CBCS
HTTP://AGROBIOL.SGGW.WAW.PL/CBCS

COMMUNICATIONS IN BIOMETRY AND CROP SCIENCE
Vol. **11, NO. 2, 2016, PP. 164–180**

INTERNATIONAL JOURNAL OF THE FACULTY OF AGRICULTURE AND BIOLOGY,
Warsaw University of Life Sciences **– SGGW, POLAND**

REGULAR ARTICLE

How many trials are needed in crop variety testing?

Johannes Forkman

Swedish University of Agricultural Sciences, Department of Crop Production Ecology, Box 7043, SE-750 07 Uppsala, Sweden.

E-mail: johannes.forkman@slu.se

CITATION: Forkman J. (2016). How many trials are needed in crop variety testing? *Communications in Biometry and Crop Science* 11, 164–180.

Received: 19 April 2016, Accepted: 6 October 2016, Published online: 14 October 2016

ABSTRACT

A method is proposed for determining the required number of trials in crop variety testing. The main focus is on estimating the correct ranking of the varieties. Correlation between observed and correct ranks, as a function of the number of trials, can be studied by simulation. Similarly, the effect of the number of trials on the expected genetic advance under selection can be explored. The result from such simulations can be used as a basis for selecting the number of trials. Swedish one-year and multi-year series in spring barley and winter wheat are used as examples.

Key Words: *crop variety trials; response to selection; sample size calculation; value for cultivation and use; variance components.*

INTRODUCTION

In crop science, the following question is recurrent: How many trials are needed? The present article investigates this problem. Specifically, it is investigated how many trials are needed in series of Swedish spring barley and winter wheat variety trials. With a series of experiments, we mean a group of identical experiments, which will be analyzed together. It will be assumed that varieties are compared with regard to yield, but generally any continuous response variable can be considered.

One common approach for selecting the sample size is to choose the size such that interesting confidence intervals become as small as desired. In comparative experiments, differences between treatments are the main interest, which means that this approach would focus on confidence intervals for differences between treatment means. Equivalently, one could require that the least significant difference should be smaller than some specified limit. Forkman, Amiri and von Rosen (2012) used this approach for Swedish series of spring barley and winter wheat.

Another common approach for selecting the sample size is to choose the size such that a specified minimum power is reached. The power is the probability of rejecting the null hypothesis when the null hypothesis is false. Here, the null hypothesis postulates no differences between treatments or, in a simpler version, no difference between two treatments. The power is an increasing function of the true difference between these treatments. Consequently, the required sample size depends on the true difference between the treatments, which is not known. In practice, the researcher must specify a hypothetical true difference in order to compute the required number of trials. In clinical trials, this is "the clinically relevant difference" or "the difference one would not like to miss" (Senn 2002, p. 170), which is an applicable definition also in crop variety testing. Kupper and Hafner (1989) showed that this statistical power approach is more reliable than the confidence interval approach.

However, the main objective of crop variety testing is perhaps not to test hypotheses about treatment differences. Rather, the aim is to identify which varieties are best and which ones are less good. The varieties are ranked, i.e., the variety that produced the highest yield obtains rank 1, the variety that produced the second highest yield obtains rank 2, and so on. Researchers are often more interested in the observed ranking of the varieties than in which varieties differ significantly. The present article proposes a method to determine the number of trials, so that the ranking of the varieties become reliable with high probability.

The correct ranking of the varieties is the ranking that would have been obtained in a series with infinitely many trials. In practice, only a smaller number of trials can be performed. As a result, the ranking obtained in the series will usually not be completely correct. However, the more trials the series include, the better the ranking tends to be. In other words, when the number of trials is increased, observed ranks tend to correlate better with correct ranks.

Theoretically, the relationship between rank correlation and number of trials is not easily explored. The distribution of the rank correlation coefficient is complicated. For example, it has been shown that no tractable formula exists for the variance of the rank correlation coefficient when data is normally distributed (Kendall and Gibbons 1990, p. 170). It is much easier to study the rank correlation through computer simulation.

When correlation between correct ranks and simulated observed ranks is computed for series with N trials, it is found that median rank correlation increases with N. At the same time, variation in rank correlation decreases. In series with few trials, rank correlation might become high or low. By making many trials, the risk of a low rank correlation is reduced. The results section reports these findings.

Although this article is mainly focused on rank correlation, expected genetic advance under selection is also considered. Assuming the top varieties are selected, the expected genetic advance under selection can be defined as the difference between the mean of the selected varieties and the mean over all varieties (Galwey 2006). The expected genetic advance under selection increases with the number of trials. The Results section illustrates how this relationship can be explored by simulation. The outcomes can be used to determine the required number of trials.

The expected genetic advance under selection is a commonly used criterion in plant breeding. A much investigated question in plant breeding programs is how to define the rules for selecting genotypes in a multi-stage selection procedure. This question has often been investigated by simulation, and several software packages are available for determining designs for multi-stage breeding trials: the PLABSIM software package (Frisch, Bohn and Melchinger 2000), the PLABSOFT software package (Maurer, Melchinger and Frisch 2008), the MBP software package (Gordillo and Geiger 2008), and the selectiongain R package (Mi et al. 2014). Using these packages, genetic parameters and breeding budget can be considered in addition to design parameters. The software package SELSYS (Robinson 1984), was developed for official U.K. cereal variety testing. This package uses simulation and is intended for design of multi-stage selection series with several years and locations. The SELSYS program was written in FORTRAN, which is nowadays less used. Kleinknecht et al. (2016) provided a new and partly extended implementation of the SELSYS approach using the SAS System.

The present article shows how simulations can be carried out using the open source software R, when the focus is rank correlation or expected genetic advance under selection. One-year series and multi-year series are considered.

MATERIALS AND METHODS

STATISTICAL MODELS

In Sweden, during the five-year period 2011–2015, altogether 126 variety trials were performed in spring barley and 96 in winter wheat. Tables 1 and 2 detail the number of trials per year in these crops. These tables also include the number of varieties per trial. Some winter wheat trials with severe damage due to winter weather conditions were not included in Table 2. In Swedish variety testing, such trials with low survival rates are not included in analyses of series.

