	0.05979 (0.07419)	0.04236 (0.07973)	0.03597 (0.08067)	0.07716 (0.07186)
			1.69	0.06465 (0.01418)	0.05235 (0.01494)	0.03665 (0.02198)	0.02727 (0.02900)	0.08165 (-0.00201)
			2.04	0.07872 (-0.01117)	0.06167 (-0.00320)	0.04278 (-0.00037)	0.03082 (0.00734)	0.09547 (-0.01245)
			2.44	0.10379 (-0.03445)	0.07883 (-0.02237)	0.05458 (-0.01093)	0.04148 (-0.00686)	0.12250 (-0.03934)
			3.00	0.16227 (-0.04436)	0.13172 (-0.04245)	0.09563 (-0.02445)	0.06849 (-0.01251)	0.19462 (-0.05876)
	7,6,5	6,5,4	1.26	0.04125 (0.12158)	0.03469 (0.12271)	0.02704 (0.12910)	0.02328 (0.12910)	0.04970 (0.11726)
			1.69	0.04222 (0.05444)	0.03562 (0.06082)	0.02263 (0.06439)	0.01825 (0.06439)	0.05014 (0.05375)
			2.04	0.04914 (0.02742)	0.04106 (0.02680)	0.02759 (0.03503)	0.02046 (0.03503)	0.05999 (0.02198)
			2.44	0.06413 (-0.00731)	0.05089 (0.00684)	0.03690 (0.01262)	0.02730 (0.01262)	0.07753 (-0.01270)
			3.00	0.10190 (-0.03207)	0.08427 (-0.02011)	0.06008 (-0.01063)	0.04713 (-0.01063)	0.12590 (-0.03709)

Table 3. The MSEs of the empirical and kernel-based estimators of odds ratio based on SSRS and SRSS.

