Judgment Post-stratified Assessment Combining Ranking Information from Multiple Sources, with a Field Phenotyping Example

Omer OZTURK and Olena KRAVCHUK

This paper presents novel estimators for a judgment post-stratified (JPS) sample, which combine the ranking information from different methods or rankers. A JPS sample divides the units in the original simple random sample (SRS) into several ranking groups based on the relative positions (ranks) of the units in their individual small comparison sets. Ranks in the comparison sets may be assigned with several different ranking procedures. When considered separately, each ranking method leads to a different JPS sample estimator of the population mean or total. Here we introduce equally or unequally weighted estimators, which combine the ranking information from multiple sources. The unequal weights utilize the standard errors of the individual ranking methods estimators. The weighted estimators provide a substantial improvement over an SRS estimator and a JPS estimator based on a single ranking method. The new estimators are applied to crop establishment phenotypic data from an agricultural field experiment.

Supplementary materials accompanying this paper appear online.

Key Words: Agreement scores; Meta-analysis; Multi-ranker model; Ranked-set sampling; Jackknife estimator.

1. INTRODUCTION

A stratified sample combines probability samples (e.g., SRS) from each stratum in a population. The total variation in a stratified sample can thus be partitioned into within- and between-stratum variation. In certain settings, it may not be straightforward to stratify the population if the stratification variable is subjective, such as visual inspection, and not available for all population units at once. In such cases, the stratification principle can be used at the sample level. MacEachern et al. (2004) introduced the concept of the JPS sample

---

Omer Ozturk (✉)Department of Statistics, The Ohio State University, 1958 Neil Avenue, Columbus, OH 43210, USA (E-mail: omer@stat.osu.edu). O. KravchukSchool of Agriculture, Food and Wine, University of Adelaide, Adelaide, Australia.

© 2021 The Author(s)

Journal of Agricultural, Biological, and Environmental Statistics, Volume 26, Number 3, Pages 329–348
<https://doi.org/10.1007/s13253-021-00439-1>

to improve the information content of an SRS by post-stratifying based on a (subjective) ranking variable.

To construct a JPS sample in an infinite population, one first selects an SRS sample and measures each unit in the sample for the characteristic of interest. For each unit, $H - 1$ additional units are then selected, without measurement, at random from the population to form a comparison set of size H . Relative positions (ranks) of every unit in the original SRS are determined in their comparison sets using a ranking procedure on an auxiliary variable, such as visual ranking. Under a consistent ranking procedure, regardless of its quality, the JPS sampling yields unbiased estimators for the population mean and total (Ozturk 2016a).

The main difference between the stratified and JPS sampling procedures is that the former combines SRSs selected across all strata, while the latter partitions a single SRS from the population into ranking groups post-experimentally, based on the ranking information from additional comparison sets. The stratum sample sizes are constant in the stratified sampling, while the judgment ranking group sample sizes are random in the JPS sampling. The JPS sample is similar to a post-stratified (PS) sample, with a clear distinction. The PS sample is stratified based on the population strata membership information, while the JPS sample is segregated into the judgment (ranking) groups based on the position (ranking) information in small comparison sets. Hence, a JPS sample can be used for populations for which no strata membership is available.

In the context of modern digital plant research (Araus and Cairns 2014), JPS sampling presents a range of potential applications in agricultural and environmental research and practice. We provide here a motivating example of the field estimation of early crop establishment of faba bean in an agronomic experiment conducted in the University of Adelaide, Australia (Kasprzak 2021). The aim was to compare the crop performance, including faba bean emergence, under a range of planting densities, controlled by the rate of sowing and row spacing. It is believed better early crop establishment leads to better yield and profitability. The study measured several variables, including the seedling emergence observed over the first two weeks after sowing. An expert human observer would typically inspect five random transects of 1m length in an experimental plot. For automated observations, the entire research field was divided into 2640 grid cells (approximately 50 cm long and 25 cm wide) for aerial photograph observations with a drone (DJI's Phantom 4 Pro v2.0) in addition to the observations by the human observer. The drone had an on-board camera and recorded the image of each cell. These images were later processed using a machine learning algorithm to predict the actual number of seedlings emerged.

For the purposes of this paper, we consider the data collected in a single year of the trial over the entire field: (1) the manual counts by the expert observer of seedling emergence (Y) on each grid cell approximately two weeks after sowing; (2) the automated counts X predicted from the image data. The ranking variable X is highly correlated with Y , $\rho_{Y,X} = 0.766$. The lack of perfect correspondence is explained by the uncertainty in the manual and automated counts due to the presence of ground cover (stubble) concealing the dusty green faba beans seedlings.

In the seedling emergence data, the auxiliary variable X can be used as a ranking variable for JPS. Even though the X values would be different from the Y values, they provide reasonably accurate ranks for faba bean emergence on the grid cells selected for comparison sets.

In a JPS sample, rank classes for individual units are assigned after the simple random sample is collected. In certain settings, it is possible to have more than one set of ranks for the same measured unit in the SRS. Ranks can be constructed with different ranking mechanisms. To illustrate this, in our example of the faba bean emergence, the ranks of Y on each measured unit can be obtained K , $K \geq 1$, times, with a minimal ranking cost, by constructing K different comparison sets (simply selecting grid cells at random from automated images) for each Y -value. In a different setting, units in each comparison set may be ranked by K rankers using different visual inspection or assessment protocols, or by K machine learning algorithms requiring, say, different amounts of training. In some other cases, for the n units selected and measured for a JPS, $n(H - 1)$ units can be additionally selected in the population, and the same person may visually rank the units in K different comparison sets of size H for each unit in the JPS by permuting the $n(H - 1)$ unmeasured units into n comparison sets K times.

The ranking information plays a significant role in the construction of a JPS sample. In a post-stratified sample (PS), the observations are binned based on their membership in stratum populations. However, knowledge of the strata membership may be too strong an assumption. When the membership information is not available for all population units, position information may be readily available in a small set of units in comparison sets. This position information is used in JPS sampling to create judgment groups of homogeneous observations. Each measured unit in a JPS sample possesses two pieces of information: the value of the characteristic of interest Y , and the relative position of the unit among the other $H - 1$ unmeasured units in its comparison set. The latter presents additional information through the ranks of the measured units. If the ranking quality is good and H is large, the information content of a JPS sample can be substantially higher than the information content of a PS sample of the same size.

The problem of reducing the impact of ranking error on the estimators in the JPS sampling design for an infinite population has generated extensive research. Wang et al. (2006) used the concomitants order statistics to estimate the population mean. Frey (2016) and Frey and Feeman (2012, 2013) constructed estimators for the population mean and variance by conditioning on the judgment group samples sizes. These new estimators improve the unconditional JPS estimators. Judgment ranks induce a stochastic ordering among judgment group means in a JPS sample. Chen et al. (2014), Frey and Ozturk (2011), Wang et al. (2008, 2012), and Stokes et al. (2007) constructed constrained estimators using isotonic regression under the restriction of stochastic ordering among judgment ranking group means. Judgment group sample sizes are random variables in a JPS sample. Hence, sample sizes for some judgment groups may be zero. Ozturk (2017) constructed conditional ranks with smaller set sizes by conditioning on the original ranks in a JPS sample. He showed that these conditional ranks in smaller comparison sets reduce the impact of empty judgment classes. The main approach in the aforementioned papers is either to average the ranking information in the sample or to smooth it through isotonic regression. Omidvar et al. (2018) estimated the prevalence of osteoporosis using a JPS sample in a finite mixture model. Zamanzade and Wang (2017) constructed several estimators for the population proportion using a JPS sample.

Recently, there has been increased research on JPS sampling in the finite population setting. [Ozturk \(2016a\)](#); [Ozturk \(2016b\)](#); [Ozturk \(2019a\)](#) used design-based inference to construct JPS estimators for the population mean and total. The JPS sample can be constructed under sampling with or without replacement. It has been shown that the estimator of the variance of the sample mean requires a finite population correction factor in sampling without replacement. [Ozturk \(2019b\)](#) constructed a JPS sample using probability proportional-to-size sampling to assign higher selection probabilities to important units in the population. In a model-based inference, [Ozturk \(2018a\)](#); [Ozturk and Bayramoglu Kavlak \(2018b\)](#) constructed predictors for the population mean and total using a super-population model.

Regardless of how the ranks are created, the ranking information from multiple consistent sources can be combined to draw statistical inference. If the ranking cost is negligible, combining sets of ranks improves the information content of a JPS sample without impacting its sampling cost. In a slightly different context, [Ozturk \(2013\)](#) and [Ozturk and Demirel \(2016\)](#) used the tie structures among K rankers to construct a ranked set sample and estimate the population mean and variance. [Stokes et al. \(2007\)](#) used the raking principle in contingency tables to incorporate ranking information into a JPS sample.