Table 1. Number of Swedish variety trials in spring barley and number of varieties per trial.

Year	Number of trials	Number of varieties
2015	26	33
2014	26	29
2013	26	12
2012	23	11
2011	25	23

In 2015, two of the trials included 16 varieties and one of the trials 15 varieties; in 2014, three of the trials included 17 varieties; and in 2013, one of the trials included 11 varieties.

Table 2. Number of Swedish variety trials in winter wheat and number of varieties per trial.

Harvest year	Number of trials	Number of varieties
2015	23	29
2014	24	37
2013	14	29
2012	18	32
2011	17	31

In 2012, one of the trials included 31 varieties, two of the trials included 30 varieties and one of the trials included 27 varieties; and in 2011, five of the trials included 30 varieties.

For each crop and year, the following linear random-effects model was fitted:

$$y_{ij} = \mu + t_i + v_j + e_{ij}, \qquad (1)$$

where y_{ij} is the yield of the jth variety in the ith trial, μ is the expected mean yield, t_i is a random effect of the ith trial, v_j is a random effect of the jth variety, and e_{ij} is a random residual error. In this model, $t_i \sim N(0, \sigma_T^2)$, $v_j \sim N(0, \sigma_V^2)$ and $e_{ij} \sim N(0, \sigma_E^2)$, and these terms are independent.

Multi-year analyses were performed, for the two crops separately, using the model

$$y_{ijk} = \mu + s_i + t_{ij} + v_k + (sv)_{ik} + e_{ijk}, \qquad (2)$$

where y_{ijk} is the yield of the kth variety in the jth trial of the ith year, μ is the expected mean yield, t_{ij} is a random effect of the jth trial of the ith year, v_k is a random effect of the kth variety, $(sv)_{ik}$ is a random effect of variety-by-year interaction, and e_{ijk} is a random residual error. In this model, $s_i \sim N(0, \sigma_S^2)$, $t_{ij} \sim N(0, \sigma_T^2)$, $v_k \sim N(0, \sigma_V^2)$, $(sv)_{ik} \sim N(0, \sigma_{SV}^2)$ and $e_{ij} \sim N(0, \sigma_E^2)$, and these terms are independent.

The observations, y_{ij} in (1) and y_{ijk} in (2), were least-squares means obtained in separate analyses of the trials. Thus, unweighted two-stage analyses (Möhring and Piepho, 2009) were applied. In the first stage, each experiment was analyzed using a linear mixed-effects model with fixed effects of varieties and random effects of replicates and incomplete blocks. The trials comprised four replicates. However, two of the replicates were treated against fungicide, primarily as protection against powdery mildew. For the present study, these replicates were not included in the analysis. Thus, results refer to yield of spring barley and winter wheat without protection against fungicide. Models were fitted using the restricted likelihood method (REML) and the mixed procedure of the SAS System (SAS Institute, 2008).

SIMULATION OF ONE-YEAR SERIES

In simulations of one-year data, parameter values were chosen based on the estimates (see the Results section) from the fits of Model (1). Parameters μ = 7400, σ_V^2 = 68 000, σ_E^2 = 130 000, and μ = 9500, σ_V^2 = 310 000, σ_E^2 = 500 000 were used, for spring barley and winter wheat, respectively. The size of σ_T^2 is of no importance for ranking of varieties. Trials with M varieties were simulated, where M was set to 25 and 30 for spring barley and winter wheat, respectively, since Swedish trials in spring barley often includes somewhat fewer varieties than Swedish trials in winter wheat (Tables 1 and 2).

Series with N = 1, 3, 5, 7, 10, 12, 15, 20, 25, 30, 50 and 100 trials were simulated. For each value of N, 1000 series were generated, each in the following way. First, M variety means, $\mu_1, \mu_2, \dots, \mu_M$, were simulated from a normal distribution with expected value μ and variance σ_V^2. Next, M normally distributed random numbers, $\hat{\mu}_1, \hat{\mu}_2, \dots, \hat{\mu}_M$, were generated with expected value μ_j, $j = 1, 2, \dots, M$, and variance σ_E^2/N. In this way, the "observed" variety means, $\hat{\mu}_1, \hat{\mu}_2, \dots, \hat{\mu}_M$, were simulated directly, since it was not necessary to first simulate the data according to model (1), and then calculate the averages. Spearman's rank correlation between $\mu_1, \mu_2, \dots, \mu_M$ and $\hat{\mu}_1, \hat{\mu}_2, \dots, \hat{\mu}_M$ was computed for each simulated series. For each crop and value of N, 1000 rank correlations were obtained. Their distributions were reported using quantiles and boxplots. Let $\hat{\mu}_{(1)}, \hat{\mu}_{(2)}$ and $\hat{\mu}_{(3)}$ denote the three highest values of $\{\hat{\mu}_1, \hat{\mu}_2, \dots, \hat{\mu}_M\}$, and let $\mu_{(1)}, \mu_{(2)}$ and $\mu_{(3)}$ denote their corresponding expected values. The genetic advance under selection was estimated as the difference between the average of $\mu_{(1)}, \mu_{(2)}$ and $\mu_{(3)}$ and the average of all μ_j, i.e., $\left(\mu_{(1)} + \mu_{(2)} + \mu_{(3)}\right)/3 - \sum_j^M \mu_j / M$. For each crop and value of N, 1000 estimates of the genetic advance under selection were obtained. The expected genetic advance under selection was estimated as the average of these 1000 estimates. Simulations were performed using R, version 3.1.1. Box 1 includes the basic code.