				Empirical estimator					Kernel-based estimator				
					SRSS					SRSS			
Distribution	m_h	r_h	x_0	SSRS	$\lambda = 0.5$	$\lambda = 0.7$	$\lambda = 0.9$	$\lambda = 1$	SSRS	$\lambda = 0.5$	$\lambda = 0.7$	$\lambda = 0.9$	$\lambda = 1$
Normal	1,2,1	4,3,5	-0.59	3.5695	2.9823	2.0300	1.6701	1.2972	1.0431	0.6749	0.5450	0.3524	0.2505
			-0.30	2.9268	2.4868	2.1185	1.1266	0.8694	0.7495	0.6500	0.4365	0.3145	0.2311
			0.00	2.6961	2.0986	1.5387	0.9312	0.6636	0.7765	0.6148	0.4020	0.2538	0.1996
			0.25	2.4695	1.9304	1.1308	0.7029	0.4920	0.7607	0.5238	0.3745	0.2426	0.1711
			1.36	4.1775	2.7385	1.8426	1.5060	1.0934	1.6797	0.9004	0.6534	0.4040	0.3723
	2,3,4	3,3,3	-0.59	1.0030	0.8398	0.8341	0.5009	0.4770	0.2938	0.2443	0.2108	0.1669	0.1441
			-0.30	0.6834	0.5939	0.5546	0.4980	0.3642	0.3008	0.2111	0.1755	0.1297	0.1155
			0.00	0.7512	0.5756	0.4717	0.3773	0.3184	0.2922	0.2415	0.2002	0.1575	0.1364
			0.25	0.6702	0.5147	0.4694	0.3297	0.2852	0.2701	0.2684	0.2622	0.2033	0.1623
			1.36	1.1456	0.9720	0.9680	0.6699	0.5981	0.5675	0.4922	0.3964	0.3296	0.2746
	4,4,4	3,4,5	-0.59	0.3484	0.2948	0.2188	0.1795	0.1532	0.1618	0.1249	0.1083	0.0780	0.0633
			-0.30	0.2995	0.2536	0.1959	0.1581	0.1363	0.1368	0.1211	0.0972	0.0691	0.0523
			0.00	0.2533	0.2189	0.1770	0.1451	0.1167	0.1313	0.1064	0.0888	0.0636	0.0479
			0.25	0.2435	0.2104	0.1766	0.1352	0.1091	0.1307	0.1017	0.0862	0.0615	0.0501
			1.36	0.4809	0.4517	0.4114	0.3697	0.3530	0.3831	0.3630	0.3510	0.3304	0.3215
	7,6,5	6,5,4	-0.59	0.1624	0.1413	0.1233	0.0977	0.0807	0.0907	0.0791	0.0662	0.0495	0.0389
			-0.30	0.1205	0.1016	0.0875	0.0682	0.0563	0.0725	0.0576	0.0464	0.0334	0.0242
			0.00	0.0867	0.0792	0.0647	0.0504	0.0419	0.0560	0.0453	0.0357	0.0246	0.0178
			0.25	0.0753	0.0644	0.0559	0.0414	0.0358	0.0472	0.0378	0.0309	0.0224	0.0167
			1.36	0.1592	0.1546	0.1483	0.1440	0.1397	0.1386	0.1375	0.1325	0.1318	0.1292
Gamma	1,2,1	4,3,5	1.26	2.8514	2.6510	2.6789	1.4205	1.3250	0.6340	0.6171	0.3838	0.2312	0.2133
			1.69	3.5393	2.2837	1.8959	1.1779	0.8976	0.6845	0.4862	0.3488	0.2276	0.1589
			2.04	3.1128	2.5447	1.4779	0.8229	0.5836	0.7205	0.5987	0.4047	0.2579	0.1863
			2.44	2.9367	2.3339	1.8812	1.0518	0.6994	0.9599	0.7769	0.5559	0.3563	0.2564
			3.00	5.5146	3.7958	3.0157	2.2418	2.3222	2.3996	1.7379	1.4236	0.7596	0.5181
	2,3,4	3,3,3	1.26	1.3265	1.4999	1.0352	0.8149	0.8460	0.2690	0.2443	0.2108	0.1669	0.1441
			1.69	0.7757	0.6075	0.5514	0.3865	0.3239	0.2921	0.2111	0.1755	0.1297	0.1155
			2.04	0.7262	0.5409	0.4270	0.3620	0.3041	0.2884	0.2415	0.2002	0.1575	0.1364
			2.44	1.0361	0.6176	0.4855	0.3764	0.3547	0.3553	0.3084	0.2622	0.2033	0.1623
			3.00	1.4659	1.1192	0.8709	0.7516	0.5806	0.5684	0.4922	0.3964	0.3296	0.2746
	4,4,4	3,4,5	1.26	0.4121	0.3835	0.2966	0.2357	0.2019	0.1388	0.1174	0.1042	0.0868	0.0780
			1.69	0.2414	0.2202	0.1875	0.1508	0.1291	0.1251	0.1002	0.0823	0.0624	0.0507
			2.04	0.2430	0.2106	0.1667	0.1285	0.1127	0.1395	0.1163	0.0904	0.0672	0.0521
			2.44	0.2763	0.2348	0.2194	0.1500	0.1306	0.1721	0.1528	0.1183	0.0828	0.0660
			3.00	0.4304	0.3905	0.2969	0.2260	0.2105	0.2492	0.2110	0.1697	0.1291	0.1093
	7,6,5	6,5,4	1.26	0.0980	0.0902	0.0791	0.0643	0.0550	0.0497	0.0412	0.0346	0.0270	0.0232
			1.69	0.0841	0.0721	0.0616	0.0489	0.0414	0.0501	0.0422	0.0356	0.0226	0.0182
			2.04	0.0892	0.0777	0.0651	0.0482	0.0414	0.0599	0.0491	0.0410	0.0275	0.0204
			2.44	0.1041	0.0993	0.0816	0.0604	0.0509	0.0775	0.0641	0.0508	0.0369	0.0273
			3.00	0.1834	0.1537	0.1250	0.1038	0.0854	0.1259	0.1019	0.0842	0.0600	0.0471

7. Illustration using base deficit data

This section's main purpose is to investigate the sampling routine on non-simulated population from the base deficit (BD) data. The base deficit in human physiology means a deficit in the total serum concentration of bicarbonate. It can be indicative of metabolic acidosis or compensatory respiratory alkalosis. The use of base deficit as a guide to volume resuscitation in trauma patients, was first established in 1988 by [10]. Since then, the base deficit has been correlated to many variables in the trauma population, such as mechanism of injury, the presence of intra-abdominal injury, transfusion requirements, mortality, the risk of complications, and the number of days spent in the intensive care unit (see, [35] and [9]). It should be highlighted that [32] obtained ordinary OR of the overall BD data using moving extreme RSS and compared it with SRS based on the empirical distribution function. Here, we obtain the Cochran

Mantel-Haenszel OR of the same data using different stratified sampling schemes based on the CDF's kernel-based estimation.