This paper adds to the JPS research for finite populations and presents new estimators of the population mean and total which combine the ranking information obtained from different sources in a JPS design. The design-based inference is presented in detail, and the model-based inference is also provided in the Supplementary Material. Section 2 details the construction of JPS sampling designs *with* and *without* replacement, D_1 and D_2 . In design D_1 , units in a comparison set are selected without replacement. The units are ranked, and the rank of the measured unit is identified. Once the rank is identified, all units, including the measured ones, are returned to the population before constructing the next comparison set, and the process is repeated till all the units in the JPS are assigned their judgment rank. In this design, the same unit can appear in the sample more than once, and all observations in the sample are independent. Hence, this sample is equivalent to a JPS sample constructed from an infinite population. The design D_2 is similar to design D_1 . The main difference is that all units in the comparison set are removed from the population, including both the measured and unmeasured ones, before selecting the next comparison set. The designs D_1 and D_2 induce different kinds of the correlation structure among the measured observations in a JPS sample. All observations in D_1 are independent, while observations in D_2 are negatively correlated. Section 2 also provides a brief review of the known results for the distributional properties of the sample mean of the JPS sampling design. Section 3 introduces equally and unequally weighted estimators, which combine the ranking information from K different sets of ranks. The unequal weights depend on the standard errors of the JPS estimators based on individual ranking sources. Section 4 provides the variance estimator for the weighted estimators of the population mean. Section 5 investigates the empirical properties of the proposed estimators. Section 6 applies the methodology to the faba bean seedling emergence data to estimate the average density of seedlings. Section 7 provides some concluding remark. Supplementary Material presents developments for model-based inference and agreement-weight estimators.

2. SAMPLING DESIGNS

In this paper, we assume the finite population setting unless specified otherwise. The finite population setting under sampling with replacement also covers the JPS sampling designs in the infinite population setting. A finite population of size N will be denoted by $\mathcal{P}$. Let Y be the variable of interest. The values of Y on population units will be denoted by $y_1, \dots, y_N$. The population mean and variance are defined as follows

$$\mu = \frac{1}{N} \sum_{i=1}^N y_i, \quad \sigma^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \mu)^2.$$

Design D_1 : we first select an SRS with replacement and measure each of the n units in the sample, $Y_1, \dots, Y_n$. For each unit (Y_i), we then select additional $H - 1$ units under sampling without replacement from the population to form a comparison set $\{Y_i, Y_1^*, \dots, Y_{H-1}^*\}$, where $Y_h^* \neq Y_i, h = 1, \dots, H - 1$. We rank all units in the comparison set from smallest to largest based on the perceived value of the characteristic Y (or some other ranking information available) and identify the rank of Y_i , $R_{i,1}$. All units in the comparison set, including the one we measured initially, are returned to the population before the construction of the next comparison set. A population size of at least H units is required to complete the design. This process creates the following JPS sample

$$D_1 = \{Y_i, R_{i,1}\}; i = 1, \dots, n.$$

Design D_2 : the JPS sample is constructed in a fashion similar to D_1 . We first select an SRS of size n without replacement, and measure all the items, $Y_1, \dots, Y_n$. For each measured item (Y_i), we construct a comparison set, $\{Y_i, Y_1^*, \dots, Y_{H-1}^*\}$ with the constraint that $Y_h^* \neq Y_i, h = 1, \dots, H - 1$, and identify the ranks ($R_{i,2}$) for $i = 1, \dots, n$. In this design, all units in a comparison set are removed from the population before selecting a comparison set for another item. Hence, all comparison sets are disjoint, and a population size of at least Kn units is required. The JPS sample created is

$$D_2 = \{Y_i, R_{i,2}\}; i = 1, \dots, n.$$

We note that samples D_1 and D_2 have different distributional properties. If the population size is sufficiently large, the pairs in sample D_2 become approximately independent, and samples D_1 and D_2 are approximately equivalent.

The conditional mean and variance of Y_j given $R_{j,r} = h$, and the conditional covariance of Y_j, Y_t given that $R_{j,r} = h, R_{t,r} = h'$ will be denoted by

$$\begin{aligned} \mu_{[h],r} &= E(Y_{[h],r}) = E(Y_j | R_{j,r} = h), \text{ for } r = 1, 2 \\ \sigma_{[h],r}^2 &= \text{Var}(Y_{[h],r}) = \text{Var}(Y_j | R_{j,r} = h) \text{ for } r = 1, 2 \end{aligned}$$

and

$$\sigma_{[h,h'],2} = \text{cov}(Y_{[h],2}, Y_{[h'],2}) = \text{cov}(Y_j, Y_t | R_{j,2} = h, R_{t,2} = h') \text{ for } r = 2.$$

For design D_1 , $\sigma_{[h,h'],1}^2 = 0$ since the observations are independent.

In the expressions above, the square brackets correspond to judgment class statistics. Under perfect ranking, the brackets can be replaced with round parentheses corresponding to order statistics.

Unbiased estimators of the population mean and total based on samples D_1 and D_2 are given by

$$\bar{Y}_{\text{JPS},r} = \sum_{h=1}^H \frac{I_{h,r} J_{h,r}}{d_{n,r}} \sum_{i=1}^n I(R_{i,r} = h) Y_i, \quad r = 1, 2,$$

where the subscript r denotes that the JPS sample is generated by design D_r and

$$J_{h,r} = \begin{cases} 1/n_{h,r} & n_{h,r} > 0, \\ 0 & n_{h,r} = 0, \end{cases} \quad n_{h,r} = \sum_{i=1}^n I(R_{i,r} = h), \quad I_{h,r} = I(n_{h,r} > 0), \quad d_{n,r} = \sum_{h=1}^H I_{h,r}.$$

The estimator $\bar{Y}_{\text{JPS},r}$ can be written in a slightly different form as a weighted average of the judgment class means

$$\bar{Y}_{\text{JPS},r} = \sum_{h=1}^H w_{h,r} \bar{Y}_h, \quad \bar{Y}_h = J_{h,r} \sum_{i=1}^n I(R_{i,r} = h) Y_i, \quad w_{h,r} = I_{h,r} / d_{n,r},$$

where the weights $w_{h,r}$ make the estimator unbiased for any set size H and sample size n . In JPS sampling, ranks have a discrete uniform distribution with the support on integers $1, \dots, H$. Hence, $n_{h,r}$, $d_{n,r}$, and $w_{h,r}$ are all random variables since they are functions of ranks $R_{i,r}$, $i = 1, \dots, n$. The sample size vector of judgment class groups, $\mathbf{n}_r = (n_{1,r}, \dots, n_{H,r})$, has a multinomial distribution with the number of trials n and the success probability vector $(1/H, \dots, 1/H)$. The following theorem provides some useful results for these random variables.

Lemma 1. *The following equalities hold for the distribution of $\frac{I_{h,r}}{d_{n,r}}$, $h = 1, \dots, H$, $r = 1, 2$ (Ozturk 2014; Dastbaravarde et al. 2016):*

- $E\left(\frac{I_{1,r}}{d_{n,r}}\right) = 1/H$
- $E\left(\frac{I_{1,r}}{d_{n,r}^2}\right) = \frac{1}{H^2} \sum_{k=1}^H \left(\frac{k}{H}\right)^{n-1}$
- $\text{Var}\left(\frac{I_{1,r}}{d_{n,r}}\right) = \frac{1}{H^2} \sum_{k=1}^{H-1} \left(\frac{k}{H}\right)^{n-1}$
- $\text{cov}\left(\frac{I_{1,r}}{d_{n,r}}, \frac{I_{2,r}}{d_{n,r}}\right) = -\frac{1}{H-1} \text{var}\left(\frac{I_1}{d_{n1}}\right)$
- $E\left(\frac{I_{1,r}^2 J_{1,r}}{d_{n,r}^2}\right) = \frac{1}{H^n} \left(\frac{1}{n} + \sum_{k=2}^H \sum_{j=1}^{k-1} \sum_{t=1}^{n-k+1} \frac{(-1)^{j-1}}{k^{2t}} \binom{H-1}{k-1} \binom{k-1}{j-1} \binom{n}{t} (k-j)^{n-t}\right).$

We note that the expected values, variances and covariances in Lemma 1 do not depend on the values of Y on population units. They can be computed for any given sample and set sizes. We do not assume that the assigned rank $R_{i,r}$ is the same as the actual rank of Y_i

in its comparison set, but we require a consistency in the ranking procedure. The ranking procedure is called consistent if it satisfies the following equality

$$\frac{1}{H} \sum_{h=1}^H E \{Y_i | I(R_{i,r} = h)\} = E(Y_i) \text{ for } r = 1, 2.$$

The consistency of a ranking procedure in JPS sampling holds since $P(R_{i,r} = h) = 1/H$ for $h = 1, \dots, H$. For convenience, we will consider ranking in the ascending order throughout the paper.

