

# Step-up and Step-down Procedures Controlling the Number and Proportion of False Positives

Paul N. Somerville  
University of Central Florida

Claudia Hemmelmann<sup>1</sup>  
Institute of Medical Statistics, Computer Sciences and Documentation  
Friedrich-Schiller University of Jena

## Abstract

In multiple hypotheses testing, it is important to control the probability of rejecting “true” null hypotheses. A standard procedure has been to control the family-wise error rate (FWER), the probability of rejecting at least one true null hypothesis.

For large numbers of hypotheses, using FWER can result in very low power for testing single hypotheses. Recently, powerful multiple step FDR procedures have been proposed which control the “False Discovery Rate” (expected proportion of Type I errors). More recently van der Laan, Dudoit and Pollard (2004) proposed controlling a generalized family-wise error rate  $k$ -FWER (also called  $g$ FWER( $k$ )), defined as the probability of at least  $(k+1)$  Type I errors ( $k=0$  for the usual FWER).

Lehmann and Romano (2005) suggested both a single-step, and a step-down procedure for controlling the generalized family-wise error rate. They make no assumptions concerning the  $p$ -values of the individual tests. The step-down procedure is simple to apply, and cannot be improved without violation of control of the  $k$ -FWER.

In this paper, by limiting the number of steps in step-down or step-up procedures, new procedures are developed to control  $k$ -FWER (and the proportion of false positives). Using data from the literature, the procedures are compared with those of Romano and Lehmann (2005), and, under the assumption of a multivariate normal distribution of the test statistics, show considerable improvement in the reduction of the number and proportion of false positives.

Keywords: multiple testing, generalized family-wise error rate, step-up and step-down procedures, reduced step procedures, false discovery rate, proportion of false positives.

## 1. Introduction

In multiple testing, the problem is to test a set, or family of hypotheses  $H_1, H_2, \dots, H_m$ . Let  $t_i$  be the test statistic corresponding to  $H_i$ . In stepwise testing, the  $t$ 's are compared with critical constants  $d_1 \leq d_2 \leq \dots \leq d_m$ , and tested sequentially, one at a time, either stopping or continuing to the next step depending on the result observed. Testing may be step-up or step-down. Stepwise testing has been in use for a long time, for example, the Newman-Keuls test and Duncan's (1951) multiple range test.

In a series of papers, Dunnett and Tamhane (1991, 1992, 1995) developed stepwise multiple testing methods (both step-up and step-down) that are appropriate when normal distribution theory applies. The critical constants were calculated so that the family-wise error rate FWER was controlled at a level  $\alpha$ .

When  $m$ , the number of hypotheses is very large, such as in micro-array analyses, using FWER can result in very low power for an individual test. Benjamini and Hochberg (1995) proposed controlling FDR, the “False Discovery Rate”, instead of the FWER. The false discovery rate is defined to be the expected value of the proportion of rejected hypotheses which are actually true, with the false discovery rate defined to be zero when no hypotheses are rejected. They proposed a step-up FDR procedure, valid for independent statistics. Benjamini and Liu (1999) presented a step-down FDR procedure valid under the same conditions. Sarkar (2002) showed that both procedures were valid under positive dependency. Troendle (2000) developed both step-up and step-down FDR procedures which asymptotically control the FDR when the test statistics have a multivariate- $t$  distribution. Somerville (2004c), assuming a multivariate- $t$

---

<sup>1</sup> Work supported by DFG Project VO 683/2-1

distribution and common correlation of the test statistics used least favorable configurations to develop step-down and step-up FDR procedures. Somerville's step-down procedure yielded the same critical values as Troendle, except that instead of setting negative values to zero, Somerville used a Minimum Critical Value (MCV). Both possess asymptotic FDR control. Extensive calculations strongly supported the assumed locations of the least favorable configurations. However a proof was not realized. The FDR procedures of Troendle and Somerville have been shown by Horn and Dunnett (2004) and Somerville (2004c) to have superior power, under the assumption that the test statistics have a multivariate-t distribution.

Van der Laan, Dudoit and Pollard (2004) proposed the generalized family-wise error rate  $gFWER(k)$ . It is defined as the probability of at least  $(k+1)$  Type 1 errors ( $k=0$  for the usual FWER), i.e.  $P[U \leq k] \geq 1 - \alpha$  where  $U$  is the number of true hypotheses which are rejected. Augmentation procedures proposed by van der Laan, Dudoit and Pollard (2004) and Dudoit, van der Laan and Birkner (2004) can then be used to control PFP (proportion of false positives), where  $P[PFP \leq \gamma] \geq 1 - \alpha$  for some pre-specified  $\gamma$ .

More recently Lehmann and Romano (2005) suggested new methods of controlling  $gFWER(k)$  (called by them  $k$ -FWER) and PFP (called by them FDP - False Discovery Proportion). Both single-step and step-down procedures were derived which control the  $k$ -FWER. The procedures make no assumptions concerning the dependence structure or the  $p$ -values of the individual tests. The step-down procedure is simple to apply, and cannot be improved without violation of control of the  $k$ -FWER. Lehmann and Romano (2005) also proposed two methods for controlling the PFP. The first holds under "mild conditions on the dependence structure of  $p$ -values" while the second requires no dependence assumptions.

Somerville (2004c) noticed that by limiting the number of steps in his procedures, the FDR could be reduced with a small subsequent loss of power. In this paper we use the methodology of Somerville (2004c), and by limiting the maximum number of steps in step-down and step-up procedures, develop new procedures to control  $k$ -FWER. Using the data of Perou et al (2000) and the data of Hedenfalk et al (2001) in the New England Medical Journal, the methods are compared with those of Lehmann and Romano (2005). Under the assumption for the test statistics, outlined in Theorem 6.1 of section 6, using the methodology developed here, considerable improvement in the reduction of false positives, and in control of PFP, was accomplished.

The procedures developed in this paper can be applied, beginning with an arbitrary set of critical constants  $d_1 \leq d_2 \leq \dots \leq d_m$ . (The constants need not be related to any step-down or step-up procedure, and if they are, the procedure need not have satisfied either an FDR or FWER requirement). Using these arbitrary critical constants,  $d_1 \leq d_2 \leq \dots \leq d_m$ , and the values of the user's test statistics (assumed to possess a multivariate-t distribution with a common correlation value) the Fortran 95 program FDRPWR can be used to determine the maximum number of steps in a step-down or step-up procedure, for which  $P[U \leq k] \geq 1 - \alpha$ . The values  $k$  and  $\alpha$  may be selected by the user.

