Conference on Applied Statistics in Agriculture 2006 - 18th Annual Conference Proceedings

A COMPARISON OF MULTIPLE TESTING METHODS: SPINOSAD AS A TREATMENT FOR LICE ON CATTLE

Zhanglin Cui

Daniel H. Mowrey

Alan G. Zimmermann

Douglas E. Hutchens

See next page for additional authors

Recommended Citation

Cui, Zhanglin; Mowrey, Daniel H.; Zimmermann, Alan G.; and Hutchens, Douglas E. (2006). "A COMPARISON OF MULTIPLE TESTING METHODS: SPINOSAD AS A TREATMENT FOR LICE ON CATTLE," *Conference on Applied Statistics in Agriculture*. https://doi.org/10.4148/2475-7772.1117

Author Information

Zhanglin Cui, Daniel H. Mowrey, Alan G. Zimmermann, and Douglas E. Hutchens

A COMPARISON OF MULTIPLE TESTING METHODS: SPINOSAD AS A TREATMENT FOR LICE ON CATTLE

Zhanglin Cui, Daniel H. Mowrey, Alan G. Zimmermann, and Douglas E. Hutchens
Eli Lilly and Company, 2001 W. Main St., Greenfield, IN 46140

ABSTRACT

A common problem in statistics is making multiple tests of hypotheses without controlling for the type I error rate. SAS has identified several different methods to adjust p-values for multiple testing. To compare the effect of these methods, an animal health dataset that deals with the treatment of cattle lice was examined. Clinical trials were conducted in Illinois and Wisconsin to evaluate the efficacy of two formulations of a new product Spinosad, two commercially available positive controls, and an untreated negative control. A baseline lice count was recorded prior to the treatment. After treatment, weekly measurements of lice counts were taken for 8 weeks. Counts of 4 lice species were recorded separately. A linear mixed model analysis was conducted for each species of lice after transforming the counts with a natural logarithm transformation. Simple contrasts between treatment groups at each week were performed. Treatment differences were also compared using 5 multiple testing methods: Bonferroni, Sidak, Holm's step-down Bonferroni, Hochberg's step-up Bonferroni, and false discovery rate. Seventy-one out of 96 simple tests showed significant differences among the treatment groups. The five multiple testing methods confirmed only 48-67 significances out of the 96 tests. Comparatively, Bonferroni and Sidak methods provided similar and the most conservative multiplicity test results, i.e. fewest significant differences. The Holm's step-down and Hochberg's step-up Bonferroni methods provided similar but less conservative results. Finally, the false discovery rate method provided the least conservative results.

Key words: multiple tests, Bonferroni, Sidak, Holm's step-down Bonferroni, Hochberg's step-up Bonferroni, false discovery rate

1. INTRODUCTION

Multiplicity is common in clinical studies in which there are multiple dose levels, multiple treatment groups, multiple endpoints, repeated measures over time, interim analyses during the course of a clinical study, stepwise methods to find an optimal analysis, or sub-population analyses. When multiplicity happens, families of tests of hypotheses cannot be avoided in clinical studies. For example, dose response contrasts, pairwise comparisons of treatment groups, pairwise comparisons with the control or best treatment group, pairwise comparisons of repeated measures with the baseline, pairwise comparisons of repeated measures with negative/positive/best treatment group are all cases in which multiple tests of hypotheses occur.

Oftentimes, multiple tests of hypotheses are conducted without controlling for the type I error rate, resulting in incorrect conclusions due to the inflated type I error. As an example,

considering an efficacy evaluation in a clinical study, multiplicity could result in declaring effectiveness which is not true. Researchers could then run the risk of a wrong decision to move a compound forward in the drug development pipeline. In a safety evaluation of a clinical study, multiplicity could result in declaring a safety issue which does not exist. In this case, researchers would run the risk of the loss of a good product because of the false safety issue.