Ozturk (2016a) showed that the estimator $\bar{Y}_{\text{JPS},r}$ is unbiased for the population mean under a consistent ranking scheme and provided a closed form expression for its variance under designs D_1 and D_2 . The following theorem is stated for easy access in the remainder of the paper.

Theorem 1. Let $(Y_i, R_i); i = 1, \dots, n$, be a JPS sample of set size H constructed under a consistent ranking scheme based on either design D_1 or design D_2 from the finite population $\mathcal{P}$ of size $N > nH$. (i) The estimators $\bar{Y}_{\text{JPS},r}$ are unbiased for μ . (ii) the variance of the estimators $\bar{Y}_{\text{JPS},r}$ are

$$\sigma_1^2 = \frac{H}{H-1} \text{Var} \left(\frac{I_{1,1}}{d_{n,1}} \right) \sum_{h=1}^H (\mu_{[h],1} - \mu)^2 + E \left(\frac{I_{1,1}^2 J_{1,1}}{d_{n,1}^2} \right) \sum_{h=1}^H \sigma_{[h],1}^2$$

for design D_1 and

$$\sigma_2^2 = C_1(n, H) \left\{ \sum_{h=1}^H \sigma_{[h],2}^2 - \sum_{h=1}^H \sigma_{[h,h],2} \right\} + C_2(n, H, N) \frac{H^2 \sigma^2}{H-1} \quad (1)$$

for design D_2 , where

$$C_1(n, H) = \left\{ \frac{1}{H(H-1)} + E \left(\frac{I_{1,2}^2 J_{1,2}}{d_{n,2}^2} \right) - \frac{H}{H-1} E \left(\frac{I_{1,2}^2}{d_{n,2}^2} \right) \right\}$$

$$C_2(n, H, N) = \left\{ \text{Var} \left(\frac{I_{1,2}}{d_{n,2}} \right) - \frac{1}{N-1} \left\{ \frac{1}{H} - E \left(\frac{I_{1,2}^2}{d_{n,2}^2} \right) \right\} \right\}.$$

We now consider an unbiased estimator for the variance of $\bar{Y}_{\text{JPS},r}$. Let

$$U_{1,r} = \frac{1}{E \left(\frac{I_{1,r} I_{2,r}}{d_{n,r}^2} \right)} \sum_{h=1}^H \sum_{h' \neq h}^H \frac{I_{h,r} I_{h',r} J_{h,r} J_{h',r}}{d_{n,r}^2} \sum_{i=1}^n \sum_{j=1}^n (X_i - X_j)^2 I(R_{i,r} = h) I(R_{j,r} = h'),$$

$$U_{2,r} = \sum_{h=1}^H \frac{H I_{h,r}^* J_{h,r} J_{h,r}^*}{d_{n,r}^*} \sum_{i=1}^n \sum_{j \neq i}^n (X_i - X_j)^2 I(R_{i,r} = h) I(R_{j,r} = h),$$

and

$$\hat{\sigma}_{SRS}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2,$$

where

$$J_{h,r}^* = \begin{cases} \frac{1}{n_{h,r}-1} & n_{h,r} > 1 \\ 0 & n_{h,r} \leq 1, \end{cases}$$

and $I_{h,r}^* = I(n_{h,r} > 1)$, $d_{n,r}^* = \sum_{h=1}^H I_{h,r}^*$. From Lemma 1, one can easily establish that $E(I_{1,r} I_{2,r} / d_{n,r}^2) = (1/H - E(I_{1,r} / d_{n,r}^2)^2) / (H-1)$. Hence, $U_{1,r}$ is a statistic that depends only on the data. The following theorem (adapted to the notation of this paper) is stated from Ozturk (2016a).

Theorem 2. Let $(Y_i, R_{i,r})$, $i = 1, \dots, n$, be JPS samples of set size H from designs D_r , $r = 1, 2$. Assume that at least one of the judgment groups has at least two measured observations, $d_{n,r}^* > 0$. Unbiased estimators of σ_1^2 and σ_2^2 are given by, respectively,

$$\hat{\sigma}_1^2 = \frac{\text{Var}(I_{1,1}/d_{n,1})}{2(H-1)} U_{1,1} + \left\{ E\left(\frac{I_{1,1}^2 J_{1,1}}{d_{n,1}^2}\right) - \text{Var}\left(\frac{I_{1,1}}{d_{n,1}}\right) \right\} \frac{U_{2,1}}{2}, \quad (2)$$

$$\hat{\sigma}_2^2 = C_1(n, H) U_{2,2}/2 + C_2(n, H, N) \frac{H^2 \hat{\sigma}_{SRS}^2}{H-1}, \quad (3)$$

$$\bar{\sigma}_2^2 = C_1(n, H) U_{2,2}/2 + C_2(n, H, N) \frac{(N-1)(U_{1,2} + U_{2,2})}{2N(H-1)}, \quad (4)$$

Theorem 2 provides two unbiased estimators: $\hat{\sigma}_2^2$ and $\bar{\sigma}_2^2$ for σ_2^2 . Ozturk (2016a) showed that for some values of n , H , and N , the coefficient $C_2(n, H, N)$ could be negative. This may rarely lead to a negative value for $\hat{\sigma}_2^2$. When this happens, the following truncated variance estimator can be used

$$\tilde{\sigma}_2^2 = \begin{cases} \hat{\sigma}_2^2 & \text{if } \hat{\sigma}_2^2 > 0 \\ C_1(n, H) U_{2,2}/2 & \text{if } \hat{\sigma}_2^2 \leq 0. \end{cases} \quad (5)$$

Further development on this estimator can be found in Ozturk (2016a).

Statistical inference can also be drawn using a model-based approach under a super-population model using a multi-ranker model. Detailed development for this approach is provided in the Supplementary Material.

3. COMBINING RANKING INFORMATION

In this section, we look at JPS sampling from a different perspective. We assume that each measured unit is assigned K conditionally independent ranks given the comparison set. The JPS sampling still has a single measurement (Y_i) for each unit, but it has K different

ranks. The definition of the JPS sample in the previous section is now extended to cover the source of ranking information

$$D_{r,k} = \{Y_i, R_{i,k,r}\}, i = 1, \dots, n, r = 1, 2, k = 1, \dots, K, r = 1, 2,$$

where $R_{i,k,r}$ is the rank assigned to Y_i by ranking method k in design D_r . The subscript $r = 1, 2$ denotes that the sample is generated from designs D_1 or D_2 , respectively. As we indicated in Sect. 1, ranks can be constructed using information from different sources. Our objective here is to combine the information contained in these K sets of ranks to construct a better estimator than a JPS sample estimator based on a single ranking method. We consider two different approaches for combining the ranking information. The first approach uses the standard errors of the estimators. The second approach combines the ranking information using equal weights for each of the K sets of ranks.

Let $\bar{Y}_{k,r}$ and $\hat{\sigma}_{k,r}^2$ be the JPS estimator and its variance estimate, respectively, based on ranking method k and design D_r . For design D_1 , we combine the ranking information from different sources by constructing weighted estimators. We use either equal weight (E) or unequal weight of inverse standard error (S).

$$\bar{Y}_{E,1} = \frac{1}{K} \sum_{k=1}^K \bar{Y}_{k,1}, \quad \bar{Y}_{S,1} = \frac{\sum_{k=1}^K \frac{\bar{Y}_{k,1}}{\hat{\sigma}_{k,1}^2}}{A_1},$$

where

$$A_1 = \sum_{k=1}^K \frac{1}{\hat{\sigma}_{k,1}^2}.$$

Similar estimators can be constructed for design D_2 . There are two unbiased estimators for the variance of $\bar{Y}_2$ (σ_2^2). Our small scale simulation study showed that the variance estimator $\tilde{\sigma}_{k,2}^2$ performs slightly better than $\hat{\sigma}_{k,2}^2$. Therefore, we use $\tilde{\sigma}_{k,2}^2$ to construct the estimator that combines the ranking information from K ranking methods.

$$\bar{Y}_{E,2} = \frac{1}{K} \sum_{k=1}^K \bar{Y}_{k,2} \text{ and } \bar{Y}_{S,2} = \frac{\sum_{k=1}^K \frac{\bar{Y}_{k,2}}{\tilde{\sigma}_{k,2}^2}}{A_2}, \text{ where } A_M = \sum_{k=1}^K \frac{1}{\tilde{\sigma}_{k,M}^2}.$$

The combined estimators can be classified into two groups. The first group gives equal weight to each individual ranking estimator. The second group assigns greater weights to the ranking methods producing smaller variances. In the Supplementary Material, we demonstrate another approach to combining the ranking information at each observation level using the agreement scores.