We develop and illustrate the procedures using the critical constants of Somerville (2004). Special approximation formulas are developed for the case where the critical constants are those of Somerville (2004).

## 2. Step-down and Step-up Procedures

Step-down test procedures may be described as follows. Let  $t_1, t_2, \dots, t_m$  be the test statistics corresponding to the null hypotheses  $H_1, H_2, \dots, H_m$ . Denote by  $T_i$  the random variable associated with  $t_i$ . Let  $T_{(1)} \leq T_{(2)} \leq \dots \leq T_{(m)}$  be the ordered values for the test statistics and denote the corresponding hypotheses  $H_{(1)}, H_{(2)}, \dots, H_{(m)}$ . If  $T_{(m)} \geq d_m$ , reject  $H_{(m)}$  and continue for  $i = m-1, m-2, \dots$ , comparing  $T_{(i)}$  with  $d_i$ , rejecting  $H_{(i)}$  if  $T_{(i)} \geq d_i$  and continuing until for the first time  $T_{(i)} < d_i$ . If  $T_{(m)} < d_m$ , reject no hypotheses. If  $T_{(i)} \geq d_i$ , for all values of  $i$ , reject all the hypotheses. The values  $d_1 \leq d_2 \leq \dots \leq d_m$  are constants (called critical values).

For step-up procedures  $T_{(i)}$  is compared with  $d_i$ , beginning with  $i = 1$ , continuing for  $i = 2, 3, \dots$  until, for the first time,  $T_{(i)} \geq d_i$ .  $H_{(i)}, H_{(i+1)}, \dots, H_{(m)}$  are then rejected.

The procedures may also be formulated in terms of p-values and corresponding critical p-values.

### 3. Least favorable configurations for the FDR step-down and step-up procedures of Somerville (2004)

Although the formulas for calculating the critical values for the step-down procedures of Troendle (2000) and Somerville (2004) are the same, the formulas for calculating the critical values for the step-down procedure of Somerville were derived using  $m$  assumed least favorable configurations (LFCs) of the expected means of the test statistics.  $LFC_i$  ( $i = 1, 2, \dots, m$ ) were defined to be the configuration of the vector of the  $m$  expected means of the test statistics where  $i$  of the expected means are zero and the remainder increase without limit. A rigorous proof of the least favorable configuration, although strongly supported by Monte Carlo results, and calculations, has not been achieved.

### 4. Calculation of the critical values

The critical values are calculated sequentially from  $d_1$  to  $d_m$ .  $d_1$  is calculated as the smallest value such that  $FDR \leq \alpha$  under the configuration  $LFC_1$ .  $d_i$  is calculated as the smallest value such that  $FDR \leq \alpha$ , given the values  $d_1, \dots, d_{i-1}$ , and the configuration  $LFC_i$  and subject to  $d_1 \leq d_2 \leq \dots \leq d_i$ . The calculation procedure is described in Somerville (2004c). Somerville has developed the Fortran 95 program SEQ to calculate the critical values for his step-down and step-up procedures, and the program has been submitted for publication.

### 5. Reduced Step Procedures

An  $s$  step step-down procedure is defined as follows: Compare  $T_{(i)}$  with  $d_i$  for  $i=m, m-1, \dots, m-s+1$  until for the first time  $T_{(i)} < d_i$ , in which case reject  $H_{(m)}, \dots, H_{(i+1)}$ . If  $T_{(m)} < d_m$ , reject no hypotheses. If  $i = m-s+1$ , reject all hypotheses for which  $T_i \geq d_i$ .

For the  $s$  step, step-up procedure, begin with the comparison of  $T_{(m-s+1)}$  with  $d_{m-s+1}$ . If  $T_{(m-s+1)} \geq d_{m-s+1}$ , reject all hypotheses  $H_i$  for which  $T_i \geq d_{m-s+1}$ , and otherwise compare  $T_{(m-s+2)}$  with  $d_{m-s+2}, \dots$ , until for the first time  $T_{(i)} \geq d_i$ .  $H_{(i)}, H_{(i+1)}, \dots, H_{(m)}$  are then rejected.

An  $s$  step procedure is equivalent to an  $m$  step procedure where the  $m-s$  smallest critical values are replaced with the value  $d_{m-s+1}$ . Thus, only the largest  $s$  critical values are used in the comparisons. The critical value  $d_{m-s+1}$  is the Minimum Critical Value (MCV). The concept of an MCV was introduced by Somerville (2004).

Figure 5.1 illustrates the effect of step reduction in a step-down procedure to control  $k$ -FWER.