Box 1. R code for simulation of rank correlation

```
### Simulation of a series in spring barley with 30 trials and 25 varieties
# Parameters
mean <- 7400
s_A <- sqrt( 68000)
s_E <- sqrt(130000)
NVarieties <- 25
NTrials <- 30
NSimul <- 1000
# Simulation
vec.mu <- rnorm(n = NVarieties*NSimul, mean = mean, sd = s_A)
vec.Y <- rnorm(n = NVarieties*NSimul, mean = vec.mu, sd = s_E/sqrt(NTrials))
mu <- matrix(vec.mu, ncol = NSimul)
Y <- matrix(vec.Y, ncol = NSimul)
Spear <- matrix(NA, nrow = 1, ncol = NSimul)    # Spearman’s rank correlation
GAS <- matrix(NA, nrow = 1, ncol = NSimul)      # Genetic advance under selection
for (j in 1:NSimul){
  Spear[ , j] <- cor(Y[ , j], mu[ , j], method = "spearman")
  GAS[ , j] <- (mu[ , j][rank(Y[ , j])== NVarieties] + mu[ , j][rank(Y[ , j])==
NVarieties-1]
  + mu[ , j][rank(Y[ , j])== NVarieties-2])/3 - mean(mu[ , j])}
# Result
quantile(Spear, probs = c(0.05, 0.10, 0.25, 0.5, 0.75, 0.90, 0.95))
boxplot(t(Spear), ylim = c(0, 1))
abline(h = seq(0, 1,0.1), v = 1, col = "gray", lty = 3)
mean(GAS)
```

SIMULATION OF FIVE-YEAR SERIES

For simulation of five-year data, parameter values were chosen based on the estimates (see the Results section) from the fits of Model (2). Parameters μ = 7500, σ_V^2 = 75 000, σ_S^2 = 141 000, σ_{VS}^2 = 7600, σ_T^2 = 1 655 000 σ_E^2 = 131 000, and μ = 9500, σ_V^2 = 367 000, σ_S^2 = 763 000, σ_{VS}^2 = 52 000, σ_T^2 = 2 304 000 σ_E^2 = 542 000 were used, for spring barley and winter wheat, respectively. In Swedish variety testing, a variety must have been tested the last year and at least one additional year, to be included in the five-year analysis. Every year, less performing varieties are replaced by new varieties. For this reason, five-year data is unbalanced. To mimic a typical five-years series, it was for spring barley assumed seven varieties were present all five years, five were present only the four last years, five only the three last years and eight only the two last years. For winter wheat, it was assumed ten varieties were present all five years, five were present only the four last years, five only the three last years, and ten only the two last years. Thus, M = 25 varieties were tested in spring barley, and M = 30 varieties were tested in winter wheat. For simplicity, it was assumed that exactly N trials were performed each year. Series with N = 1, 3, 5, 7, 10, 12, 15, 20, 25, 30, 50 and 100 trials were simulated using model (2).

In Swedish variety testing, five-year data is analyzed using Model (2), but with fixed effects of varieties. In the simulation study, this mixed-effects model was fitted using the lmer function of R, with the default REML method. Least square means, $\hat{\mu}_1, \hat{\mu}_2, \dots, \hat{\mu}_M$, were computed for the varieties. These were compared with the true means: $\mu_1, \mu_2, \dots, \mu_M$ generated through the simulation. The kth true mean, μ_k, is the simulated intercept plus the simulated random effect of the the kth variety. Spearman's rank correlation between $\mu_1, \mu_2, \dots, \mu_M$ and $\hat{\mu}_1, \hat{\mu}_2, \dots, \hat{\mu}_M$ was computed for each simulated series. For each crop and value of N, 1000 rank correlations were obtained.

POPULATION CORRELATION COEFFICIENTS

For one-year series, population correlation coefficients can be computed exactly. Given the variances σ_V^2 and σ_E^2, the Pearson correlation coefficient ρ can be computed, for various values of N, using the equation

$$\rho = \frac{Cov(\mu_i, \hat{\mu}_i)}{\sqrt{\mathrm{var}(\mu_i)\mathrm{var}(\hat{\mu}_i)}} = \frac{\sigma_V^2}{\sqrt{\sigma_V^2(\sigma_V^2 + \sigma_E^2/N)}}, \tag{3}$$

Given (3), the expected value of the Spearman rank correlation ρ_S can be computed using the exact equation (Moran, 1948)

$$E(\rho_S) = \frac{6}{\pi}\left(\frac{M-2}{M+1}\arcsin\frac{\rho}{2} + \frac{1}{M+1}\arcsin\rho\right). \tag{4}$$

RESULTS

PARAMETER ESTIMATES

Tables 3a and 4a reports, for spring barley and winter wheat, respectively, estimates of expected mean yield and variance components using model (1). Tables 3b and 4b presents the estimates of the variance components expressed as coefficients of variation: $\mathrm{CV} = \sqrt{\sigma^2}/\mu$.

Table 3a. Estimates, for spring barley, of expected yield μ (kg/ha), variance between varieties σ_V^2, variance between trials σ_T^2, and residual variance σ_E^2.

Parameter	2015	2014	2013	2012	2011	Mean
μ	7912	7462	7561	7620	6560	7423
σ_V^2	141632	76009	18187	54707	49534	68014
σ_T^2	926594	2623146	1729659	1599108	1378978	1651497
σ_E^2	149648	113014	104356	161065	117829	129183

Table 3b. Variances of Table 3a (spring barley) expressed as coefficients of variation (CV).

	2015	2014	2013	2012	2011	Mean
Variety	0.048	0.037	0.018	0.031	0.034	0.035
Trial	0.122	0.217	0.174	0.166	0.179	0.173
Error	0.049	0.045	0.043	0.053	0.052	0.048

Table 4a. Estimates, for winter wheat, of expected yield μ (kg/ha), variance between varieties σ_V^2, variance between trials σ_T^2, and residual variance σ_E^2.

Parameter	2015	2014	2013	2012	2011	Mean
μ	10714	10279	9433	8917	8106	9490
σ_V^2	291248	289494	189195	656820	122160	309783
σ_T^2	1395589	1817879	2895728	2001840	4105331	2443273
σ_E^2	417852	761293	279357	799923	241835	500052

Table 4b. Variances of Table 4a (winter wheat) expressed as coefficients of variation (CV).