4. VARIANCE ESTIMATOR OF THE COMBINED ESTIMATOR

In order to assess the uncertainty of the combined estimators, we need to estimate their variances. We first consider the jackknife variance estimate for the combined estimators. This estimate is appealing in settings where it may not be clear how to calculate a good estimator or the standard deviation of the estimator. Such situations occur when the delta method may not work properly due to the absence of a clear theoretical or functional relationship between the data and the estimator. In our study, the JPS sampling generates one set of response measurements and K different sets of ranks for each measured unit in its comparison set, and the ranking methods could vary substantially. For example, the same comparison set in a field research project may be ranked through visual inspection, an auxiliary field variable and computer assisted automated ranking methods.

Let $\bar{Y}_{E,r}^{(-i)}$ and $\bar{Y}_{S,r}^{(-i)}$, for $r = 1, 2$, be the combined estimators after removing the i -th observation along with all its ranks from the sample. The jackknife variance estimate for each of the combined estimators is then given by

$$\hat{\sigma}_{E,r}^2 = \sum_{i=1}^n (\hat{Y}_{E,r}^{(-i)} - \bar{J}_{E,r})^2 ((n-1)/n)^2, \quad r = 1, 2,$$

$$\hat{\sigma}_{S,r}^2 = \sum_{i=1}^n (Y_{S,r}^{*(-i)} - \bar{J}_{S,r})^2 ((n-1)/n)^2, \quad r = 1, 2,$$

where $\bar{J}_{E,r} = \sum_{i=1}^n \bar{Y}_{E,r}^{(-i)} / n$ and $\bar{J}_{S,r} = \sum_{i=1}^n \bar{Y}_{S,r}^{(-i)} / n$. In the jackknife variance estimators, the coefficient $(n-1)/n$ is replaced with $((n-1)/n)^2$.

The variance estimates allow us to construct an approximate $100(1 - \alpha)\%$ confidence interval for the population mean and total,

$$\bar{Y}_{E,r} \pm t_{n-1, 1-\alpha/2} \hat{\sigma}_{E,r}, \quad \bar{Y}_{S,r} \pm t_{n-1, 1-\alpha/2} \hat{\sigma}_{S,r},$$

$$\bar{Y}_{JPS,r} \pm t_{n-1, 1-\alpha/2} \hat{\sigma}_{JPS,r}, \quad \bar{Y}_{SRS,r} \pm t_{n-1, 1-\alpha/2} \hat{\sigma}_{SRS,r}$$

for $r = 1, 2$, where $t_{n-1, a}$ is the a -th upper quantile of the t-distribution with $n-1$ degrees of freedom, and $\bar{Y}_{SRS,r}$ and $\hat{\sigma}_{SRS,r}$ are the sample mean and sample variance of a simple random sample under design D_r , $r = 1, 2$. An R-package *RankedSetSampling* is developed to compute the estimators and construct confidence intervals for Design D₁ and D₂, Ozturk et al. (2021). It is available to download at <https://biometryhub.github.io/RankedSetSampling>.

5. EMPIRICAL COMPARISONS OF COMBINED ESTIMATORS

In this section, we performed a simulation study to empirically investigate the properties of the estimators. We considered four populations of various distribution shapes likely to occur in practice. The populations include uniform (also $\text{beta}(p = 1, q = 1)$), $\text{beta}(4, 4)$, $\text{beta}(4, 13)$ and the standard exponential distribution ($\exp(1)$). These distributions correspond to populations with short tail (uniform, $\text{beta}(1, 1)$), symmetric ($\text{beta}(4, 4)$), slightly positively

skewed (beta(4,13)) and positively skewed (exponential, exp(1)). For convenience, uniform, beta(4,4), beta(4,13), and exp(1) distributions are denoted as Group I, II, III, and IV populations, respectively. The values of the population units in simulations are generated by the quantiles of these four distributions

$$y_i = F_G^{-1}(i/(N+1)), i = 1, \dots, N, G = \text{I, II, III, IV},$$

where $N = 400$ is the size of the finite population and F_G is the cumulative distribution function of the Group G population, $G = \text{I, II, III, IV}$. These four populations are also used in McIntyre (1952, 2005) to illustrate the efficiency of ranked set sample designs in agricultural experiments.

The JPS samples are generated using designs D_1 and D_2 with set sizes $H = 3, 6$ and sample sizes $n = 30$. The number of rankers and simulation size are selected to be $K = 3, 12$ and 2000, respectively. The quality of ranking information is modeled through the Dell and Clutter (1972) model. This model creates a ranking variable V in each comparison set. Let $\mathbf{Y}_i^\top = (Y_i, Y_1, \dots, Y_{H-1})$ be the comparison set for Y_i generated from a population with variance σ^2 . The Dell and Clutter model generates another random vector, $\boldsymbol{\epsilon}_i^\top = (\epsilon_i, \epsilon_1, \dots, \epsilon_{H-1})$, from a normal distribution with mean zero and variance σ_ϵ^2 . We then add these two vectors to construct the ranking vector $\mathbf{V}_i^\top = (V_i, V_1, \dots, V_{H-1})$. Units in the comparison set are ranked based on the values of V , and the rank of V_i is taken as the judgment rank of Y_i . The correlation coefficient between Y_i and V_i is given by $\rho = 1/\sqrt{1 + \sigma_\epsilon^2/\sigma^2}$. The magnitude of σ_ϵ^2 controls the quality of ranking information. If $\sigma_\epsilon^2 = 0$, the rank of Y_i is the same as the rank of V_i , hence, no ranking error occurs. Large values of σ_ϵ^2 correspond to random ranking. In our simulation study, we considered six different ranking models.

M_1 : All K ranking methods use $\rho = 0.1$.

M_2 : All K ranking methods use $\rho = 0.5$.

M_3 : All K ranking methods use $\rho = 0.8$.

M_4 : Two-thirds of the ranking methods use $\rho = 0.8$ and the remaining one-third use $\rho = 0.3$.

M_5 : One-third of the ranking methods use $\rho = 0.8$ and the remaining two-thirds use $\rho = 0.3$.

M_6 : The bottom, middle and upper one-third of ranking methods use $\rho = 0.8, 0.5, 0.3$, respectively.

The relative efficiencies are defined as

$$\text{RE}_{E,r} = \frac{\text{MSE}(\bar{Y}_{\text{SRS},r})}{\text{MSE}(\bar{Y}_{E,r})}, \quad \text{RE}_{S,r} = \frac{\text{MSE}(\bar{Y}_{\text{SRS},r})}{\text{MSE}(\bar{Y}_{S,r})}, \quad \text{RE}_{J,r} = \frac{\text{MSE}(\bar{Y}_{\text{SRS},r})}{\text{MSE}(\bar{Y}_{\text{JPS},2})},$$

for $r = 1, 2$. We note that the estimators in each design D_r , $r = 1, 2$ are compared with the simple random sample mean estimator. Values of $\text{RE}_{\dots}$ greater than one indicate that the rank-based estimators have higher efficiencies than the SRS estimator of the same size.

Table 1. Efficiency of multi-ranker combined estimators relative to the estimator $\bar{Y}_{\text{SRS},r}$, $\text{RE}_{E,r} = \text{MSE}(\bar{Y}_{\text{SRS},r})/\text{MSE}(\bar{Y}_{E,r})$, $r = 1, 2$