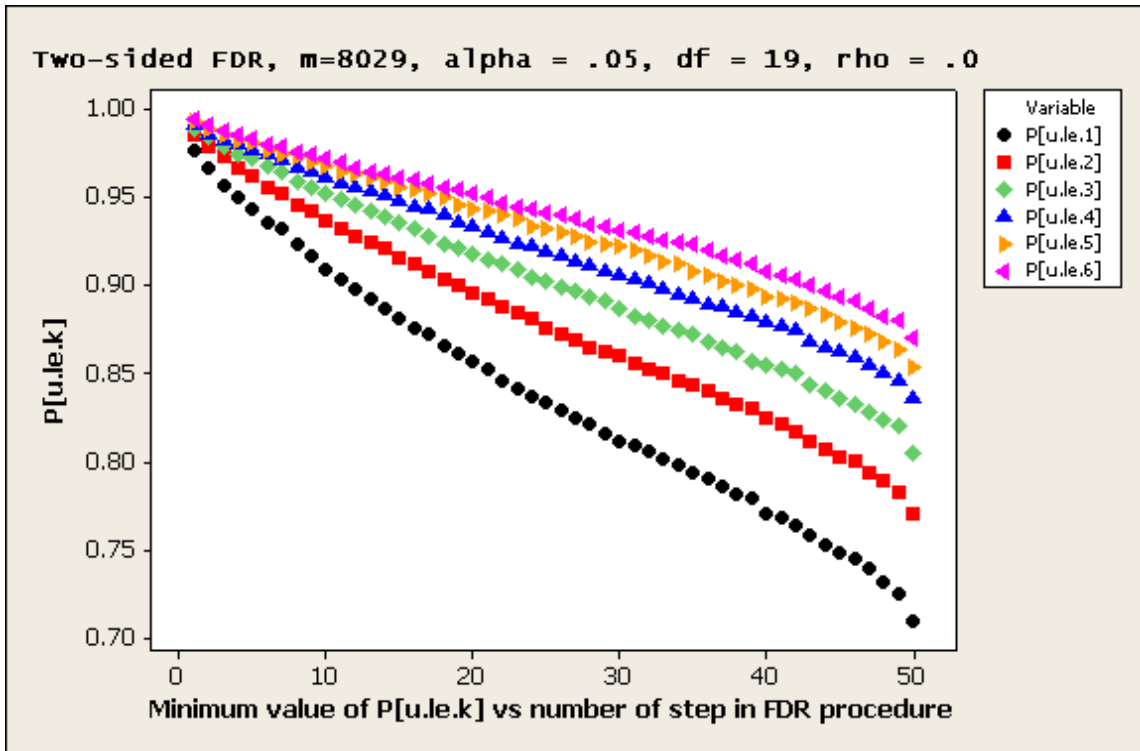


Figure 5.1

## 6. Controlling the Number of False Discoveries

Given an FDR procedure, reducing the number of steps reduces the power (reduces rejection probabilities) while at the same time reduces the number of false discoveries. We use the tables of Somerville (2006), and the Fortran program FDRPWR to obtain the maximum number of steps  $s$ , which will guarantee that  $P[u \leq k] \geq 1 - \alpha$  for various values of  $k$ ,  $m$ ,  $\rho$  and  $v$ . Approximate formulas are obtained for  $s$  as a function of  $k$  for  $v = 15$  and  $\infty$ . Additional tables are produced and given in the APPENDIX. Interpolation and extrapolation may be used for the critical values in the tables in Somerville (2006). Each critical value  $d_i$  is approximately linear in  $\log(m)$ , and in  $1/v$ . We begin with the following theorem.

**Theorem 6.1** Suppose null hypotheses  $H_1, H_2, \dots, H_m$  are simultaneously tested, using either a step-down or step-up test procedure. Let the random variable  $T_i$ , associated with the test statistic  $t_i$  have mean  $\mu_i$ , and a distribution such that

$$P[T_i \geq a \mid \mu_i = \mu] = P[T_i \geq (a + \delta) \mid \mu_i = (\mu + \delta)] \quad \text{where } a \text{ and } \delta \text{ are arbitrary constants.}$$

Suppose  $T_1, T_2, \dots, T_m$  have means  $\mu_1, \mu_2, \dots, \mu_{m1}, 0, \dots, 0$ , respectively (there are  $m1$  false hypotheses). If  $u$  is the number of true hypotheses which are rejected, then, for the 1-sided case,  $P[U \leq k]$  is minimized for  $k \in [0, m-m1]$ , when  $\mu_i \rightarrow \infty$ , for  $i = 1, \dots, m1$ .

**Proof:** Obtain a random realization of the  $m$  random variables  $T_1, T_2, \dots, T_m$  by generating a random  $m$ -dimensional vector (mean zero) and adding to it the vector  $(\mu_1, \mu_2, \dots, \mu_{m1}, 0, \dots, 0)$ .

**Step-down case:** In a step-down testing procedure, exactly  $r$  hypotheses are rejected if and only if the following holds:  $T_{(m)} \geq d_m, T_{(m-1)} \geq d_{m-1}, \dots, T_{(m-r+1)} \geq d_{m-r+1}, T_{(m-r)} < d_{m-r}$  for  $0 < r < m$ . (If  $T_{(m)} < d_m, r=0$ , and if  $T_{(i)} \geq d_i$  for all  $i, (i=1, \dots, m), r=m$ .) Suppose exactly  $r$  hypotheses are rejected. Let  $T_i'$  (with mean  $\mu_i'$ ) be one of the random variables corresponding to a false hypothesis.

Case i)  $T_i'$  corresponds to a rejected hypothesis. Then  $T_i' \geq d_{m-r}$ , and a larger value of  $\mu_i'$  would not change the rejection status of any hypothesis.

Case ii)  $T_i'$  does not correspond to a rejected hypotheses. Then  $T_i' < d_{m-r}$ . An increase in  $\mu_i'$  by an amount  $d_{m-r} - T_i'$  such that  $T_i' = d_{m-r}$  results in at least one additional rejection (an additional false hypotheses). Increase  $\mu_i'$  by  $(d_{m-r} - T_i')$ . Then  $T_i' = d_{m-r}$ . Re-label the  $m-r$  smallest  $T_i$ , and the corresponding hypotheses. Compare  $T_{(j)}$  with  $d_j$  for  $j = m-r, m-r-1, \dots$ , until for the first time  $T_{(j)} < d_j$ . If  $j$  then is less than  $m-r$ , not only the false hypothesis  $H_{(m-r)}$  will be rejected, but additional hypotheses  $H_{(m-r-1)}, \dots, H_{(j+1)}$ , will be rejected, some of which may be true hypotheses.

Thus for the step-down case, and any realization of the random variables corresponding to the test statistics, an increase in the mean of a random variable corresponding to a false hypothesis may result in an additional rejection of a true hypotheses, but can never result in a decrease. We may conclude that for the step-down case,  $P[u \leq k]$  ( $k = 0, 1, \dots, m-1$ ) is minimized when the mean of the random variables corresponding to the false hypotheses increases without limit. QED.

**Step-up case:** In a step-up testing procedure, exactly  $r$  hypotheses are rejected if and only if the following hold:  $T_{(1)} < d_1, T_{(2)} < d_2, \dots, T_{(m-r)} < d_{m-r}, T_{(m-r+1)} \geq d_{m-r+1}$  for  $0 < r < m$ . (If  $T_{(1)} \geq d_1$ , all hypotheses are rejected, and if  $T_{(i)} < d_i$  for all  $i$ , no hypotheses are rejected.

The remainder of the proof for the step-up case is the same as that for the step-down case, if the two sentences beginning with “Compare  $T_{(j)}$  with  $d_j$  for  $j = \dots$ ”, are replaced with “Compare  $T_{(j)}$  with  $d_j$  for  $j = 1, 2, \dots, m-r$ , until for the first time  $T_{(j)} \geq d_j$ . If  $j$  is less than  $m-r$ , not only the false hypothesis  $H_{(m-r)}$  will be rejected, but additional hypotheses  $H_{(j)}, \dots, H_{(m-r-1)}$ , will be rejected, some of which may be true hypotheses.”