The inflation of type I error can be measured by the familywise error rate (FWE). Assuming there are m tests of hypotheses, H_{01}, H_{02}, …, H_{0m}, in a multiplicity situation, the FWE is defined as the probability of rejecting at least one of the null hypotheses given the null hypotheses are all true:

$$\mathbf{FWE = P(reject\ at\ least\ one\ of\ H_{01}, H_{02}, \ldots, H_{0m} \mid H_{01}, H_{02}, \ldots, H_{0m}\ are\ all\ true)}$$

If the m tests are independent, the FWE = $1-0.95^m$ which is 0.64 when m=20 and each of the m hypotheses is tested at a non-adjusted type I error rate of 5%; If m tests are not all independent, $0.05<FWE< 1-0.95^m$.

It is necessary to control the FWE when a researcher wants to ensure that any claimed effects are real, reproducible, or repeatable, with the standard 95% confidence.

2. P-VALUE ADJUSTMENT METHODS FOR MULTIPLE TESTING

Several p-value adjustment methods for multiple testing are implemented in SAS procedures such as PROC MULTTEST (SAS Institute Inc, 1999). A brief introduction to the Bonferroni, Sidak, Holm's step-down Bonferroni, Hochberg's step-up Bonferroni, and false discovery rate methods is provided in this section. The following notation will be used: suppose there are m null hypotheses in a clinical study, H_{01}, H_{02}, …, $H_{0m;}$ let p_1, p_2, …, p_m be the respective p-values from the tests of these null hypotheses; let $p_{(1)}$, $p_{(2)}$,…, $p_{(m)}$ be the ordered p-values corresponding to the ordered hypotheses $H_{0(1)}$, $H_{0(2)}$, …, $H_{0(m)}$, where m = total number of tests.

2.1. BONFERRONI

$$\mathbf{p_{BONj} = mp_j}$$

where p_j is the non-adjusted raw p-value for Hypothesis H_{0j}. If the adjusted p-value exceeds 1, it is set to 1 (Westfall et al., 1999).

2.2. SIDAK

$$\mathbf{p_{SID\ j} = 1 - (1 - p_j)^m}$$

where p_j is the non-adjusted raw p-value for Hypothesis H_{0j} (Sidak, 1967).

2.3. HOLM'S STEP-DOWN BONFERRONI

Holm's step-down Bonferroni method adjusts the raw p-values in a stepwise fashion starting with the smallest raw p-value (Holm, 1979; Hochberg and Tamhane, 1987). Suppose $p_{(1)} < p_{(2)} < ,\ldots, < p_{(m)}$ are the ordered p-values then,

$$p_{HOLM(1)} = mp_{(1)}$$
$$p_{HOLM(2)} = \max\{p_{HOLM(1)}, (m-1)p_{(2)}\}$$
$$p_{HOLM(3)} = \max\{p_{HOLM(2)}, (m-2)p_{(3)}\}$$
$$\ldots\ldots$$
$$p_{HOLM(m-1)} = \max\{p_{HOLM(m-2)}, 2p_{(m-1)}\}$$
$$p_{HOLM(m)} = \max\{p_{HOLM(m-1)}, 1p_{(m)}\}$$

The adjusted p-value is set equal to its predecessor if its calculated value is less. As always, if any adjusted p-value exceeds 1, it is set to 1. Consequently, all comparisons that follow the first non-significant comparison must be non-significant.

2.4. HOCHBERG'S STEP-UP BONFERRONI

Hochberg's Step-up Bonferroni p-value adjustment is also in a stepwise fashion, but starting with the largest raw p-value (Hochberg, 1988). Suppose $p_{(1)} < p_{(2)} < ,\ldots, < p_{(m)}$ are the ordered p-values then,

$$p_{HOC(m)} = 1p_{(m)}$$
$$p_{HOC(m-1)} = \min\{p_{HOC(m)}, 2p_{(m-1)}\}$$
$$p_{HOC(m-2)} = \min\{p_{HOC(m-1)}, 3p_{(m-2)}\}$$
$$\ldots\ldots$$
$$p_{HOC(2)} = \min\{p_{HOC(3)}, (m-1)p_{(2)}\}$$
$$p_{HOC(1)} = \min\{p_{HOC(2)}, mp_{(1)}\}$$

An adjusted p-value is set equal to its predecessor if its calculated value is greater. Consequently, Each adjusted p-value must be no greater than its predecessor.