	2015	2014	2013	2012	2011	Mean
Variety	0.050	0.052	0.046	0.091	0.043	0.059
Trial	0.110	0.131	0.180	0.159	0.250	0.165
Error	0.060	0.085	0.056	0.100	0.061	0.075

On average, coefficient of variation for error was larger in winter wheat (0.075) than in spring barley (0.048). This was expected, since winter weather can have a large impact on variation. It was noted that coefficient of variation for varieties was also larger for winter wheat (0.059) than for spring barley (0.035).

Considering multi-year analysis, Table 5 reports estimates of expected mean yield and the variance components using model (2). Variety-by-year interaction was larger in winter wheat than in spring barley.

Table 5. Estimates of expected yield μ (kg/ha), variance between varieties σ_V^2, variance between years σ_S^2, variance for variety-by-year interaction σ_{VS}^2, variance between trials σ_T^2, and residual variance σ_E^2. Variances are also expressed as coefficients of variation (CV).

Parameter	Spring barley	CV	Winter wheat	CV
μ	7457		9463	
σ_V^2	74826	0.037	366546	0.064
σ_S^2	140860	0.050	763174	0.092
σ_{VS}^2	7658	0.012	52027	0.024
σ_T^2	1655329	0.173	2304336	0.160
σ_E^2	131380	0.049	542054	0.078

RANK CORRELATION IN ONE-YEAR SERIES

Table 6 shows obtained correlations between estimated and correct ranks in one-year series with N trials, for spring barley. The 50th percentile is the median. The 25th and 75th percentiles are also known as the first and third quartile, respectively. For example, when series with N = 10 trials were simulated, the median of the obtained 1000 rank correlations was 0.90. Since the fifth percentile was 0.80, the rank correlation was smaller than 0.80 in 5% of the simulated series of N = 10 trials.

Table 6. Correlation between estimated ranks and correct ranks was computed in 1000 simulated one-year series of N trials in spring barley. This was made for twelve values of N. The table reports percentiles for the distributions of the obtained rank correlations.

	Percentile						
N	5	10	25	50	75	90	95
1	0.28	0.36	0.46	0.57	0.66	0.74	0.77
3	0.56	0.61	0.69	0.76	0.82	0.86	0.88
5	0.68	0.71	0.77	0.84	0.88	0.90	0.92
7	0.74	0.77	0.82	0.87	0.90	0.93	0.94
10	0.80	0.83	0.86	0.90	0.93	0.95	0.95
12	0.82	0.85	0.88	0.91	0.94	0.95	0.96
15	0.85	0.87	0.90	0.93	0.95	0.96	0.97
20	0.87	0.89	0.92	0.94	0.96	0.96	0.97
25	0.89	0.91	0.93	0.95	0.96	0.97	0.98
30	0.91	0.92	0.94	0.96	0.97	0.98	0.98
50	0.93	0.95	0.96	0.97	0.98	0.99	0.99
100	0.96	0.97	0.98	0.98	0.99	0.99	0.99

Table 7 shows obtained correlations between estimated and correct ranks in one-year series with N trials, for winter wheat. Still using series with 10 trials as an example, the median rank correlation was 0.91 for winter wheat, and in 5% of the 1000 simulated series, the rank correlation was lower than 0.83.

Table 7. Correlation between estimated ranks and correct ranks was computed in 1000 simulated one-year series of N trials in winter wheat. This was made for twelve values of N. The table reports percentiles for the distributions of the obtained rank correlations.

	Percentile						
N	5	10	25	50	75	90	95
1	0.34	0.41	0.51	0.60	0.68	0.73	0.78
3	0.62	0.66	0.73	0.79	0.84	0.87	0.88
5	0.72	0.76	0.81	0.85	0.89	0.91	0.92
7	0.78	0.81	0.85	0.89	0.91	0.93	0.94
10	0.83	0.85	0.89	0.91	0.94	0.95	0.96
12	0.85	0.87	0.90	0.93	0.94	0.96	0.96
15	0.87	0.89	0.92	0.94	0.95	0.96	0.97
20	0.90	0.93	0.94	0.95	0.96	0.98	0.98
25	0.92	0.94	0.95	0.96	0.97	0.98	0.98
30	0.93	0.94	0.95	0.96	0.97	0.98	0.98
50	0.95	0.96	0.97	0.98	0.98	0.99	0.99
100	0.97	0.97	0.98	0.99	0.99	0.99	0.99

Generally, rank correlation increases with the number of trials. At the same time, variation in rank correlation decreases. This is evident from Tables 6 and 7, but is even more obvious in Figures 1 and 2, which report simulation results as box plots, for spring barley and winter wheat, respectively. In these box plots, the median is indicated as a bold horizontal line within the rectangular boxes. The boxes are limited upwards by the 75th percentile and downwards by the 25th percentile. Thus, 50% of simulated rank correlations are included in the box and the width of the box is the interquartile range, which is a measure of the variation. The boxplot whiskers extend out from the box at most one and a half interquartile range.

The Appendix exemplifies what correlation may look like for various rank correlations. These examples illustrate the importance of the rank correlation being high. For example, Figure A5 is a plot of observed ranks versus correct ranks in one of the simulated series containing 25 varieties of spring barley. The observed ranks are the ranks obtained in the simulated series. The correct ranks are the ranks that theoretically would have been observed if the series included infinitely many trials. In this series the rank correlation was 0.80. Although this rank correlation might seem high, the second best variety (i.e., the correct rank is 2), was observed as the seventh best variety in the series (the observed rank is 7). Furthermore, the 18th best variety out of 25 (correct rank 18) was observed as the fifth best variety in the series (observed rank 5).