For the 2-sided case, we may replace the test statistic  $t_i$  with the absolute values of  $t_i$  and the theorem follows.

It may be noted that Theorem 6.1 is true when the test statistics have the multivariate-t distribution.

## 7. Maximizing Power While Controlling the Number of False Discoveries

It is clear that increasing the number of steps in a step-down or step-up procedure increases the power. We need to determine the maximum number of steps which still controls the number of false discoveries  $k$ .

We use the Fortran 95 program FDRPWR. Given an arbitrary set of critical values, the program assumes the test statistics have a multivariate-t distribution with an arbitrary common correlation coefficient. For  $m$ , number of degrees of freedom, number of false hypotheses, common mean delta for the false hypotheses, the program determines three kinds of power, the probability of rejection of  $i$  true (and false) hypotheses, and the probability that  $k$  or fewer true (and false) hypotheses are rejected. The determinations are made simultaneously for all possible values of  $i$  and  $k$ . The program uses Monte Carlo, and the user can decide which outputs he wishes to examine.

For our use of the program to determine the largest number of steps  $s$  to guarantee  $P[u \leq k] \geq 1 - \alpha$ , we make use of the theorem above and use a large value for delta (assuring with high probability, the rejection of all false hypotheses). For a fixed number of steps  $s$ , the values of the smallest  $m-s$  critical values are replaced with the value  $d_{m-s+1}$  for the program input. For a specified value of  $k$ , the program is then used to determine the number of false hypotheses for which  $P[u \leq k]$  is smallest. If this smallest probability is greater than our requirement, we increment  $s$  and repeat until we obtain the largest  $s$  which still meets our requirements.

In the determination above of the “worst possible case” for the number of false hypotheses, it has been observed that, except for small values of  $m$  (e.g. 50),  $P[u \leq k]$  is a concave upward function of the number of false hypotheses, and this can be used to reduce the number of cases examined. For large values for  $k$ , small differences in the number of false hypotheses produce very small differences in  $P[u \leq k]$  near its smallest value.

Section C of Somerville (2006) gives, for  $\alpha = .05$ , the 200 largest critical values for  $m = 1000$ , for the 1- and 2-sided cases, for degrees of freedom  $v = 15$  and  $v = \infty$ , and for  $\rho = .0$ . Section D gives the 300 largest

critical values for  $m = 10000$  for the same cases. FDRPWR was used to obtain the maximum values of  $s$  such that  $P[u \leq k] \geq .95$  for  $k = 1, 2, \dots$  for the four cases for each of  $m = 1000$  and  $m = 10000$ . Using FDRPWR as described above, the maximum value of  $s$  for which the FDR requirement of  $.05$  was met was determined for  $k = 1, 2, \dots$ .

Plots of the maximum value of  $s$  vs.  $k$  suggested little difference between  $m = 1000$  and  $m = 10000$  and between the 1- and 2-sided cases. There were however large differences between  $v = 15$  and  $\infty$ .

Combining the data for the 1- and 2-side cases, and for  $m = 1000$  and  $m = 10000$ , quadratic regressions of the maximum value of  $s$  versus  $k$  were obtained for  $v = 15$  and  $v = \infty$ . (with R-squared values of  $.995$  and  $.998$  respectively). Data for the regression calculations used only the cases where the "maximum steps" was less than 200. The regression equations are given below.

$$v = 15: \text{ Max Steps} = -0.24923 + 3.85682*k - .000425084*k^2 \quad (\text{use for } k \leq 33).$$

$$v = \infty: \text{ Max Steps} = -5.0138 + 10.1836*k + 0.3642844*k^2 \quad (\text{use for } k \leq 19).$$

Using the above equations, the program FDRPWR is not needed to obtain the (approximate) maximum number of steps when  $v = 15$  or  $\infty$ , and  $1000 \leq m \leq 10000$ . Previous work by Somerville (2004) has suggested linear interpolation in  $\log(m)$  for critical values of the FDR multiple comparison procedure, and thus the equations should be appropriate for an order of magnitude larger  $m$  values. For values of  $v$  other than 15 or  $\infty$ , linear interpolation in  $1/v$  is suggested.

FDRPWR was used in the same manner to develop Tables A1 and A2 of the Appendix. The tables give the maximum value of  $s$  ( the number of steps in the step-down FDR procedure) such that  $P[u \leq k] \geq .95$ . Table A1, for the 1-sided case, gives the maximum value of  $s$  for  $k \leq 7$ , common correlation coefficient  $\rho$  of  $.0, .1$  and  $.5$ ,  $v = 15$  and  $\infty$ , and  $m$  values of 50, 100, 250, 500, 1000, 2500, 5000 and 10000. Table A2, for the 2-sided case, gives the maximum value of  $s$  for the same values of  $k, v$ , and  $m$ , but only for  $\rho$  of  $.0$ , and  $.1$ . The tabled values may be in error by a step, and for large  $k$  for more than a step. However, the resulting values of  $P[u \leq k]$  should be close to  $.95$ .

## 8. Controlling the Proportion of False Positives (PFP)

Theorem 2 in van der Laan, Dudoit and Pollard (2004) and extensions in Dudoit, van der Laan and Birkner (2004) control PFP, the proportion of false positives. We augment the set of rejected hypotheses with  $r_1$  additional hypotheses where  $r_1$  is the maximum value of  $i$  such that  $(k+i)/(r_0+i) \leq \gamma$ . The additional hypotheses should be those corresponding to the smallest p-values (or largest test statistic values) which have not already been rejected, and we may then state that

$$P[\text{PFP} \leq \gamma] \geq 1 - \alpha.$$

## 9. Examples

### Example 1

Korn et al (2004) used data from Perou et al (2000) to illustrate their Procedures A and B. Procedure A controls  $u$ , the number of false discoveries so that the  $P[u \leq k] \geq 1 - \alpha$ . Procedure B controls the proportion of false discoveries with confidence approximately  $1 - \alpha$ . The microarray dataset consisted of measurements on approximately 9000 genes in paired tumor specimens, collected both before and after chemotherapy on 20 breast cancer patients. Korn et al, after elimination of cases with missing data, included 8029 genes for analysis. The null hypothesis was that chemotherapy did not affect gene expression.