2.5. FALSE DISCOVERY RATE

Benjamini and Hochberg's false discovery rate (FDR) adjusts the raw p-values also in a stepwise fashion starting with the largest raw p-value (Benjamini and Hochberg, 1995). It is a step-up method. Suppose $p_{(1)} < p_{(2)} < ,\ldots, < p_{(m)}$ are the ordered p-values then,

$$p_{FDR(m)} = 1p_{(m)}$$
$$p_{FDR(m-1)} = \min\{p_{FDR(m)}, [m/(m-1)]p_{(m-1)}\}$$
$$p_{FDR(m-2)} = \min\{p_{FDR(m-1)}, [m/(m-2)]\, p_{(m-2)}\}$$
$$\ldots\ldots$$
$$p_{FDR(2)} = \min\{p_{FDR(3)}, [m/2]p_{(2)}\}$$

$$p_{FDR(1)} = \min\{p_{FDR(2)}, mp_{(1)}\}$$

The false discovery rate adjusts p-values that control the "false discovery rate," described by Benjamini and Hochberg (1995). These adjustments are potentially much less conservative than the Hochberg adjustments; however, they do not necessarily control the familywise error rate. Furthermore, they are guaranteed to control the false discovery rate only with independent *p*-values that are uniformly distributed under their respective null hypotheses. The false discovery rate adjusted *p*-values are defined in step-up fashion, like the Hochberg adjustments, but with less conservative multipliers.

3. A CASE STUDY: SPINOSAD AS A TREATMENT FOR LICE ON CATTLE

To compare the effect of the multiple testing methods, an animal health dataset that deals with the treatment of cattle lice was examined.

3.1. ANIMAL CLINICAL FIELD STUDY

Lice species are common pests that live on beef cattle in America. Several different lice species can be found on cattle. The Lice species bite skin tissue or suck blood from the cattle. The control of lice in the cattle population is economically important. A new product Spinosad is developed for the treatment of lice on beef cattle. Spinosad can be applied to cattle as a spray formulation or pour-on formulation (Lloyd et al., 1996).

Clinical trials were conducted in Illinois and Wisconsin in the spring of 2000 to evaluate the efficacy of two formulations of Spinosad, two commercially available positive controls, and an untreated negative control, for the treatment of lice on beef cattle under natural field conditions. There were five treatments in the study. Treatments A and B are the 400 ppm Spinosad spray formulation and 2 mg/kg Spinosad pour-on formulation, respectively. Treatments C and D are the positive controls, 5.8% Co-Ral® spray formulation and 1% Cylence™ pour-on formulation, respectively. Finally, treatment E is the untreated negative control (Campbell et al, 2001; Colwell, 2002; Holste et al., 1997).

There were seven cattle in each treatment group at each trial site. Each treatment group (herd) of cattle was maintained in separate outdoor pens at each site. Baseline lice counts on cattle were recorded prior to the treatment. After treatment, weekly lice counts on cattle were taken for eight weeks. Counts of four lice species (*Bovicola bovis, Haematopinus eurysternus, Linognathus vituli,* and *Solenopotes capillatus*) were recorded separately.

The objectives of the clinical study were to determine the efficacy of the test substance, Spinosad, applied as a diluted spray or as a neat pour-on against sucking and chewing lice species on naturally infested cattle under field conditions; and to compare the efficacy of Spinosad to commercially available positive controls, Co-Ral and Cylence, used as per the manufacturer's instructions for lice control and applied under the same field conditions. The

purpose of this paper is to compare the five multiple testing p-value adjustment methods using the study data.

3.2. STATISTICAL ANALYSIS METHODS

A natural logarithmic transformation of the lice counts ($\log_e$ (count+1)) was applied since the data were positive integers (counting numbers) which usually have the standard deviation proportional to the mean.

A linear mixed model analysis (PROC MIXED, SAS v8.2) was conducted for each species of lice after transforming the counts with a natural logarithmic transformation. Treatment, Week and Treatment*Week were treated as fixed effects and Site, Site*Treatment, and Site*Treatment*Week as random effects. After examining several covariance structures, a repeated measures analysis on each animal within site*treatment combination was modeled using a compound symmetric covariance structure. The log-transformed baseline lice counts were used as a covariate in the model. Each lice species was analyzed separately.