Group	H	M_i	$K = 3$						$K = 12$					
			Design D_1			Design D_2			Design D_1			Design D_2		
			$\text{RE}_{E,1}$	$\text{RE}_{S,1}$	$\text{RE}_{J,1}$	$\text{RE}_{E,2}$	$\text{RE}_{S,2}$	$\text{RE}_{J,2}$	$\text{RE}_{E,1}$	$\text{RE}_{S,1}$	$\text{RE}_{J,1}$	$\text{RE}_{E,2}$	$\text{RE}_{S,2}$	$\text{RE}_{J,2}$
I	3	M_1	0.984	0.976	0.941	0.982	0.974	0.928	1.002	0.994	0.930	0.997	0.988	0.912
	3	M_2	1.182	1.170	1.019	1.218	1.203	1.055	1.260	1.259	1.063	1.291	1.292	1.036
	3	M_3	1.710	1.696	1.396	1.794	1.771	1.404	1.835	1.858	1.374	2.016	2.043	1.386
	3	M_4	1.480	1.506	1.358	1.493	1.524	1.403	1.578	1.658	1.426	1.666	1.754	1.398
	3	M_5	1.221	1.248	1.360	1.221	1.252	1.342	1.321	1.392	1.451	1.329	1.403	1.458
	3	M_6	1.315	1.337	1.391	1.367	1.396	1.453	1.381	1.442	1.444	1.392	1.459	1.426
	6	M_1	0.935	0.919	0.805	0.927	0.909	0.794	0.995	0.977	0.797	0.995	0.981	0.791
	6	M_2	1.259	1.242	0.966	1.264	1.241	0.975	1.387	1.390	0.953	1.421	1.417	0.962
	6	M_3	2.287	2.254	1.506	2.428	2.368	1.537	2.715	2.727	1.496	3.116	3.138	1.542
	6	M_4	1.705	1.797	1.439	1.776	1.901	1.526	1.996	2.196	1.496	2.115	2.361	1.520
	6	M_5	1.320	1.415	1.519	1.309	1.397	1.449	1.461	1.627	1.522	1.525	1.730	1.510
	6	M_6	1.418	1.489	1.484	1.465	1.542	1.547	1.597	1.756	1.448	1.672	1.849	1.548
II	3	M_1	0.964	0.957	0.921	0.980	0.975	0.939	1.001	0.996	0.920	0.997	0.992	0.898
	3	M_2	1.166	1.162	1.044	1.231	1.224	1.070	1.266	1.270	1.070	1.306	1.313	1.064
	3	M_3	1.629	1.619	1.331	1.746	1.737	1.376	1.893	1.913	1.452	1.922	1.941	1.411
	3	M_4	1.451	1.478	1.377	1.552	1.595	1.446	1.563	1.631	1.350	1.614	1.692	1.355
	3	M_5	1.230	1.258	1.326	1.269	1.307	1.401	1.308	1.379	1.410	1.326	1.397	1.367
	3	M_6	1.273	1.303	1.309	1.328	1.361	1.397	1.362	1.414	1.358	1.378	1.442	1.401
	6	M_1	0.951	0.944	0.840	0.919	0.904	0.767	0.981	0.975	0.747	0.987	0.979	0.774
	6	M_2	1.244	1.227	0.955	1.235	1.226	0.962	1.373	1.381	0.971	1.429	1.432	0.971
	6	M_3	2.119	2.099	1.414	2.322	2.284	1.463	2.606	2.656	1.523	2.945	2.972	1.529
	6	M_4	1.695	1.789	1.461	1.739	1.840	1.449	1.907	2.079	1.481	2.013	2.229	1.498
	6	M_5	1.295	1.372	1.415	1.367	1.466	1.548	1.446	1.602	1.433	1.511	1.710	1.541
	6	M_6	1.357	1.425	1.360	1.462	1.558	1.568	1.570	1.700	1.462	1.634	1.803	1.524

Population values are generated from $y_i = F_G^{-1}(i/(N+1))$, where F_G is the CDF of group G , $G = \text{I}(\text{uniform}), \text{II}(\text{beta}(4,4)), \text{III}(\text{beta}(4,13)), \text{IV}(\text{exp}(1))$. The population size is $N = 400$, the sample size $n = 30$. Simulation size is 2000

Table 1 provides the relative efficiencies for Group I (uniform) and II (beta(4,4)) populations when $K = 3$ and $K = 12$. It is clear that the JPS estimators perform better than the SRS estimators for models M_2, M_3, M_4, M_5 and M_6 . In ranking model M_1 , there is a weak correlation, $\rho = 0.1$, between the response and ranking variables. Hence, for this model, the JPS estimators have RE values less than one and appear to be slightly less efficient than the SRS estimator. These low efficiency values are related to the quality of ranking information. Since the judgment group sample sizes are random variables in a JPS sample, they bring an additional variation to the variance of the estimator. When there is no ranking information to separate the data into homogeneous ranking groups, the JPS sample estimators may be slightly less efficient. On the other hand, the multi-ranker combined JPS estimators with $K = 12$ and $H = 6$ are almost as efficient as the SRS estimator even for ranking model M_1 . This shows that the multi-ranker JPS estimators perform as well as the SRS estimator even if the ranking quality is poor. If there is a reasonable information gain to improve the ranking quality, the multi-ranker JPS estimators always outperform the SRS and a JPS estimator with $K = 1$.

The impact of ranking quality can be clearly seen by considering the efficiency achieved for models M_1 , M_2 and M_3 , where all the ranking methods have the same correlation $\rho = 0.1, 0.5$, and 0.8 , respectively. The efficiency values are ordered for models M_1 , M_2 and M_3 in all columns in Table 1. The efficiencies of all the estimators are increasing functions of ρ , for a given design D_r and the number of ranking methods K . The efficiencies for model M_1 ($\rho = 0.1$) are smaller than 1 when $K = 3$, but nearly equal one when $K = 12$. When ρ increases to 0.5 (M_2) and 0.8 (M_3) the efficiencies also increase for both designs (D_1, D_2) with the number of ranking methods $K = 3$ and $K = 12$.

If the ranking correlation ρ is the same for all ranking methods (ranking models M_1 , M_2 , and M_3), the estimators $\bar{Y}_{S,r}$ and $\bar{Y}_{E,r}$ have similar efficiency results. On the other hand, if ranking methods have different ρ values (ranking models M_4 , M_5 and M_6), the estimator weighted with the inverse standard error outperforms the equal weight estimator. For example, the values of $RE_{S,r}$ and $RE_{E,r}$ are essentially equal for models M_1 , M_2 and M_3 , while $RE_{S,r}$ are larger than $RE_{E,r}$ for models M_4 , M_5 and M_6 .

It is also clear that the efficiencies are an increasing function of the set size when there is a reasonable ranking information to rank the units in the comparison sets. For example, the efficiencies of all the estimators for set size $H = 6$ are higher than the efficiencies of the estimators with $H = 3$ for ranking models M_i , $i = 1, 2, \dots, 6$. For model M_1 , the efficiencies of the estimators of set size $H = 6$ (row 7, Table 1) are lower than the efficiencies of the estimators of set size $H = 3$ (row 1). The reason for this is that when $H = 6$ and $n = 30$, the judgment ranking group sample sizes are more uneven than when $H = 3$ and $n = 30$. This leads to some ranking groups being empty, and inflates the variance of the estimator. It is less likely to have empty ranking groups when $H = 3$.

Both Groups I and II populations are symmetric, but the Group I distribution has a short tail (or heavy tail) in comparison to that of Group II. The Group III population is slightly right-skewed, while the Group IV distribution has a strong skewness to the right. The efficiencies of the estimators for Groups III and IV are given in Table 2. Even though the shapes of these populations are different, the efficiency patterns are similar. The main difference is that the improvement in efficiencies decreases with the strength of skewness. The uniform distribution yields the maximum improvement in efficiency. The least improvement happens for the exponential distribution. This result is consistent with the efficiency results of ranked set samples in McIntyre (1952), McIntyre (2005). McIntyre reported that the efficiencies are the highest for the uniform distribution and decrease with skewness. A similar result is also reported in Takahasi and Wakimoto (1968). Under perfect ranking and infinite population settings, they proved that the upper bound of relative efficiency of a ranked set sample mean with respect to the simple random sample mean is $(H + 1)/2$. This upper bound is achieved for the uniform distribution.

We also computed the simulation standard errors of the multi-ranker estimators of population mean. The maximum value of the simulation standard error for simulation size 2000 was 0.002. Due to space considerations, no tables of the estimated simulation standard errors are reported here.