Using the Fortran 95 program SEQ, the critical values for the FDR two-sided step-down procedure were obtained using  $m = 8029$ ,  $\alpha = .05$ , degrees of freedom 19,  $\rho = .0$  and  $.1$ , and 51 "unique" critical values. Table 9.1 gives the critical values using  $\rho = .0$ , and the absolute values of the 38 largest paired t-test values. The number of genes identified is given for number of steps from 1 to 38. The table shows that a maximum of 36 genes may be identified using the FDR step-down procedure of Somerville (2004), assuming the test statistics have a multivariate-t distribution with common correlation zero.

Table 9.2 compares the methodology of this paper and that of Lehmann and Romano (2005) with respect to the maximum number of hypotheses rejections which will guarantee that  $k\text{-FWER} \leq \alpha$ . The number in parentheses is the maximum number of steps  $s$  for the procedure of this paper. In the line prefixed by \*, FDRWPR was used in the determination of the number of steps to be used. In the line prefixed by \*\*, the regression equation was used, with interpolation used to obtain the critical value between the steps.

From Table 9.2, we see that with 9 steps we can reject 24 hypotheses while controlling 2-FWER, and still control FDR. The augmentation procedures van der Laan et al (2004) and Dudoit et al (2004) are not useful in demonstrating that  $P[\text{PFP} \leq .10] \geq .95$  when we reject 24 hypotheses. The procedure of Theorem 3.1 of Lehmann and Romano (2005) allow 17 rejections, while Theorem 3.3 allows 3 for  $P[\text{PFP} \leq .10] \geq .95$ . The first theorem holds “under mild conditions on the dependence structure of p-values” while the second holds “without any dependence assumptions.” The results of this paper assume that the test statistics have a multivariate-t distribution with zero correlation.

Although not shown here, a similar analysis, assuming the test statistics have a common correlation of .1 yielded essentially the same result.

It is worth noting that Procedure A of Korn et al (2004) identified 17, 23 and 28 genes, controlling 0-FWER,1-FWER and 2-FWER at level .05. Their procedure identified 28 genes with  $P[\text{PFP} \leq 28] \geq .95$  (approximately).

Steps	t-values	CV's	Identified	Steps	t-values	CV's	Identified
1	7.905	5.552	18	20	5.404	4.612	30
2	7.423	5.370	20	21	5.140	4.594	31
3	7.329	5.251	20	22	5.126	4.576	31
4	7.021	5.159	20	23	5.085	4.559	31
5	7.021	5.099	22	24	4.910	4.543	32
6	6.977	5.028	23	25	4.876	4.528	33
7	6.699	4.987	23	26	4.835	4.513	33
8	6.576	4.934	23	27	4.813	4.499	33
9	6.259	4.893	24	28	4.778	4.485	33
10	6.215	4.857	25	29	4.712	4.472	34
11	6.215	4.826	26	30	4.625	4.459	34
12	6.139	4.798	27	31	4.605	4.447	34
13	6.139	4.770	28	32	4.544	4.435	35
14	6.073	4.744	28	33	4.542	4.423	35
15	5.965	4.719	28	34	4.481	4.412	35
16	5.905	4.696	29	35	4.439	4.400	35
17	5.868	4.673	29	36	4.393	4.389	36
18	5.581	4.652	29	<b>37</b>	<b>4.334</b>	<b>4.377</b>	<b>36</b>
19	5.451	4.631	29	<b>38</b>	<b>4.333</b>	<b>4.366</b>	<b>36</b>

Table 9.1

38 Largest t-values and Critical Values and Number of Genes Identified  
Data from Perou et al

<b>k</b>	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>
L-R		11	17	17	17	19	20	20	20	20	22
*Paper	18(1)	20(4)	24(9)	27(12)	29(17)	31(22)	33(27)	34(29)	35(32)	35(34)	36(39)
**Paper	18(1)	20(3.8)	23(7.7)	27(11.5)	29(15.4)	30(19.2)	31(23.1)	33(26.9)	35(30.7)	35(34.6)	36(38.5)

Table 9.2

Maximum Number of Hypotheses Which Can be Rejected in Controlling k-FWER ( $\alpha = .05$ )  
Comparison of Methods of Lehmann and Romano and This Paper  
Number of Steps (this paper) Used in Parentheses

\* FDRPWR used in determination of maximum number of steps

\*\* Regression formula used for determination of number of steps

### Example 2

Hedenfalk et al (2001) analysed gene expression data from 7 carriers of the BRCA1 mutation and 8 carriers of the BRCA2 mutation in an hereditary breast cancer study. The null hypothesis was that gene expressions were the same for the two mutations. There were 3226 genes that met the criteria for inclusion in the analysis.

Transformed data ( $\log_2$ ) was used in our analysis. Means and standard deviations for the 3226 genes were calculated for the two mutations. Histograms of the standard deviations for each of BRCA1 and BRCA2 suggested that it was not unreasonable to assume that the standard deviations were from populations with standard deviations .6123 and .6666 respectively. Calculations of correlations between the means for the two mutations suggested very little correlation, if any. Accordingly z statistics were calculated for the mean differences for each of the 3226 genes.

The Fortran 95 program SEQ was used to determine the critical values for the step-down FDR procedure using an MCV of 2.757 (exactly 327 "unique" critical values),  $m = 3226$ ,  $v = \infty$ ,  $\alpha = .05$ ,  $\rho = 0$ , 2-sided. The value of MCV was chosen so as to produce a number of "unique" critical values which was at least as large as the anticipated number of gene differences. Using this set of critical values, 133 genes were identified (using 133 steps). Figure 9.1 compares the 140 smallest p-values with the corresponding critical p-values.