Simple linear contrasts between treated groups and the negative control at each week were constructed. Treatment differences were compared using 5 multiplicity test methods: Bonferroni, Sidak, Holm's step-down Bonferroni, Hochberg's step-up Bonferroni, and false discovery rate. P-values were adjusted for two different families of simple tests: one, 32 tests (data not shown) consisting of 4 pairs of treatment comparisons (A vs. E, B vs. E, C vs. E, and D vs. E) at each of the 8 treatment weeks for each lice species; and two, 96 tests consisting of 4 pairs of treatment comparisons (A vs. E, B vs. E, C vs. E, and D vs. E) at each of the 8 treatment weeks across the three major species (*B. bovis, L. vituli,* and *S. capillatus*). The data for the family one is not shown due to the fact that the lice species were parasitoids on the same cattle. The lice species were related to each other. However, the results from the two families were similar. In addition, only Dunnett type treatment comparisons were selected for the purpose of this paper. Other meaningful treatment comparisons were omitted from this paper.

4. RESULTS AND DISCUSSION

Based on the non-adjusted p-values, significant overall treatment differences averaged across weeks, pairwise treatment differences averaged across weeks, and overall treatment differences at some weeks existed for lice species *B. bovis, L. vituli,* and *S. capillatus* but not *H. eurysternus* (Tables 1 and 2). The lice species *H. eurysternus* was dropped from further analysis due to insufficient number of animals infected and lack of significance. *B. bovis*, *L. vituli*, and *S. capillatus* proceeded for further testing of pairwise treatment differences at each week.

The non-adjusted p-values for pairwise treatment differences at each week for the three major lice species are shown in Table 3. The adjusted p-values for pairwise treatment differences at each week for the three major lice species using Bonferroni, Sidak, Holm's step-down Bonferroni, Hochberg's step-up Bonferroni, and false discovery rate methods adjusted for a family of 96 tests are shown in Tables 4, 5, 6, 7, and 8, respectively.

The non-adjusted multiple testing produced 71 significant p-values (Table 3). The result indicated that Spinosad exhibited 8 weeks of effectiveness for *B. bovis,* 7-8 weeks for *L. vituli,* and 2-4 weeks for *S. capillatus.*

The adjusted p-values from Bonferroni, Sidak, Holm's step-down Bonferroni, and Hochberg's step-up Bonferroni methods revealed the same test results (Tables 4-8). Only 48 p-values remained significant after adjustment. These methods indicated that Spinosad had 5 weeks of effectiveness for *B. bovis,* 7-8 weeks of effectiveness for *L. vituli,* and no effectiveness for *S. capillatus.*

The false discovery rate method produced more significant p-values (67) than the other multiple testing methods (Table 8), indicating that Spinosad exhibited 8 weeks of effectiveness for *B. bovis,* 7-8 weeks for *L. vituli,* and 1-3 weeks for *S. capillatus.*

Although Bonferroni, Sidak, Holm's step-down Bonferroni, and Hochberg's step-up Bonferroni methods produced the same number of significant p-values, as expected the magnitude of the adjusted p-values differed, in general, in the following order from largest to smallest:

> Bonferroni > Sidak > Holm's step-down Bonferroni > Hochberg's step-up Bonferroni > False discovery rate > Non-adjusted raw p-value.

To compare the magnitude of the non-adjusted raw p-values and the adjusted p-values from the five different multiple testing methods, the non-adjusted p-values (Table 3) and their corresponding adjusted p-values (Tables 4-8) were merged to adjoining columns in a data set and sorted by the raw p-value in ascending order. The sorted p-values were plotted in a graph for comparison (Figure 1).

When the raw p-values were small (right panel of Figure 1), the magnitude of adjusted p-values from Boferroni method was close to that from Sidak method. These two methods represented the most conservative approaches (i.e. with fewest significant treatment differences). Likewise, Holm's step-down Bonferroni and Hochberg's step-up Bonferroni produced similar and moderate p-values. Whereas, the false discovery rate method produced the smallest adjusted p-values that were close to the non-adjusted raw p-values. The false discovery rate method represented the least conservative method.