Let $C_{E,r}$, $C_{S,r}$ and $C_{J,r}$ be the coverage probabilities of the confidence intervals of population mean based on the estimators $\bar{Y}_{E,r}$, $\bar{Y}_{S,r}$ and $\bar{Y}_{J,r}$, respectively. Tables 3 and 4 present these coverage probabilities for the sample size of $n = 30$ and the population

Table 2. Efficiency of multi-ranker combined estimators relative to the estimator $\bar{Y}_{\text{SRS},r}$, $\text{RE}_{E,r} = \text{MSE}(\bar{Y}_{\text{SRS},r})/\text{MSE}(\bar{Y}_{E,r})$, $r = 1, 2$

Dist	H	M_i	$K = 3$						$K = 12$					
			Design D_1			Design D_2			Design D_1			Design D_2		
			RE_{E,D_1}	$\text{RE}_{S,1}$	$\text{RE}_{JPS,1}$	$\text{RE}_{E,2}$	$\text{RE}_{S,2}$	$\text{RE}_{JPS,2}$	$\text{RE}_{E,1}$	$\text{RE}_{S,1}$	$\text{RE}_{JPS,1}$	$\text{RE}_{E,2}$	$\text{RE}_{S,2}$	$\text{RE}_{JPS,2}$
III	3	M_1	0.973	0.970	0.913	0.976	0.971	0.922	0.999	0.999	0.919	1.007	1.006	0.911
	3	M_2	1.197	1.199	1.045	1.201	1.201	1.094	1.266	1.276	1.066	1.299	1.310	1.046
	3	M_3	1.636	1.624	1.326	1.690	1.683	1.320	1.825	1.840	1.343	1.901	1.917	1.390
	3	M_4	1.390	1.425	1.305	1.449	1.468	1.338	1.482	1.540	1.312	1.581	1.651	1.347
	3	M_5	1.228	1.262	1.344	1.242	1.278	1.329	1.282	1.341	1.328	1.301	1.368	1.291
	3	M_6	1.263	1.279	1.258	1.307	1.328	1.357	1.369	1.418	1.420	1.391	1.443	1.325
	6	M_1	0.957	0.945	0.818	0.948	0.937	0.816	0.989	0.982	0.801	0.995	0.991	0.818
	6	M_2	1.210	1.198	0.961	1.259	1.242	0.991	1.376	1.388	0.971	1.405	1.415	0.947
	6	M_3	2.108	2.078	1.446	2.339	2.289	1.491	2.488	2.488	1.356	2.887	2.924	1.609
	6	M_4	1.669	1.757	1.477	1.671	1.745	1.406	1.850	2.031	1.436	2.015	2.244	1.451
	6	M_5	1.317	1.401	1.364	1.331	1.424	1.480	1.416	1.561	1.374	1.470	1.635	1.493
	6	M_6	1.390	1.459	1.419	1.370	1.450	1.486	1.529	1.663	1.443	1.595	1.741	1.408
IV	3	M_1	0.995	0.996	0.936	0.994	0.997	0.931	1.003	1.008	0.932	1.004	1.010	0.917
	3	M_2	1.151	1.158	1.022	1.177	1.191	1.030	1.218	1.233	0.993	1.231	1.249	1.049
	3	M_3	1.470	1.464	1.222	1.492	1.498	1.222	1.560	1.568	1.170	1.629	1.637	1.204
	3	M_4	1.377	1.406	1.308	1.351	1.372	1.262	1.390	1.427	1.203	1.422	1.464	1.199
	3	M_5	1.164	1.189	1.226	1.184	1.210	1.233	1.241	1.290	1.211	1.249	1.294	1.264
	3	M_6	1.207	1.230	1.242	1.225	1.247	1.241	1.292	1.327	1.214	1.311	1.354	1.257
	6	M_1	0.941	0.933	0.829	0.927	0.919	0.784	0.993	0.995	0.767	0.991	0.990	0.793
	6	M_2	1.255	1.249	0.982	1.251	1.236	0.967	1.320	1.332	0.925	1.371	1.394	0.951
	6	M_3	1.836	1.800	1.349	1.978	1.939	1.305	2.137	2.126	1.272	2.221	2.214	1.265
	6	M_4	1.534	1.580	1.349	1.592	1.645	1.282	1.787	1.900	1.351	1.834	1.976	1.296
	6	M_5	1.252	1.313	1.330	1.269	1.331	1.340	1.350	1.447	1.272	1.392	1.525	1.292
	6	M_6	1.328	1.382	1.327	1.371	1.413	1.338	1.495	1.591	1.260	1.530	1.657	1.316

The population values are generated from $y_i = F_G^{-1}(i/(N+1))$, where F_G is the CDF of group G , $G = I(\text{uniform}), II(\text{beta}(4,4)), III(\text{beta}(4,13)), IV(\text{exp}(1))$. The population size is $N = 400$, the sample sizes $n = 30$. The simulation size is 2000

distributions of Groups I, II, III and IV. It is clear that all the coverage probabilities are reasonably close to the nominal coverage probability of 0.95, for both symmetric and skewed distributions. We also performed another simulation study with smaller sample size $n = 15$. In that case, the coverage probabilities are reasonably close to 0.95 for Groups I, II and III. The coverage probabilities for Group IV (exponential distribution) were about 0.90, which is smaller than the nominal value 0.95. The results are presented in Supplementary Material. This additional simulation study shows that to achieve a reasonable coverage probability for skewed distributions, the multi-ranker JPS samples require larger sample sizes.

6. FAB BEAN CROP ESTABLISHMENT EXAMPLE

In this section, we performed another simulation study using the faba bean seedling emergence data introduced in Sect. 1. The population means and variances of the actual seedling counts (Y) and automated seedling counts (X) for the entire field of $N = 2640$ transects are $\mu_Y = 1.051$ seedling/50 cm, $\sigma_Y^2 = 0.926$ (seedling/50 cm) ^{2} , and $\mu_X = 0.769$

Table 3. Coverage probabilities of jackknife confidence intervals of population mean

Dist	H	M_i	$K = 3$						$K = 12$					
			Design D_1			Design D_2			Design D_1			Design D_2		
			$\bar{C}_{E,1}$	$\bar{C}_{S,1}$	$\bar{C}_{J,1}$	$\bar{C}_{E,2}$	$\bar{C}_{S,2}$	$\bar{C}_{J,2}$	$\bar{C}_{E,1}$	$\bar{C}_{S,1}$	$\bar{C}_{J,1}$	$\bar{C}_{E,2}$	$\bar{C}_{S,2}$	$\bar{C}_{J,2}$
I	3	M_1	0.949	0.950	0.949	0.958	0.958	0.955	0.941	0.942	0.944	0.959	0.957	0.953
	3	M_2	0.947	0.946	0.944	0.962	0.961	0.962	0.958	0.960	0.955	0.965	0.966	0.962
	3	M_3	0.946	0.946	0.946	0.962	0.962	0.957	0.952	0.952	0.949	0.964	0.964	0.948
	3	M_4	0.960	0.958	0.951	0.956	0.962	0.958	0.950	0.951	0.943	0.970	0.970	0.962
	3	M_5	0.951	0.951	0.948	0.956	0.958	0.955	0.963	0.964	0.955	0.957	0.962	0.957
	3	M_6	0.946	0.948	0.950	0.969	0.970	0.968	0.954	0.956	0.952	0.962	0.963	0.966
	6	M_1	0.956	0.955	0.948	0.959	0.958	0.954	0.955	0.956	0.945	0.962	0.963	0.952
	6	M_2	0.954	0.953	0.946	0.959	0.960	0.948	0.954	0.958	0.943	0.967	0.967	0.953
	6	M_3	0.960	0.961	0.950	0.970	0.971	0.957	0.960	0.961	0.938	0.979	0.978	0.962
	6	M_4	0.959	0.958	0.944	0.968	0.972	0.953	0.960	0.962	0.951	0.969	0.972	0.955
	6	M_5	0.949	0.959	0.939	0.967	0.970	0.958	0.942	0.949	0.948	0.970	0.976	0.951
	6	M_6	0.961	0.963	0.938	0.968	0.963	0.956	0.960	0.959	0.934	0.964	0.969	0.953
II	3	M_1	0.948	0.947	0.950	0.959	0.960	0.963	0.952	0.950	0.949	0.953	0.952	0.950
	3	M_2	0.951	0.949	0.947	0.966	0.966	0.962	0.954	0.952	0.950	0.967	0.967	0.960
	3	M_3	0.950	0.948	0.945	0.970	0.970	0.963	0.953	0.954	0.957	0.966	0.969	0.961
	3	M_4	0.952	0.954	0.951	0.966	0.969	0.967	0.952	0.954	0.941	0.964	0.963	0.961
	3	M_5	0.952	0.954	0.944	0.965	0.966	0.963	0.939	0.943	0.950	0.963	0.966	0.960
	3	M_6	0.947	0.947	0.939	0.968	0.968	0.965	0.947	0.946	0.947	0.963	0.965	0.959
	6	M_1	0.957	0.957	0.952	0.963	0.962	0.951	0.952	0.954	0.939	0.962	0.962	0.953
	6	M_2	0.960	0.959	0.948	0.960	0.962	0.954	0.946	0.946	0.944	0.962	0.966	0.963
	6	M_3	0.966	0.966	0.945	0.976	0.974	0.958	0.965	0.965	0.958	0.975	0.977	0.958
	6	M_4	0.962	0.962	0.941	0.968	0.971	0.953	0.962	0.958	0.953	0.967	0.971	0.953
	6	M_5	0.963	0.968	0.954	0.959	0.962	0.951	0.958	0.962	0.949	0.973	0.973	0.961
	6	M_6	0.960	0.961	0.940	0.962	0.966	0.954	0.946	0.949	0.941	0.964	0.969	0.962

Population values are generated from $y_i = F_G^{-1}(i/(N+1))$, where F_G is the CDF of group G , $G = \text{I(uniform), II(beta(4,4)), III(beta(4,13)), IV(exp(1))}$. The population size is $N = 400$, the sample size $n = 30$, the simulation size 2000

seedling/50 cm, $\sigma_X^2 = 0.649$ (seedling/50 cm)², respectively. The correlation coefficient between X and Y is relatively high, $\rho_{Y,X} = \text{cor}(X, Y) = 0.766$.