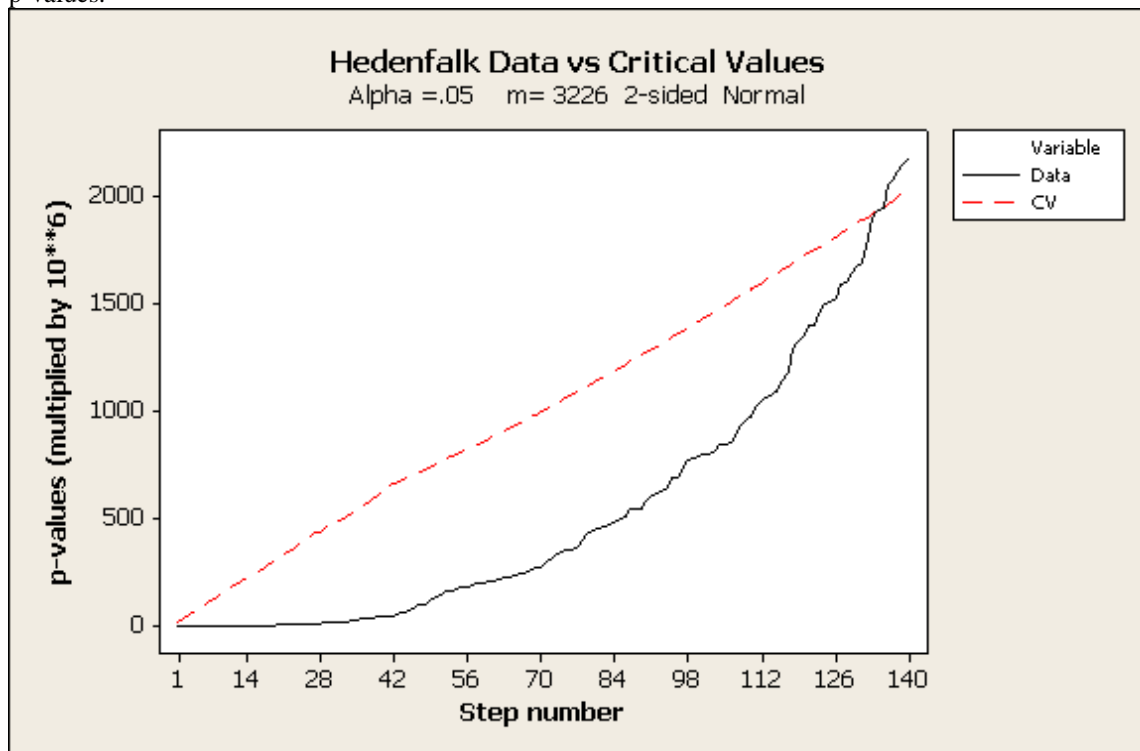


Figure 9.1

Figure 9.2 compares the methodology of this paper and that of Lehmann and Romano (2005) with respect to the maximum number of hypotheses rejections which will guarantee the  $k$ -FWER  $\leq \alpha$ . For  $k$  equal to 1 both procedures can reject 32 hypotheses. However, for the Hedenfalk data, for  $k$  greater than 1 the method of this paper can reject considerably more than that of Lehmann and Romano (2005). The methods of this paper require the assumption of a multivariate- $t$  distribution of the test statistics while Lehmann and Romano (2005) make no assumptions concerning the dependence structure or the p-values of the individual tests.

Table A2 in the Appendix could have been used to obtain the maximum number of steps used for  $k \leq 4$ , to

obtain the values for  $u$  in Figure 9.2.

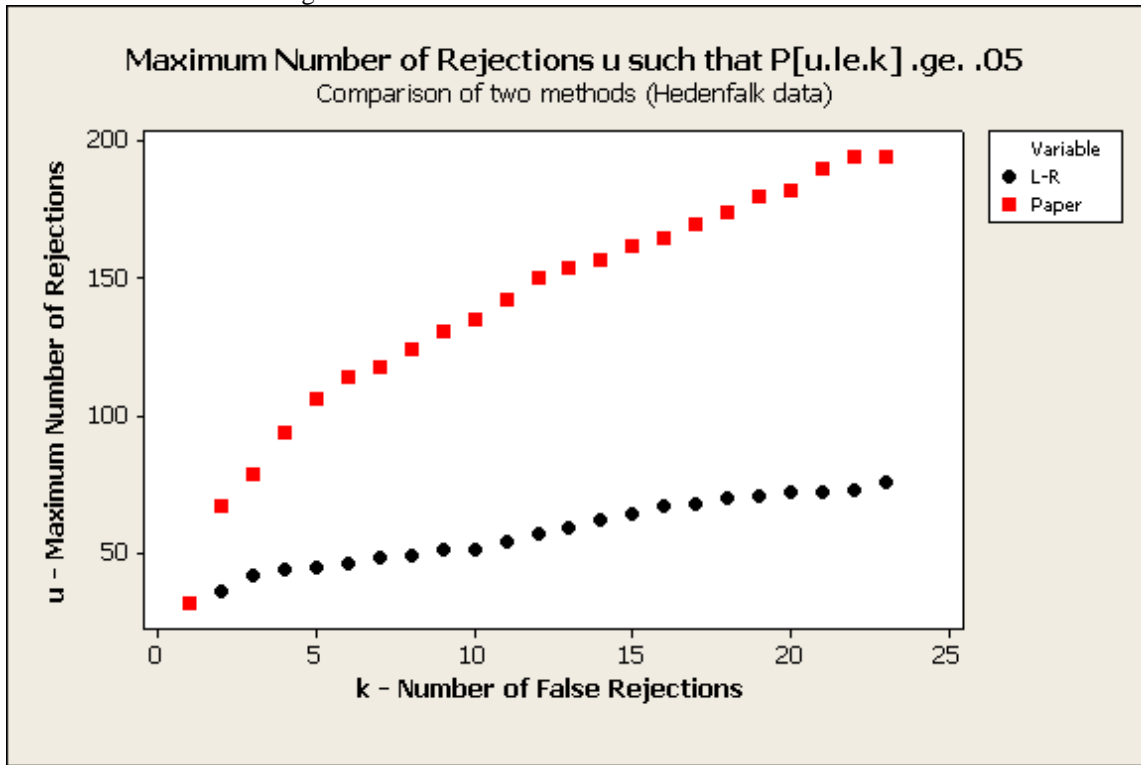


Figure 9.2

Table 9.3 gives the maximum number of steps  $s$  which can be used in the procedure of this paper and still maintain  $P[U \leq k] \geq .95$ , and the corresponding number of genes  $r_0$  for which differences in gene expression was identified.