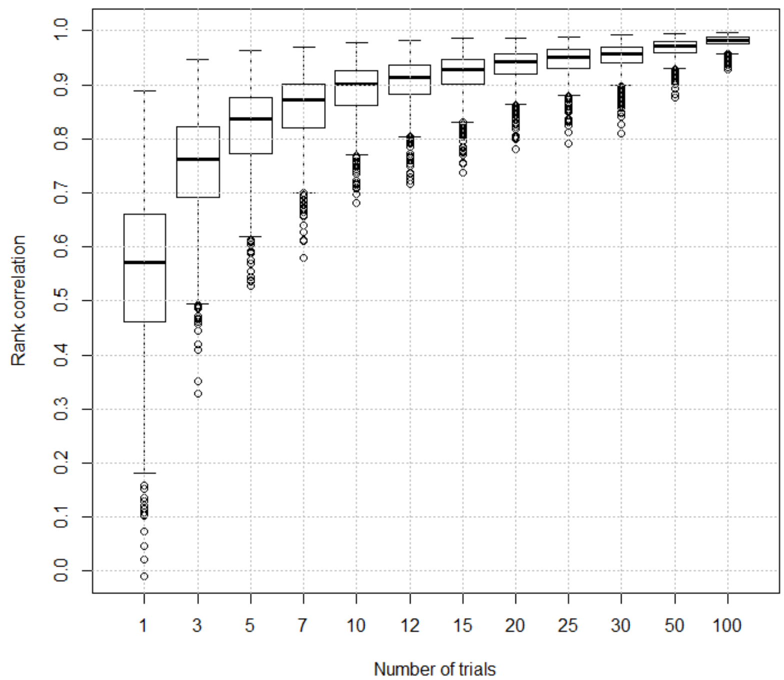

Figure 1. Correlation between estimated ranks and correct ranks was computed in 1000 simulated one-year series of N trials in spring barley. This was made for twelve values of N. The figure shows box plots for the distributions of the obtained rank correlations.

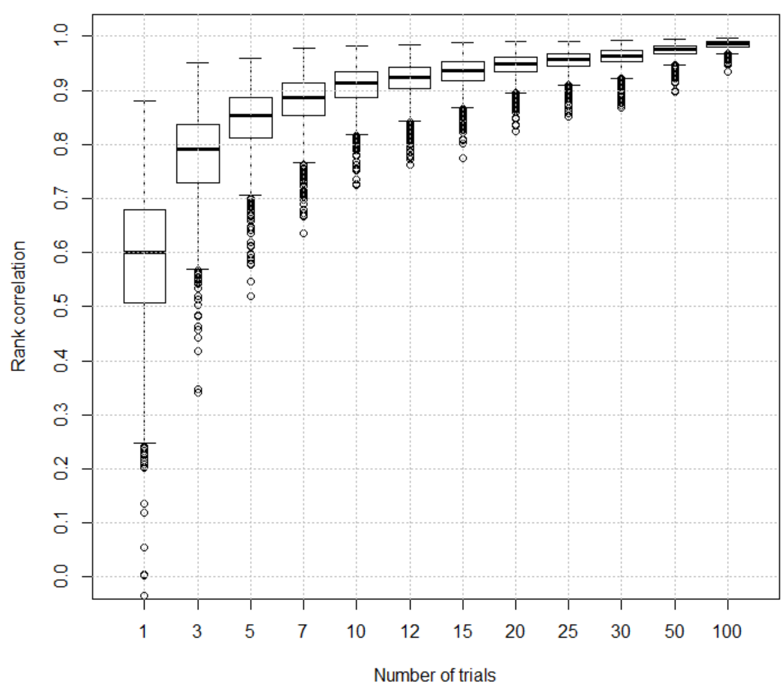

Figure 2. Correlation between estimated ranks and correct ranks was computed in 1000 simulated one-year series of N trials in winter wheat. This was made for twelve values of N. The figure shows box plots for the distributions of the obtained rank correlations.

Using (3) and (4), it is easily confirmed that the distributions shown in Figures 1 and 2 agree well with the theoretical population rank correlation coefficients. For example, in a one-year spring-barley series with $N = 3$ trials, the population Pearson correlation coefficient is

$$\rho = \frac{68\,000}{\sqrt{68\,000(68\,000 + 130\,000/3)}} = 0.7815\,,$$

and the population Spearman correlation coefficient is

$$\rho_S = \frac{6}{\pi}\left(\frac{25-2}{25+1}\arcsin\frac{0.7815}{2} + \frac{1}{25+1}\arcsin 0.7815\right) = 0.7442\,.$$

However, Figure 1 tells much more about the distribution of the rank correlation coefficient, than this single number does.

RANK CORRELATION IN FIVE-YEAR SERIES

Tables 8 and 9 present results for simulation of five-year series. These tables list the percentiles of the observed rank correlations, for spring barley and winter wheat, respectively. The same distributions are illustrated using boxplots in Figures 3 and 4. In these tables and figures, N is the number of trials per year. In five-year series, the distributions are similar when N is larger than 10.

Table 8. Correlation between estimated ranks and correct ranks was computed in 1000 simulated five-year series with N trials/year in spring barley. This was made for twelve values of N. The table reports percentiles for the distributions of the obtained rank correlations.

	Percentile						
N	5	10	25	50	75	90	95
1	0.57	0.62	0.70	0.77	0.83	0.87	0.89
3	0.77	0.80	0.85	0.89	0.92	0.94	0.94
5	0.83	0.85	0.89	0.91	0.94	0.95	0.96
7	0.86	0.88	0.90	0.93	0.95	0.96	0.97
10	0.88	0.90	0.93	0.94	0.96	0.97	0.98
12	0.88	0.90	0.93	0.95	0.96	0.97	0.98
15	0.89	0.91	0.93	0.95	0.97	0.98	0.98
20	0.91	0.92	0.94	0.96	0.97	0.98	0.98
25	0.91	0.93	0.95	0.96	0.97	0.98	0.98
30	0.92	0.93	0.95	0.96	0.97	0.98	0.98
50	0.92	0.94	0.95	0.97	0.98	0.98	0.99
100	0.93	0.94	0.96	0.97	0.98	0.98	0.99

Table 9. Correlation between estimated ranks and correct ranks was computed in 1000 simulated five-year series with N trials/year in winter wheat. This was made for twelve values of N. The table reports percentiles for the distributions of the obtained rank correlations.