In this simulation, the JPS samples were generated with set sizes $H = 3, 4, 6$ and sample sizes $n = 15, 30$. For each JPS sample in the simulation, we constructed $K = 3, 12$ different comparison sets and determined the rank of Y using the auxiliary variable X . Hence, each Y value in a sample had either $K = 3$ or $K = 12$ judgment ranks $R_{i,k}$, $\{Y_i, R_{i,k}, k = 1, \dots, K; i = 1, \dots, n\}$, where $R_{i,k}$ is the rank of Y_i in the comparison set k . The simulation size was taken to be 2000. We note that the actual values of seedling emergence Y and their predicted values are count data per a given area. Since the X -variable is a discrete random variable, ties among X -values in comparison sets may occur. When ties happen, we break them at random. Even though the correlation coefficient between Y and X is relatively high, ties among the X -values in the comparison sets may reduce the quality of ranking information.

In this population, values of the X variable are available for all $N = 2640$ population units. This provides an alternative estimator to estimate the population mean or total using a ratio estimator. For the ratio estimator, we select an SRS of size n using sampling *with* (D_1)

and *without* (D_2) replacement and measure the pairs $(Y_{i,r}, X_{i,r})$, $i = 1, \dots, n$ for design D_r , $r = 1, 2$. The ratio estimator of the population mean is then given by

$$\bar{Y}_{R,r} = \frac{\bar{Y}_r}{\bar{X}_r} \bar{x}, r = 1, 2,$$

where

$$\bar{Y}_r = \frac{1}{n} \sum_{i=1}^n Y_{i,r}; \quad \bar{X}_r = \frac{1}{n} \sum_{i=1}^n X_{i,r}; \quad \bar{x} = \frac{1}{N} \sum_{i=1}^N x_i.$$

The efficiency of the ratio estimator with respect to the SRS estimator is defined by

$$RE_{R,r} = \frac{MSE(\bar{Y}_{SRS,r})}{MSE(\bar{Y}_{R,r})}, r = 1, 2.$$

Table 5 presents the efficiencies of the estimators with respect to the estimator $\bar{Y}_{SRS,r}$. It is clear that all estimators, including the ratio estimators, provide an improvement over the SRS within each sampling design D_r . Among these estimators, the least efficient estimators ($RE_{J,r}$) are the JPS estimators with a single ranking method. The JPS estimators are in general less efficient than the ratio estimators; see the efficiency values ($RE_{J,1}$, $RE_{R,1}$) and ($RE_{J,2}$, $RE_{R,2}$).

The efficiency improvements of multi-ranker JPS estimators, even when $K = 3$, are substantial. Both equal and unequal weight JPS estimators perform better than JPS estimators with $K = 1$ and the ratio estimators in both designs D_1 and D_2 . For example, the standard error-weighted estimators are 1.252 (1.906/1.523) and 1.206 (1.907/1.581) times more efficient than the ratio estimators when $n = 30$, $K = 12$ and $H = 6$ for designs D_1 and D_2 , respectively. For the JPS estimator with $K = 1$, the same efficiency values are 1.536 (1.906/1.241) and 1.537 (1.907/1.241).

Table 5 demonstrates that the efficiency values of equally and standard error-weighted estimators are very close to each other. In this simulation, all K ranking methods use the same correlation coefficient $\rho_{X,Y} = 0.766$. Hence, the standard error weights for all ranking groups are *on average* equal. This leads to equal weights and hence similar efficiency values to those of the equally weighted estimators.

Table 5 shows very little difference between the efficiency values of the same estimators under design D_1 and design D_2 . For example, the efficiency values in columns 4 and 8, 5 and 9, and 6 and 10 are very close. This can be anticipated since the finite population correction factor 0.989 (1-30/2640) is very close to 1, and the impact of sampling without replacement is not visible.

We have also computed the empirical coverage probabilities of the confidence intervals for the population mean. All confidence intervals provided coverage probabilities reasonably close to the nominal coverage probability of 0.95. These results are not reported here.

Table 4. Coverage probabilities of jackknife confidence intervals of population mean

Dist	H	M_i	$K = 3$						$K = 12$					
			Design D_1			Design D_2			Design D_1			Design D_2		
			$\bar{C}_{E,1}$	$\bar{C}_{S,1}$	$\bar{C}_{J,1}$	$\bar{C}_{E,2}$	$\bar{C}_{S,2}$	$\bar{C}_{J,2}$	$\bar{C}_{E,1}$	$\bar{C}_{S,1}$	$\bar{C}_{J,1}$	$\bar{C}_{E,2}$	$\bar{C}_{S,2}$	$\bar{C}_{J,2}$
III	3	M_1	0.942	0.940	0.942	0.954	0.954	0.957	0.946	0.946	0.942	0.956	0.957	0.957
	3	M_2	0.950	0.950	0.946	0.955	0.955	0.952	0.948	0.950	0.946	0.956	0.956	0.952
	3	M_3	0.949	0.948	0.949	0.958	0.957	0.951	0.943	0.944	0.950	0.966	0.966	0.962
	3	M_4	0.955	0.957	0.956	0.961	0.961	0.959	0.941	0.943	0.943	0.960	0.962	0.950
	3	M_5	0.950	0.949	0.948	0.959	0.961	0.952	0.949	0.948	0.948	0.963	0.962	0.961
	3	M_6	0.955	0.952	0.945	0.962	0.965	0.962	0.946	0.945	0.948	0.960	0.961	0.955
	6	M_1	0.952	0.953	0.945	0.963	0.961	0.960	0.945	0.946	0.944	0.952	0.952	0.952
	6	M_2	0.946	0.948	0.940	0.964	0.963	0.956	0.945	0.946	0.935	0.956	0.957	0.947
	6	M_3	0.953	0.952	0.944	0.973	0.970	0.956	0.956	0.953	0.942	0.974	0.973	0.964
	6	M_4	0.957	0.954	0.942	0.969	0.970	0.957	0.943	0.947	0.938	0.973	0.975	0.954
	6	M_5	0.957	0.960	0.948	0.959	0.961	0.954	0.954	0.956	0.941	0.955	0.958	0.951
	6	M_6	0.951	0.953	0.942	0.961	0.960	0.961	0.955	0.959	0.946	0.964	0.967	0.950
IV	3	M_1	0.930	0.930	0.925	0.941	0.939	0.940	0.923	0.920	0.926	0.930	0.931	0.931
	3	M_2	0.929	0.928	0.928	0.944	0.943	0.941	0.941	0.939	0.933	0.931	0.931	0.931
	3	M_3	0.940	0.939	0.939	0.947	0.952	0.934	0.926	0.926	0.918	0.945	0.947	0.933
	3	M_4	0.930	0.930	0.927	0.944	0.940	0.945	0.928	0.929	0.925	0.954	0.952	0.946
	3	M_5	0.926	0.927	0.936	0.940	0.939	0.941	0.940	0.939	0.933	0.940	0.942	0.940
	3	M_6	0.918	0.916	0.920	0.936	0.938	0.935	0.929	0.928	0.923	0.943	0.942	0.942
	6	M_1	0.925	0.924	0.921	0.944	0.943	0.936	0.931	0.930	0.933	0.931	0.930	0.925
	6	M_2	0.929	0.930	0.911	0.949	0.950	0.944	0.932	0.932	0.927	0.955	0.951	0.937
	6	M_3	0.940	0.942	0.925	0.956	0.958	0.938	0.950	0.948	0.927	0.959	0.957	0.936
	6	M_4	0.940	0.936	0.930	0.959	0.957	0.941	0.948	0.946	0.933	0.951	0.953	0.942
	6	M_5	0.944	0.944	0.935	0.950	0.953	0.938	0.934	0.938	0.924	0.946	0.951	0.936
	6	M_6	0.935	0.938	0.930	0.947	0.950	0.941	0.945	0.945	0.929	0.956	0.961	0.951

Population values are generated from $y_i = F_G^{-1}(i/(N+1))$, where F_G is the CDF of group G , $G = \text{I(uniform), II(beta(4,4)), III(beta(4,13)), IV(exp(1))}$. The population size is $N = 400$, sample size $n = 30$, simulation size 2000