In this illustration, the SSRS and the SRSS samples are drawn from the data collected based on a retrospective study of the trauma registry at Memorial Health University Medical Center in Savannah, Georgia. All trauma patients were assessed by the trauma team between January, 1998 and May, 2000. The BD data contains different attributes such as BD, age, gender and hospital disposition type (alive or dead) from 4579 patients. After removing missing data, we have 4374 samples included 3125 males and 1222 females. Since females and males have different physiological features, we can consider the entire data set as two strata. The stratified samples from female and male strata are selected with sizes 60 ($r_1 = 5, m_1 = 12$) and 50 ($r_1 = 5, m_1 = 10$), respectively, and computed the Cochran Mantel-Haenszel OR estimate for the 20th to the 90th of quantiles overall BD. By repeating the process for 500 bootstraps resamples with replacement, we compute the OR estimates and corresponding MSE's of the selected values.

Equation (34) is used to select the samples from interested variable based on the imperfect ranking scenario. The results are reported in Table 4. Table 4 includes the following columns: different values of BD (x_0), OR of population $\theta(x_0)$, estimated values of OR (Estimate), mean squared error (MSE) and relative efficiency (RE) which defined as

$$\mathrm{RE} = \frac{\mathrm{MSE}\left(\hat{\theta}^{SSRS}(x_0)\right)}{\mathrm{MSE}\left(\hat{\theta}^{SRSS}(x_0)\right)}.$$

For different values of λ, the RE's values have been presented in Table 4. It observes that by increasing λ, the RE's increase, and the obtained results based on the real data coincide with the presented results in the simulation study.

8. Conclusion

In this paper, the kernel-based estimators for the Cochran Mantel-Haenszel odds ratio of the underlying population were suggested based on the SSRS and SRSS. The kernel function with optimal bandwidth considered in the present paper. Moreover, the closed form for the expectation, variance and MSE of the odds and odds ratio estimators obtained for both SSRS and SRSS. Further, we showed analytically that the SRSS kernel-based estimator of the odds has better performance than that of the SSRS counterpart. A simulation study performed to compare these estimators based on the bias and MSE criterions using different sample sizes of strata and in overall. The effect of imperfect rankings on the performance of the proposed estimators discussed. Moreover, the proposed estimator compered with its traditional counterpart, the empirical estimator. We found that our proposed estimators using SRSS still more efficient than using SSRS in all cases. Finally, the performance of the estimators illustrated by using the base deficit data set. According to the obtained results, we recommended that to use SRSS kernel-based estimator, whenever possible, to get more accurate estimates of the odds and odds ratio of biomarkers.

Table 4. The Cochran Mantel-Haenszel OR estimates and corresponding MSEs for BD values based on 500 resamples.

		SSRS			SRSS($\lambda = 0.5$)		SRSS($\lambda = 0.7$)			SRSS($\lambda = 0.9$)		SRSS($\lambda = 1.0$)			
$BD(x_0)$	$\theta(x_0)$	Estimate	MSE	Estimate	MSE	RE	Estimate	MSE	RE	Estimate	MSE	RE	Estimate	MSE	RE
-5.00	8.88	9.0655	7.2644	9.0102	6.2006	1.1716	8.9897	5.3813	1.3499	8.8882	3.3877	2.1443	8.8827	3.2139	2.2603
-3.60	7.14	6.8459	3.3398	6.6660	2.9721	1.1237	6.6977	2.4306	1.3741	6.5748	1.9338	1.7271	6.5308	1.8117	1.8435
-2.70	5.59	5.4109	1.9203	5.3825	1.8959	1.0129	5.3904	1.3553	1.4169	5.3172	1.0104	1.9005	5.3042	0.9016	2.1299
-2.00	4.94	4.5605	1.4200	4.4794	1.3857	1.0248	4.4274	1.2209	1.1631	4.4537	0.9286	1.5292	4.4169	0.8783	1.6168
-1.00	3.88	3.4578	0.9508	3.4528	0.8779	1.0830	3.4118	0.8530	1.1147	3.3596	0.7176	1.3250	3.3509	0.6163	1.5428
0.00	2.48	2.5706	0.6020	2.5397	0.5813	1.0356	2.6051	0.4571	1.3170	2.5609	0.3339	1.8029	2.5183	0.2855	2.1086
0.60	2.26	2.1430	0.4814	2.1032	0.3829	1.2572	2.1613	0.3558	1.3530	2.1192	0.2924	1.6464	2.0932	0.2874	1.6750
2.00	1.36	1.3008	0.3189	1.3217	0.3084	1.0340	1.2844	0.2634	1.2107	1.3289	0.1960	1.6270	1.3068	0.1885	1.6917