The augmentation procedure of van der Laan et al (2004) and Dudoit et al (2004) can be used to find the maximum number of rejections  $r$  which can be made and have  $P[\text{PFP} \leq \gamma] \geq 1 - \alpha$ , where  $\gamma = .10$ , and  $\alpha = .05$ . The set of rejected hypotheses  $r_0$ , may be augmented with  $r_1$  additional hypotheses, where  $r_1$  is the maximum value of  $i$  such that  $(k+i)/(r_0 + i) \leq \gamma$ . We may ignore  $k$ , and find that 170 hypotheses may be rejected and we will still have  $P[\text{PFP} \leq .10] \geq .95$ . To simultaneous control  $\text{FDR} \leq .05$  (using the step-down procedure of Somerville (2004), we can reject only 128 hypotheses.

$k$	Max $s$	$r_0$	$r_1$	$r$	$k$	Max $s$	$r_0$	$r_1$	$r$
0	1	32	3	35	10	139	135	3	138
1	7	48	4	52	11	157	142	3	145
2	16	67	5	72	12	173	150	3	153
3	26	78	6	84	13	189	154	2	156
4	41	94	6	100	14	207	157	1	158
5	59	106	6	112	15	222	159	1	160
6	75	113	5	118	16	238	165	0	165
7	91	117	5	122	17	253	170	0	<b>170</b>
8	106	124	4	<b>128</b>	18	267	175	-	-
9	122	131	4	135	19	281	181	-	-

Table 9.3

Table Showing Maximum  $s$  for  $P[U \leq k] \geq .95$  and Method for Control of PFP

The two procedures for control of PFP of Lehmann and Romano show  $P[\text{PFP} \leq .10] \geq .95$ , using 32 and 28 hypotheses rejections respectively. The first procedure “holds under mild conditions on the dependence structure of  $p$ -values”, while the second “holds without any dependence assumptions. The procedure of

this paper assumes the  $m$  test statistics possess a multivariate normal distribution with zero correlations. The critical values need not satisfy any FDR requirement.

## 10. Summary and Conclusions

All models being approximations to reality, all models are incorrect. The validity of the usage of a model depends very much of the usefulness of the approximation in reality. Models which make no assumptions may be correct, but models which do make assumptions can be very useful in practical situations.

The models in this paper are based on the assumption that the test statistics have a multivariate-t distribution. The validity of the use of a model is a function of the ability of the model to give useful answers to real problems. The model has been shown to give superior results when the test statistics have a multivariate-t distribution. We are of the opinion that the model will give good results in many cases where the model is merely a rough approximation, especially in the many cases where the distribution of the test statistics exhibit positive dependency in some sense. Sarkar (2002) and Benjamini and Yekutieli (2001) discuss a number of such practical cases.

In this paper we show that by limiting the number of steps in step-down or step-up procedures, we can control the number of false discoveries. We maximize power while maintaining control over the number of false discoveries. Specifically, we produce formulas, and tables in the Appendix, which specify the maximum number of steps in an FDR procedure to control the probability of a specified number of false discoveries with probability .95 when the critical constants used are those obtained using the step-down procedure of Somerville (2004). We show also how to use the FDR step-down procedure to control the proportion of false positives (PFP).

The methods for controlling  $k$ -FWER and PFP developed in this paper are compared in two examples with those recently developed by Lehmann and Romano (2005).

Using the critical values for the Benjamini and Hochberg (1995) procedure, the reduced step procedure of this paper has been used with similar results. Details are given in a paper under preparation.

It may be noted that the procedure of this paper may be implemented using the critical constants of Somerville (2004) by using the following two steps:

- i) Determine the critical values (use interpolation or extrapolation using the tables in Somerville (2006), or use the Fortran program SEQ.
- ii) Determine the maximum number of steps to use (use Table A1 of the Appendix, use the regression equations in 7. or make use of the Fortran 95 program FDRPWR).

## 12. References

Benjamini Y., Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J.R.Statist. Soc. B*, 289-300.

Benjamini, Y., Liu, W. (1999). A step-down multiple hypotheses testing procedure that controls the false discovery rate under independence. *JSPI*, **82**,163-170.

Dudoit Sandrine, van der Laan, Mark J. and Birkner, Merrill D.(December 15, 2004). Multiple Testing Procedures for Controlling Tail Probability Error Rates. U.C. Berkeley Division of Biostatistics Working Paper Series. Working Paper 166. <http://www.bepress.com/ucbbiostat/paper166/>

Duncan, D.B. (1951) A significance test for differences between ranked treatments in an analysis of variance. *Virginia Journal of Science* **2**,171-189.

Dunnett, C.W. and Tamhane. A. C.(1991). Step-down multiple tests for comparing treatments with a control in unbalanced one-way layouts. *Statistics in Medicine* **10**, 939-947.

Dunnett, C.W. and Tamhane. A. C.(1992). A step-up multiple test procedure. *Journal of the American Statistical Association* **87**, 162-170.

Dunnett, C.W. and Tamhane. A. C.(1995). Step-up multiple testing of parameters with unequally correlated estimates. *Biometrics* **51**, 217-221.

Hedenfalk, Ingrid et al (2001). Gene Expression Profiles in Hereditary Breast Cancer. *The New England Journal of Medicine*, **344**, 539-548.

Horn, M., Dunnett, C. W. (2004). Power and sample size comparisons of stepwise FWE and FDR controlling test procedures in the normal many-one case. *Recent developments in multiple comparison procedures*. Institute of Mathematical Statistics, Lecture Notes - Monograph Series, **47**, 48-64.

Korn, Edward L., Troendle, James F., McShane, Lisa M. and Simon, Richard (2004). Controlling the number of false discoveries: Application to high-dimensional genomic data. *JSPI*, **124**, 379-398.

Lehmann, E. L. and Romano, Joseph P. (2005). Generalizations of the familywise error rate. *Annals of Statistics*, **33**, No. 3, 1138-1154

Perou, C. M., Sorlie, T. et al (2000). Molecular portraits of human breast tumors, *Nature* **406**, 747-752.

Sarkar, S.K. (2002). Some results on false discovery rate in stepwise multiple testing, *Annals of Statistics* **30**, 239-257.

Somerville, Paul N. (2003) Optimum FDR procedures and MCV values. Technical Report TR-03-01, Department of Statistics, University of Central Florida.

Somerville, Paul N. (2004a) A step-down procedures for large numbers of hypotheses. Paper presented at the annual meetings of the Florida Chapter of the ASA. Department of Statistics Technical Report 2004-6.)