Table 5. Efficiency of the multi-ranker combined estimators relative to the SRS estimator $\bar{Y}_{\text{SRS},r}$, $\text{RE}_{E,r} = \text{MSE}(\bar{Y}_{\text{SRS},r})/\text{MSE}(\bar{Y}_{E,2})$ for the faba bean seedling emergence data, the population size is $N = 2640$

n	K	H	Design D_1				Design D_2			
			$\text{RE}_{E,1}$	$\text{RE}_{S,1}$	$\text{RE}_{J,1}$	$\text{RE}_{R,1}$	$\text{RE}_{E,2}$	$\text{RE}_{S,2}$	$\text{RE}_{J,2}$	$\text{RE}_{R,2}$
15	3	3	1.436	1.408	1.150	1.126	1.386	1.368	1.097	1.138
15	3	4	1.410	1.386	1.047	1.138	1.425	1.396	1.083	1.471
15	3	6	1.419	1.405	1.088	1.056	1.418	1.395	1.107	1.253
15	12	3	1.558	1.578	1.147	1.237	1.543	1.563	1.115	1.267
15	12	4	1.584	1.584	1.101	1.283	1.612	1.626	1.023	1.224
15	12	6	1.614	1.619	1.103	1.076	1.608	1.610	1.079	1.182
30	3	3	1.566	1.557	1.348	1.626	1.550	1.545	1.315	1.470
30	3	4	1.599	1.591	1.221	1.672	1.590	1.574	1.273	1.397
30	3	6	1.662	1.637	1.238	1.624	1.632	1.614	1.233	1.532
30	12	3	1.571	1.588	1.255	1.574	1.600	1.613	1.234	1.651
30	12	4	1.730	1.740	1.317	1.434	1.740	1.758	1.224	1.465
30	12	6	1.893	1.906	1.241	1.523	1.894	1.907	1.241	1.581

7. CONCLUDING REMARKS

Many survey and field sampling studies collect concurrently with the main variable or make it available in some other way; additional auxiliary variables correlated with the variable of interest. This situation is typical of modern field phenotyping in plant research. This additional information can induce a structure among the sample units, creating multiple sets of ranks for each measured observation. These ranks can be used to form judgment post-stratified samples with multiple ranks for a given data set. In this paper, we introduce novel weighted estimators, which combine the ranking information from different sources in a JPS sample. We constructed two different weight functions. The first weight function uses the standard error of each single-ranker JPS estimator. The second weight function (presented in Supplementary Material) utilizes the agreement scores of all ranking methods. The variances of the estimators are computed using the jackknife variance estimate. Approximate $(1 - \alpha)100\%$ confidence intervals are constructed.

The empirical studies show that the proposed estimators always perform better than the SRS and the JPS estimator of a single ranking method. The efficiency gain is substantial if the set size of the comparison set is large, $H \geq 3$. Combining ranking information from multiple sources can also be applied to more complex designs, such as stratified JPS and two-stage JPS designs. The implementation of such novel estimators requires new computing tools to handle the computational complexity. An R-package has been developed to compute the estimators.

ACKNOWLEDGEMENTS

Funding was provided by Australian Grain Research and Development Corporation, South Australian Grain Industry Trust Fund (AU) and Australian Grain Research and Development Corporation. Funding for open-access was kindly provided by the Australian Grains Research and Development Corporation (project UA00164) and Academic Consortium for the 21st Century (Special Project Fund, 2018).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

[Received July 2020. Revised November 2020. Accepted January 2021. Published Online February 2021.]

REFERENCES

- Araus JL, Cairns JE (2014) Field high-throughput phenotyping: the new crop breeding frontier. *Trends Plant Sci* 19:52–61
- Chen M, Ahn S, Wang X, Lim J (2014) Generalized isotonized mean estimators for judgment post-stratification with multiple rankers. *J Agric Biol Environ Stat* 19:405–418

- Dastbaravarde A, Arghami NR, Sarmad M (2016) Some theoretical results concerning non-parametric estimation by using a judgement post-stratification sample. *Commun Stat Theory Methods* 45:2181–2203
- Dell TR, Clutter JL (1972) Ranked set sampling theory with order statistics background. *Biometrics* 28:545–555
- Frey J (2016) A more efficient mean estimator for judgement post-stratification. *J Stat Comput Simul* 86(7):1404–1441
- Frey J, Feeman TG (2012) An improved mean estimator for judgment post-stratification. *Comput Stat Data Anal* 56:418–426
- Frey J, Feeman TG (2013) Variance estimation using judgment post-stratification. *Ann Inst Stat Math* 65:551–569
- Frey J, Ozturk O (2011) Constrained estimation using judgment post-stratification. *Ann Inst Stat Math* 63:769–789
- Kasprzak, P (2021) Feasibility Study of Ranked Set Sampling for Typical Modern Agronomy Trials Under Different Sampling Designs: Benefits and Challenges of Ranked Set Sampling. Thesis of Masters of Philosophy (Biometry), University of Adelaide
- MacEachern SN, Stasny EA, Wolfe DA (2004) Judgment post-stratification with imprecise rankings. *Biometrics* 60:207–215
- McIntyre GA (1952) A method for unbiased selective sampling using ranked set sampling. *Aust J Agric Res* 3:385–390
- McIntyre GA (2005) A method for unbiased selective sampling, using ranked sets. *Am Stat* 59:230–232
- Omidvar S, Jafari Jozani M, Nematollahi N (2018) Judgment post-stratification in finite mixture modeling: an example in estimating the prevalence of osteoporosis. *Stat Med* 37(30):4823–4836
- Ozturk O (2013) Combining multi-observer information in partially rank-ordered judgment post-stratified and ranked set samples. *Can J Stat* 41:304–324
- Ozturk O (2014) Statistical inference for population quantiles and variance in judgment post-stratified samples. *Comput Stat Data Anal* 77:188–205
- Ozturk O (2016a) Statistical inference based on judgment post-stratified samples in finite population. *Surv Methodol* 42:239–262
- Ozturk, (2016b) Estimation of finite population mean and total using population ranks of sample units. *J Agric Biol Environ Stat* 21:181–202
- Ozturk O (2017) Statistical inference with empty strata in judgment post stratified samples. *Ann Inst Stat Math* 69:1029–1057
- Ozturk O (2019a) Statistical inference using rank based post-stratified samples in a finite population. *Test* 28:1113–1143. <https://doi.org/10.1007/s11749-018-0618-y>
- Ozturk O (2019b) Post-stratified probability-proportional-to-size sampling from stratified populations. *J Agric Biol Environ Stat*. <https://doi.org/10.1007/s13253-019-00370-6>
- Ozturk O, Demirel N (2016) Estimation of population variance from multi-ranker ranked set sampling designs. *Commun Stat Simul Comput* 45(10):3568–3583
- Ozturk O, Bayramoglu Kavlak K (2018a) Statistical inference using stratified ranked set samples from finite populations. In: Bouza C, Al-Omari AI (eds) *Ranked set sampling: 65 years improving the accuracy in data gathering*. Elsevier, New York
- Ozturk O, Bayramoglu Kavlak K (2018b) Model based inference using ranked set samples in finite population setting. *Surv Methodol* 44:1–16
- Ozturk O, Rogers S, Kravchuk O, and Kasprzak, P (2021). RankedSetSampling: Easing the Application of Ranked Set Sampling in Practice. R package version 0.0.1. <https://biometryhub.github.io/RankedSetSampling/>
- Ozturk O, Rogers S, Kravchuk O, and Kasprzak, P (2021). RankedSetSampling: Easing the Application of Ranked Set Sampling in Practice. R package version 0.0.1. <https://biometryhub.github.io/RankedSetSampling/>
- Stokes SL, Wang X, Chen M (2007) Judgment post-stratification with multiple rankers. *J Stat Theory Appl* 6:344–359
- Takahasi K, Wakimoto K (1968) On unbiased estimates of the population mean based on sample stratified by means of ordering. *Ann Inst Math Stat* 20:1–31

- Wang X, Wang K, Lim J (2012) Isotonized CDF estimation from judgment post-stratification data with empty strata. *Biometrics* 68:194–202
- Wang X, Stokes SL, Lim J, Chen M (2006) Concomitants of multivariate order statistics with application to judgment post-stratification. *J Am Stat Assoc* 101:1693–1704
- Wang X, Lim J, Stokes SL (2008) A nonparametric mean estimator for judgment post-stratified data. *Biometrics* 64:355–363
- Zamanzade E, Wang X (2017) Estimation of population proportion for judgment post-stratification. *Comput Stat Data Anal* 112:257–269

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.