Appendix

To obtain the expectation and variance of $\hat{\theta}^{SSRS}(x_0)$, let $U = \sum_{h=1}^{L} n_h \hat{F}_{X_h}^{SRS}(x_0) \left(1 - \hat{F}_{Y_h}^{SRS}(x_0)\right)$ and $V = \sum_{h=1}^{L} n_h \hat{F}_{Y_h}^{SRS}(x_0) \left(1 - \hat{F}_{X_h}^{SRS}(x_0)\right)$. In addition, the expectation and variance for ratio of two random variables U and V are given by ([22, p.181])

$$\mathrm{E}\left[\frac{U}{V}\right] \approx \frac{\mathrm{E}(U)}{\mathrm{E}(V)} - \frac{\mathrm{Cov}(U,V)}{\mathrm{E}^2(V)} + \frac{\mathrm{E}(U)}{\mathrm{E}^3(V)}\mathrm{Var}(V), \tag{35}$$

and

$$\mathrm{Var}\left[\frac{U}{V}\right] \approx \left[\frac{\mathrm{E}(U)}{\mathrm{E}(V)}\right]^2 \left\{\frac{\mathrm{Var}(U)}{\mathrm{E}^2(U)} + \frac{\mathrm{Var}(V)}{\mathrm{E}^2(V)} - \frac{2\mathrm{Cov}(U,V)}{\mathrm{E}(U)\mathrm{E}(V)}\right\}. \tag{36}$$

On the other hand, we have

$$\begin{aligned} \mathrm{E}(U) &= \sum_{h=1}^{L} n_h \mathrm{E}\left[\hat{F}_{X_h}^{SRS}(x_0)\right] \mathrm{E}\left[1 - \hat{F}_{Y_h}^{SRS}(x_0)\right], \\ \mathrm{E}(V) &= \sum_{h=1}^{L} n_h \mathrm{E}\left[\hat{F}_{Y_h}^{SRS}(x_0)\right] \mathrm{E}\left[1 - \hat{F}_{X_h}^{SRS}(x_0)\right], \end{aligned} \tag{37}$$

$$\begin{aligned} \mathrm{Var}(U) = \sum_{h=1}^{L} n_h^2 \Bigg[&\mathrm{Var}\left(\hat{F}_{Y_h}^{SRS}(x_0)\right) \mathrm{E}^2\left[\hat{F}_{X_h}^{SRS}(x_0)\right] + \mathrm{Var}\left(\hat{F}_{X_h}^{SRS}(x_0)\right) \mathrm{E}^2\left[1 - \hat{F}_{Y_h}^{SRS}(x_0)\right] \\ &+ \mathrm{Var}\left(\hat{F}_{Y_h}^{SRS}(x_0)\right) \mathrm{Var}\left(\hat{F}_{X_h}^{SRS}(x_0)\right) \Bigg], \\ \mathrm{Var}(V) = \sum_{h=1}^{L} n_h^2 \Bigg[&\mathrm{Var}\left(\hat{F}_{X_h}^{SRS}(x_0)\right) \mathrm{E}^2\left[\hat{F}_{Y_h}^{SRS}(x_0)\right] + \mathrm{Var}\left(\hat{F}_{Y_h}^{SRS}(x_0)\right) \mathrm{E}^2\left[1 - \hat{F}_{X_h}^{SRS}(x_0)\right] \\ &+ \mathrm{Var}\left(\hat{F}_{Y_h}^{SRS}(x_0)\right) \mathrm{Var}\left(\hat{F}_{X_h}^{SRS}(x_0)\right) \Bigg], \end{aligned} \tag{38}$$