	Percentile						
N	5	10	25	50	75	90	95
1	0.63	0.66	0.72	0.79	0.84	0.88	0.90
3	0.79	0.82	0.86	0.89	0.92	0.94	0.95
5	0.84	0.86	0.89	0.92	0.94	0.95	0.96
7	0.86	0.88	0.91	0.93	0.95	0.96	0.97
10	0.88	0.90	0.92	0.94	0.96	0.97	0.97
12	0.89	0.91	0.93	0.95	0.96	0.97	0.98
15	0.90	0.91	0.93	0.95	0.96	0.97	0.98
20	0.91	0.92	0.94	0.95	0.97	0.97	0.98
25	0.91	0.92	0.94	0.96	0.97	0.98	0.98
30	0.91	0.93	0.94	0.96	0.97	0.98	0.98
50	0.92	0.93	0.95	0.96	0.97	0.98	0.98
100	0.93	0.94	0.95	0.96	0.98	0.98	0.98

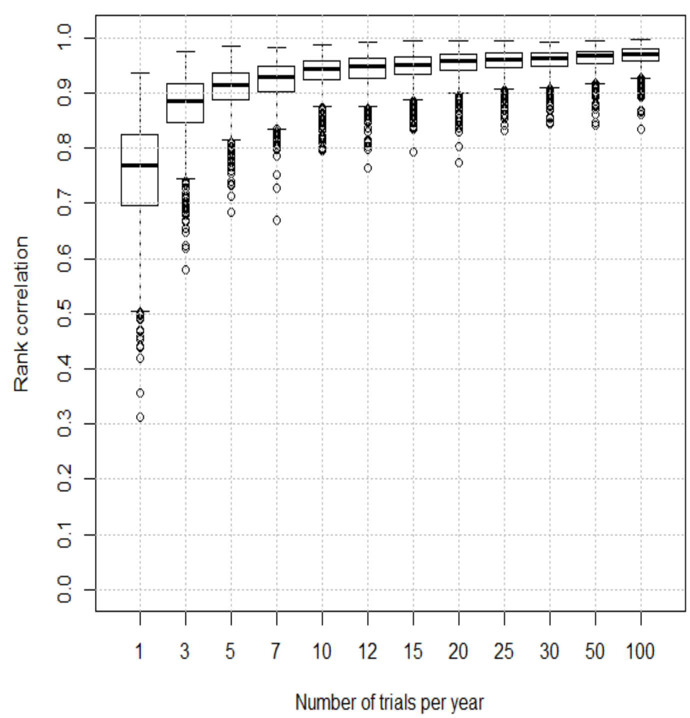

Figure 3. Correlation between estimated ranks and correct ranks was computed in 1000 simulated five-year series of N trials per year in spring barley. This was made for twelve values of N. The figure shows box plots for the distributions of the obtained rank correlations.

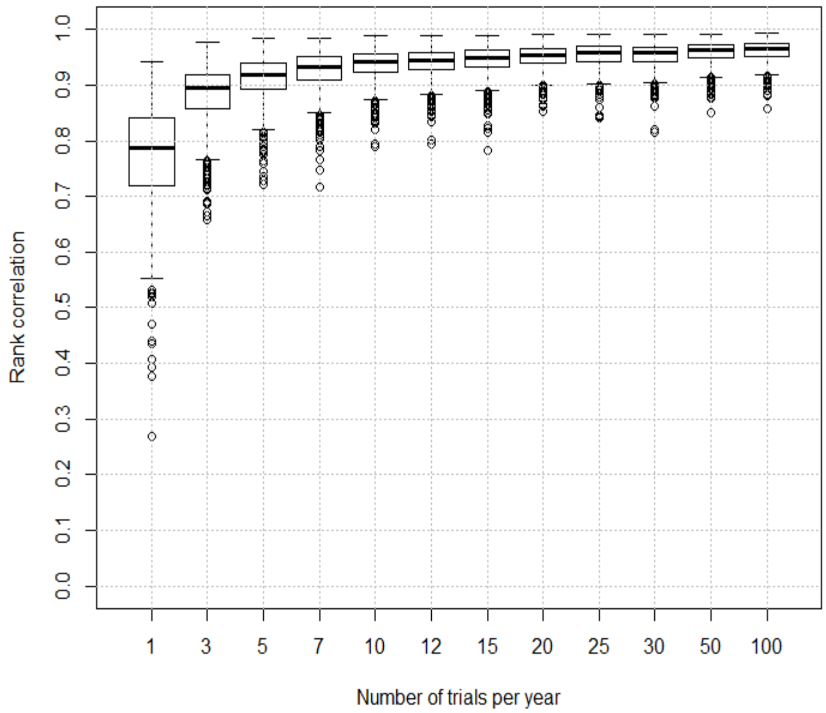

Figure 4. Correlation between estimated ranks and correct ranks was computed in 1000 simulated five-year series of N trials per year in winter wheat. This was made for twelve values of N. The figure shows box plots for the distributions of the obtained rank correlations.

Table 10. Estimated expected genetic advance under selection (kg/ha) by the number of trials (N). In one-year series, N is the total number of trials. In five-year series, N is the number of trials per year.

	One-year series		Five-year series	
N	Spring barley	Winter wheat	Spring barley	Winter wheat
1	250	601	334	831
3	329	766	393	912
5	356	824	406	942
7	370	850	402	962
10	384	876	419	970
12	389	887	418	961
15	394	896	415	971
20	399	907	422	973
25	402	913	421	984
30	406	917	421	979
50	411	927	424	978
100	414	935	421	994

EXPECTED GENETIC ADVANCE UNDER SELECTION

Table 10 reports estimates of the expected genetic advance under selection, provided the top three varieties are selected. The expected genetic advance under selection is increasing with N. As a function of N, the expected genetic advance under selection is to begin with increasing fast when N grows from 1 upwards, but as N grows, the growth rate of the genetic advance under selection decreases.

DISCUSSION

The researcher must decide the reasonable number of trials in a series, provided that a high rank correlation is desired. From the figures of this article, including those in the Appendix, it should be clear that the results of a series of crop variety trials can be misleading if the series includes only a small number of trials.