Somerville, Paul N. (2004b) Step-down procedures for large numbers of hypotheses. *Applied Parallel Computing, State of the Art in Scientific Computing*. Jack Dongarra, K. Madsen, and J. Wasniewski (Eds.): PARA 2004, LCNS 3732, pp. 949-956, 2005. © Springer-Verlag Berlin Heidelberg 2005.

Somerville, Paul N. (2004c) FDR step-down and step-up procedures for the correlated case. *Recent Developments in Multiple Comparison Procedures*. IMS Lecture Notes Monograph Series, **47**, 100 -118.

Somerville, Paul N. (2006) Some tables of critical values for Somerville's multiple comparison procedure. Technical Report 2006-1, Department of Statistics, University of Central Florida.  
<http://pegasus.cc.ucf.edu/~somervil/2006-1.pdf>

Troendle, James F. (2000). Stepwise normal theory multiple test procedures controlling the false discovery rate, *J. Statist. Plann. Inference* **84**, 139-158.

van der Laan, Mark J., Dudoit, Sandrine and Pollard, Katherine S. (2004). Augmentation Procedures for Control of the Generalized Family-Wise Error Rate and Tail Probabilities for the Proportion of False Positives, *Statistical Applications in Genetics and Molecular Biology* **3**, 1-25.

APPENDIX

Table A1

<u>m</u>	<u>ρ</u>	<u>k</u>									
		<u>0</u>	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>		
		<u>v</u>									
		<u>∞</u>	<u>15</u>	<u>∞</u>	<u>15</u>	<u>∞</u>	<u>15</u>	<u>∞</u>	<u>15</u>	<u>∞</u>	<u>15</u>
50	.0	1 1	7 4	16 10	27 18	* 28	* *	* *	* *	* *	
	.1	1 1	5 4	12 9	19 14	17 21	* 28	* *	* *	* *	
	.5	1 1	3 3	7 6	10 9	14 12	17 16	21 19	24 23		
100	.0	1 1	7 4	16 9	27 14	* 21	* 30	* *	* *		
	.1	1 1	5 4	11 8	18 12	26 17	* 23	* 29	* *		
	.5	1 1	3 3	7 6	10 9	13 12	16 15	20 18	22 21		
250	.0	1 1	7 4	16 8	27 12	* 17	* 22	* 28	* *		
	.1	1 1	5 4	11 7	17 10	24 14	30 19	* 23	* 29		
	.5	1 1	3 3	7 6	10 9	13 12	16 15	19 18	22 21		
500	.0	1 1	7 4	16 7	27 11	* 15	* 20	* 24	* 29		
	.1	1 1	5 4	11 7	16 10	23 14	30 17	* 21	* 25		
	.5	1 1	3 3	7 6	10 9	13 12	16 15	19 17	22 20		
1000	.0	1 1	7 4	16 7	27 11	* 15	* 19	* 23	* 27		
	.1	1 1	5 4	11 7	16 10	22 14	29 17	* 21	* 24		
	.5	1 1	3 3	6 6	10 9	12 12	15 15	19 17	21 19		
2500	.0	1 1	7 4	16 7	27 10	* 14	* 17	* 20	* 23		
	.1	1 1	5 4	11 7	16 10	22 13	27 16	* 19	* 22		
	.5	1 1	3 3	6 6	9 8	12 11	15 14	18 16	20 19		
5000	.0	1 1	7 4	16 7	27 10	* 13	* 16	* 19	* 23		
	.1	1 1	5 4	10 7	15 10	20 13	26 16	30 19	31 22		
	.5	1 1	3 3	6 6	9 8	12 11	15 14	18 16	20 19		
10000	.0	1 1	7 4	16 7	27 10	* 13	* 16	* 19	* 23		
	.1	1 1	4 4	9 7	14 10	19 13	25 16	29 19	31 22		
	.5	1 1	3 3	6 6	9 8	12 11	15 13	17 15	20 18		

Max number of steps for  $P[u \leq k] \geq .95$ ,  $\alpha = .05$ ,  $v = \infty$ , one-sided case ( $v = 15$  in italics)

\* At least 31

Table A2

<u>m</u>	<u>ρ</u>	<u>k</u>									
		<u>0</u>	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>	<u>8</u>	<u>9</u>
		<u>v</u>		<u>v</u>		<u>v</u>		<u>v</u>		<u>v</u>	
		<u>∞</u>	<u>15</u>	<u>∞</u>	<u>15</u>	<u>∞</u>	<u>15</u>	<u>∞</u>	<u>15</u>	<u>∞</u>	<u>15</u>
50	.0	1 1	7 4	16 9	26 15	* 22	* 30	* *	* *	* *	* *
	.1	1 1	6 4	14 9	24 14	* 21	* 30	* *	* *	* *	* *
100	.0	1 1	7 4	16 8	26 12	* 18	* 24	* 30	* *	* *	* *
	.1	1 1	6 3	14 7	23 12	* 17	* 23	* 28	* *	* *	* *
250	.0	1 1	7 4	16 7	27 11	* 15	* 20	* 24	* 29	* 28	* 29
	.1	1 1	6 3	13 7	22 11	30 15	* 19	* 23	* 28	* 27	* 28
500	.0	1 1	7 3	16 7	27 10	* 14	* 18	* 22	* 26	* 25	* 26
	.1	1 1	6 3	13 7	22 10	29 14	* 17	* 21	* 25	* 24	* 25
1000	.0	1 1	7 3	16 7	27 10	* 13	* 17	* 20	* 23	* 22	* 23
	.1	1 1	6 3	13 7	21 10	29 13	* 17	* 20	* 23	* 22	* 23
2500	.0	1 1	7 3	16 6	27 9	* 13	* 16	* 19	* 22	* 21	* 22
	.1	1 1	6 3	13 6	20 9	28 12	* 15	* 19	* 22	* 21	* 22
5000	.0	1 1	7 3	16 6	27 9	* 12	* 15	* 18	* 21	* 20	* 21
	.1	1 1	6 3	12 6	19 9	27 12	* 14	* 18	* 21	* 20	* 21
10000	.0	1 1	7 3	16 6	26 9	* 12	* 15	* 18	* 21	* 20	* 21
	.1	1 1	6 3	12 6	19 9	26 12	31 14	* 17	* 20	* 19	* 20

Max number of steps for  $P[u \leq k] \geq .95$ ,  $\alpha = .05$ ,  $v = \infty$ , two-sided case ( $v = 15$  in italics)

\* At least 31