and

$$\begin{aligned} \mathrm{Cov}(U,V) = \sum_{h=1}^{L} n_h^2 \Bigg\{ &\mathrm{Var}\left(\hat{F}_{X_h}^{SRS}(x_0)\right) \left(\mathrm{E}^2\left[\hat{F}_{Y_h}^{SRS}(x_0)\right] - \mathrm{E}\left[\hat{F}_{Y_h}^{SRS}(x_0)\right]\right) \\ &+ \mathrm{Var}\left(\hat{F}_{Y_h}^{SRS}(x_0)\right) \left(\mathrm{E}^2\left[\hat{F}_{X_h}^{SRS}(x_0)\right] - \mathrm{E}\left[\hat{F}_{X_h}^{SRS}(x_0)\right]\right) \\ &+ \mathrm{Var}\left(\hat{F}_{Y_h}^{SRS}(x_0)\right) \mathrm{Var}\left(\hat{F}_{X_h}^{SRS}(x_0)\right) \Bigg\}. \end{aligned} \tag{39}$$

Finally, by substituting (37)-(39) in (35) and (36) the results are obtained.

REFERENCES

[1] A. Agresti and M. Kateri. Categorical Data Analysis. Springer, 2011.

[2] M. Al-Odat and M. F. Al-Saleh. A variation of ranked set sampling. Journal of Applied Statistical Science, 10(2):137–146, 2001.

[3] N. Altman and C. Léger. Bandwidth selection for kernel distribution function estimation. Journal of Statistical Planning and Inference, 46(2):195–214, 1995.

[4] A. Azzalini. A note on the estimation of a distribution function and quantiles by a kernel method. Biometrika, 68(1):326–328, 1981.

[5] L. Barabesi and L. Fattorini. Kernel estimators of probability density functions by ranked-set sampling. Communications in Statistics-Theory and Methods, 31(4):597–610, 2002.

[6] L. L. Bohn and D. A. Wolfe. The effect of imperfect judgment rankings on properties of procedures based on the ranked-set samples analog of the mann-whitney-wilcoxon statistic. Journal of the American Statistical Association, 89(425):168–176, 1994.

[7] Z. Chen. Density estimation using ranked-set sampling data. Environmental and Ecological Statistics, 6(2):135–146, 1999.

[8] Z. Chen, N.-Z. Shi, and W. Gao. Nonparametric estimation of the log odds ratio for sparse data by kernel smoothing. Statistics & Probability Letters, 81(12):1802–1807, 2011.

[9] J. W. Davis, R. C. Mackersie, T. L. Holbrook, and D. B. Hoyt. Base deficit as an indicator of significant abdominal injury. Annals of Emergency Medicine, 20(8):842–844, 1991.

[10] J. W. Davis, S. R. Shackford, R. C. Mackersie, and D. B. Hoyt. Base deficit as a guide to volume resuscitation. The Journal of Trauma, 28(10):1464–1467, 1988.

[11] A. Eftekharian and M. Razmkhah. On estimating the distribution function and odds using ranked set sampling. Statistics & Probability Letters, 122:1–10, 2017.

[12] A. Eftekharian and H. Samawi. On kernel-based quantile estimation using different stratified sampling schemes with optimal allocation. Journal of Statistical Computation and Simulation, 91(5):1040–1056, 2021.

[13] A. Franke and G. Osius. The asymptotic covariance matrix of the odds ratio parameter estimator in semiparametric log-bilinear odds ratio models. Journal of Statistical Planning and Inference, 143(1):63–81, 2013.

[14] J. C. Frey. New imperfect rankings models for ranked set sampling. Journal of Statistical planning and Inference, 137(4):1433–1445, 2007.

[15] Y. Huang, J. Yin, and H. Samawi. Methods improving the estimate of diagnostic odds ratio. Communications in Statistics-Simulation and Computation, 47(2):353–366, 2016.

[16] M. Lejeune and P. Sarda. Smooth estimators of distribution and density functions. Computational Statistics & Data Analysis, 14(4):457–471, 1992.