For example, if three Swedish spring barley trials with 25 varieties are analyzed together, the probability of a rank correlation smaller than 0.61 is 10% (Table 6). This computation assumes the variance components of Table 3a. Figure A3 in the Appendix, shows what results might look like when the rank correlation is as small as 0.60, which is quite probable for an experiment with three trials. In that particular case, including 25 varieties, the variety that yielded the most in the experiment (observed rank 1) was actually the 17th best variety (correct rank 17). More trials would have revealed that this variety was not one of the top varieties, but just the 17th best out of 25.

Results of a single trial can be very misleading. In a single trial, the rank correlation between observed and correct ranks is often smaller than 0.5, which is not much informative (see Figure A1 for an example). In unfortunate cases, the rank correlation based on a single trial can even be negative (Figures 1 and 2), which in practice means that high-performing varieties can perform poorly in a single trial.

If we would like the rank correlation to be higher than 0.90 with probability 95%, then 20 trials are needed in winter wheat (Table 7), because the fifth percentile is 0.90 for one-year series with 20 trials. In spring barley, more than 25 but less than 30 trials are needed for the same specification (Table 6). If we would instead like the rank correlation to be higher than 0.85 in nine cases out of ten, then 10 trials are needed for a one-year series in winter wheat (Table 7). In spring barley, one-year series with 12 trials meet this requirement (Table 6).

The inference space is different in a multi-year series than in a one-year series. In a multi-year series, long-term differences between varieties are investigated through a “sample” of years. This is not the case in a one-year series, for in such it is impossible to draw any conclusions about other years than the one investigated. In five-year series, the distributions were similar for all investigated N larger than 10. Thus, for ranking of varieties based on five-year series, it does not help much to include more than 10 trials per year. The precision, however, would be improved by increasing the number of years.

Note that the numbers reported in Tables 6–9 are not probabilities of obtaining a completely correct ranking. The reported numbers are rank correlations. The ranking of the varieties is completely correct only when the rank correlation is exactly 1. This occurs very rarely. In one-year series with 100 trials, an incorrect ranking is obtained with at least probability 95% (Tables 6 and 7). Also, in five-year series with 100 trials per year, an incorrect ranking is obtained with at least probability 95% (Tables 8 and 9).

In Swedish variety trials, coefficient of variation for error is usually larger in winter wheat than in spring barley. For this reason, series in winter wheat were expected to require more trials than series in spring barley. However, for estimation of the correct ranking of varieties, this study revealed that more trials are needed in spring barley than in winter wheat. The reason for this surprising result is that not only error variation is larger in winter wheat, but also varieties differ more with regard to yield in winter wheat than in spring barley. The correct ranking is easier to estimate when differences between varieties are large, as in winter wheat, than when differences are smaller as in spring barley. However, also the size of the error variation is important. In multi-year series, the size of the variety-by-year interaction is crucial. The present article shows how the relative importance of these variance components can be assessed by simulation.

Since variance components vary between regions, it is advisable to perform the simulation study using variances that are typical for the region of interest. For example, a separate analysis was made for the subset of winter wheat trials performed in Sweden's southernmost province, Scania. The results of this study are not reported here, but it turned out that inter-variety variation was larger in Scania than in the rest of the country. Winter weather in southern Sweden was such that some sensitive varieties did not fare well in Scania, while other varieties fared better, especially in 2012 and 2014. At the same time, error variance was about as big in Scania as in Sweden's other farming areas. For this reason, it was concluded that for one-year series of winter wheat, somewhat less trials are needed if the trials are performed in Scania than if the trials are performed in more northern regions of Sweden.

Some winter wheat trials with severe damage due to winter weather conditions were excluded from this study, since in Sweden such trials are not included in analyses of series. This may be one explanation for winter wheat variance components being smaller in this study than in the study performed by Forkman, Amiri and von Rosen (2012).

The effects of the varieties were assumed to be normally distributed. In fact, some of the years the distributions of variety means were negatively skewed. This occurred in both crops, but especially in winter wheat in 2012 and 2014. These years, some of the winter wheat varieties gave considerable lower yield than the most. When the distribution is negatively skewed like this it is easier to rank varieties giving low yield than varieties giving high yield. If it is particularly important to rank high-performing varieties correctly and the distribution of variety means can be negatively skewed, then it is advisable to include somewhat more trials in the series than this study indicates.

Acknowledgments

I thank Dr. Jannie Hagman for having initiated this work, and an anonymous reviewer for constructive suggestions.

References

Forkman J., Amiri S., von Rosen D. (2012). Effect of region on the uncertainty in crop variety trial programs with a reduced number of trials. *Euphytica* 186, 489–500.

Frisch M., Bohn M., Melchinger A.E. (2000). Computer note. PLABSIM: Software for simulation of marker-assisted back-crossing. *Journal of Heredity* 91, 86–87.

Galwey N.W. (2006). Introduction to mixed modelling: beyond regression and analysis of variance. Chichester: Wiley.

Gordillo G.A., Geiger H.H. (2008). MBP (Version 1.0): A software package to optimize maize breeding procedures based on doubles haploid lines. *Journal of Heredity* 99, 227–231.

Kendall M., Gibbons J.D. (1990). *Rank correlation methods. 5th ed.* New York: Oxford University Press.

Kleinknecht K., Möhring J., Laidig F., Meyer U., Piepho H.P. (2016). A simulation-based approach for evaluating the efficiency of multienvironment trial desigs. *Crop Science* 56, 1–14.

Kupper L.L., Hafner K.B. (1989). How appropriate are popular sample size formulas? *The American Statistician* 43, 101–105.

Maurer H.P., Melchinger A.E., Frisch M. (2008). Population genetic simulation and data analysis with Plabsoft. *Euphytica* 161, 133–139.

Mi X., Utz H.F., Technow F., Melchinger A.E. (2014). Optimizing resource allocation for multistage selection in plant breeding with R package Selectiongain. *Crop Science* 54, 1413–1418.

Moran P.A.P. (1948). Rank correlation and product-moment correlation. *Biometrika* 35, 203–206.

Möhring J., Piepho H.P. (2009). Comparison of weighting in two-stage analysis of plant breeding trials. *Crop Science* 49, 1977–1988.