[17] J. Lim, M. Chen, S. Park, X. Wang, and L. Stokes. Kernel density estimator from ranked set samples. Communications in Statistics-Theory and Methods, 43(10-12):2156–2168, 2014.

[18] I. Locatelli and V. Rousson. Assessing interrater agreement on binary measurements via intraclass odds ratio. Biometrical Journal, 58(4):962–973, 2016.

[19] K.-J. Lui and K.-C. Chang. Notes on odds ratio estimation for a randomized clinical trial with noncompliance and missing outcomes. Journal of Applied Statistics, 37(12):2057–2071, 2010.

[20] V. Mandowara and N. Mehta. Modified ratio estimators using stratified ranked set sampling. Hacettepe Journal of Mathematics and Statistics, 43(3):461–471, 2014.

[21] G. McIntyre. A method for unbiased selective sampling, using ranked sets. Crop and Pasture Science, 3(4):385–390, 1952.

[22] A. Mood, F. Graybill, and D. Boes. Introduction to The Theory of Statistics. McGraw-hill, 1974.

[23] H. Muttlak. Median ranked set sampling. Journal of Applied Statistical Sciences, 6(4):245–255, 1997.

[24] E. A. Nadaraya. Some new estimates for distribution functions. Theory of Probability & Its Applications, 9(3):497–500, 1964.

[25] A. M. Polansky and E. R. Baker. Multistage plug-in bandwidth selection for kernel distribution function estimates. Journal of Statistical Computation and Simulation, 65(1-4):63–80, 2000.

[26] D. Rahardja, H. Wu, S. Huang, and J. Chu. Confidence intervals for the odds ratio of two independent binomial proportions using data with one type of misclassification. Journal of Statistics and Management Systems, 19(2):259–268, 2016.

[27] M. Rosenblatt. Remarks on some nonparametric estimates of a density function. The Annals of Mathematical Statistics, 27(3):832–837, 1956.

[28] H. Samawi, A. Chatterjee, J. Yin, and H. Rochani. On kernel density estimation based on different stratified sampling with optimal allocation. Communications in Statistics-Theory and Methods, 46(22):10973–10990, 2017.

[29] H. Samawi, A. Chatterjee, J. Yin, and H. Rochani. On quantiles estimation based on different stratified sampling with optimal allocation. Communications in Statistics-Theory and Methods, 48(6):1529–1544, 2018.

[30] H. M. Samawi. Stratified ranked set sample. Pakistan Journal of Statistics-All Series, 12:9–16, 1996.

[31] H. M. Samawi, M. S. Ahmed, and W. Abu-Dayyeh. Estimating the population mean using extreme ranked set sampling. Biometrical Journal, 38(5):577–586, 1996.

[32] H. M. Samawi and M. F. Al-Saleh. Valid estimation of odds ratio using two types of moving extreme ranked set sampling. Journal of the Korean Statistical Society, 42(1):17–24, 2013.

[33] H. M. Samawi and M. I. Siam. Ratio estimation using stratified ranked set sample. Metron, 61(1):75–90, 2003.

[34] S. L. Stokes and T. W. Sager. Characterization of a ranked-set sample with application to estimating distribution functions. Journal of the American Statistical Association, 83(402):374–381, 1988.

[35] L. N. Tremblay, D. V. Feliciano, G. S. Rozycki, and J. A. Morris Jr. Assessment of initial base deficit as a predictor of outcome: Mechanism of injury does make a difference/discussion. The American surgeon, 68(8):689, 2002.

[36] M. Vock and N. Balakrishnan. A jonckheere–terpstra-type test for perfect ranking in balanced ranked set sampling. Journal of Statistical Planning and Inference, 141(2):624–630, 2011.

[37] W. Wang and G. Shan. Exact confidence intervals for the relative risk and the odds ratio. Biometrics, 71(4):985–995, 2015.

[38] J. Yin, Y. Hao, H. Samawi, and H. Rochani. Rank-based kernel estimation of the area under the roc curve. Statistical Methodology, 32:91–106, 2016.

[39] M. Zhao, Y. Zhao, and I. W. McKeague. Empirical likelihood inference for the odds ratio of two survival functions under right censoring. Statistics & Probability Letters, 107:304–312, 2015.