Robinson D.L. (1984). A study on sequential variety selection systems. *Journal of Agricultural Science* 102, 119–126.

SAS Institute (2008). *SAS/STAT 9.2 User's Guide: the Mixed Procedure (book excerpt)*. Cary, NC: SAS Institute.

Senn S. (2002). *Statistical Issues in Drug Development*, Chichester: Wiley.

APPENDIX

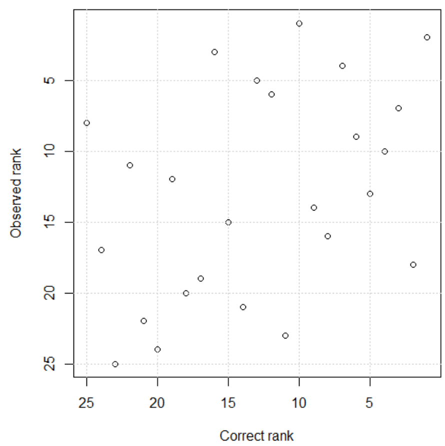

Figure A1. Rank correlation 0.40.

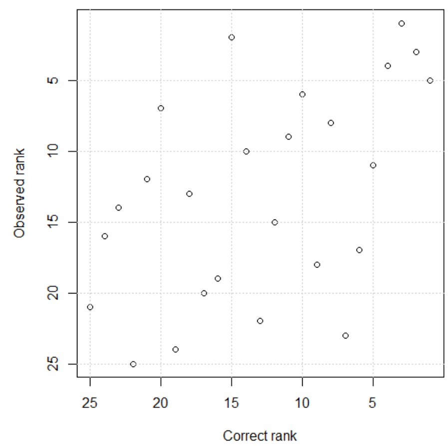

Figure A2. Rank correlation 0.50.

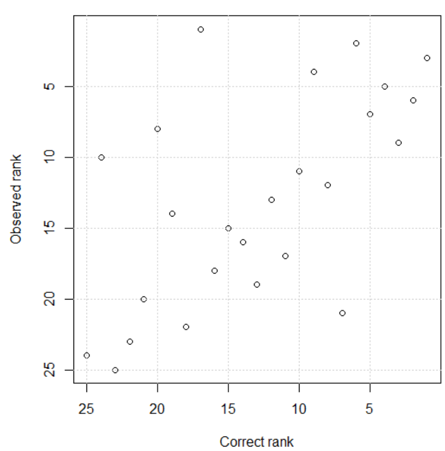

Figure A3. Rank correlation 0.60.

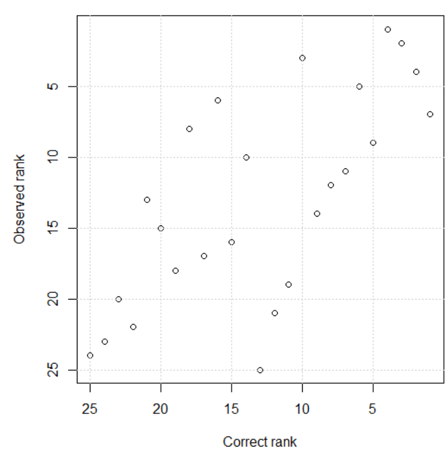

Figure A4. Rank correlation 0.70.

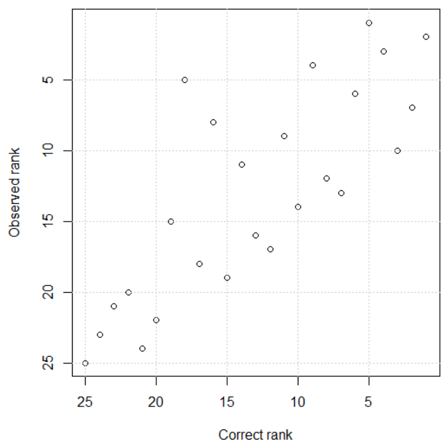

Figure A5. Rank correlation 0.80.

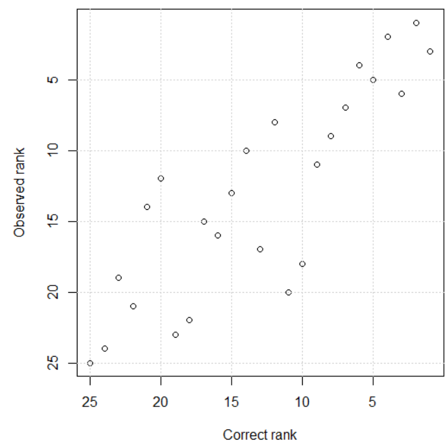

Figure A6. Rank correlation 0.85.

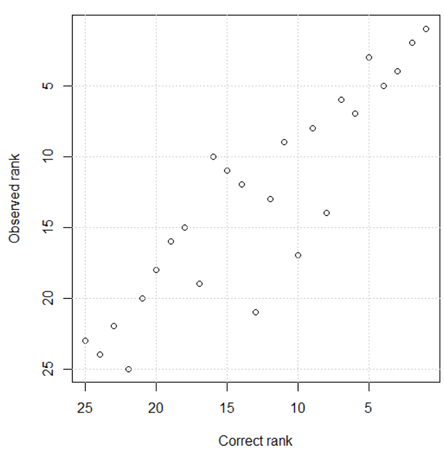

Figure A7. Rank correlation 0.90.

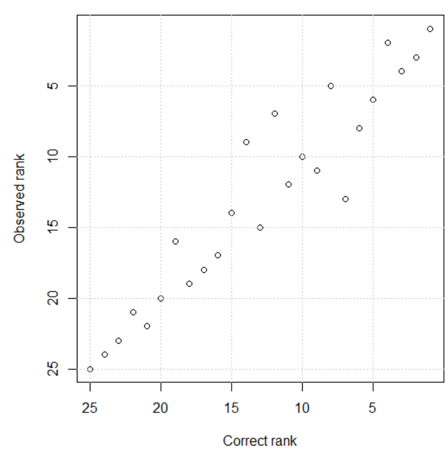

Figure A8. Rank correlation 0.95.