When the raw p-values were large (left panel of Figure 1), the adjusted p-values from Bonferroni, Sidak, Holm's step-down Bonferroni methods quickly increased and approached the maximum p-value of 1. However, the false discovery rate adjusted p-value remained close to the raw p-value and the Hochberg's step-up Bonferroni adjusted p-values were moderate with an upper boundary at the largest raw p-value.

Although Bonferroni and Sidak methods controlled familywise error rate at 5% level, the adjusted p-value became quite conservative when the number of tests were large, resulting in a

low chance to detect a true treatment effect. The Holm's step-down Bonferroni and Hochberg's Step-up Bonferroni methods improved the chance to detect the treatment differences while controlling the familywise error rate at 5% (Westfall and Young, 1993). Since the Hochberg adjusted *p*-values were smaller than or equal to Holm's p-values, the Hochberg method appeared to be more powerful to detect a potential treatment difference. However, this apparent improved power came at the cost of having to make the assumption of independence (Hochberg, 1988). It appeared that the false discovery rate was the "most powerful" of the five adjustment methods to detect potential treatment differences and thought to be very useful for screening large numbers of tests. However, it did not control the familywise error rate and faced the greatest risk of a type I error.

5. SUMMARY

Based on the data shown in Tables 4-7, Bonferroni, Sidak, step-down Bonferroni and step-up Bonferroni methods produced the same number of significant p-values. However, the magnitude of the adjusted p-values differed. Comparatively, among the five multiple testing methods, Bonferroni and Sidak provided the most consistent and conservative multiple testing results, i.e. fewest significant differences. The step-down Bonferroni and step-up Bonferroni methods provided similar but less conservative results. Finally, the false discovery rate method provided the least conservative results.

Controlling for the familywise error rate at 5% and having the improved power to detect the treatment differences, Hochberg's Step-up Bonferroni p-value adjustment showed that (1) Spinosad spray and pour-on treatments reduced *B. bovis* counts significantly for the first 5 weeks compared with the untreated control group ($P \leq 0.0163$); (2) Spinosad spray and pour-on treatments reduced *L. vituli* counts significantly for at least 7 treatment weeks compared with the untreated control group ($P \leq 0.0142$); (3) Spinosad spray and pour-on treatments did not reduce *S. capillatus* counts significantly in this study.

6. REFERENCES

Benjamini, Y. and Hochberg, Y. 1995. Controlling the false discovery rate: a new and powerful approach to multiple testing. Journal of the Royal Statistical Society, Series B 57:1289-1300.

Campbell, J.B., et al. 2001. Comparative efficacy of several insecticides for control of cattle lice (Mallophaga: Trichodectidae and Anoplur: Heamatopinidae). Veterinary Parasitology 96: 155-164.

Colwell, D.D. 2002. Persistent activity of moxidectin pour-on and injectable against sucking and biting louse infestations of cattle. Veterinary Parasitology 104: 319-326.

Holste, J.E., et al. 1997. Eprinomectin: a noval avermectin for control of lice in all classes of cattle. Veterinary Parasitology 73: 153-161.

Hochberg, Y. 1988. A sharper Bonferroni procedure for multiple tests of significance. Biometrica 75:800-802.

Hochberg, Y. and Tamhane, A.C. 1987. Multiple Comparison Procedures. New York: John Wiley & Sons.

Holm, S. 1979. A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics 6:65-70.

Lloyd, J.E., et al. 1996. Doramectin systemic activity against cattle grubs, Hypoderma lineatum and H. bovis (Diptera: Oestridae), and cattle lice, B. bovis (Mallophaga: Trichodectidae), L. vituli and S. capillatus (Anoplura: Linognathidae), and Heamatopinus eurysternus (Anoplur: Heamatopinidae), in Wyoming. Veterinary Parasitology 63:307-317.

SAS Institute Inc. 1999. SAS/STAT User's Guide, Version 8, Volume 2. Cary, NC: SAS Institute Inc.

Sidak, Z. 1967. Rectangular confidence regions for the means of multivariate normal distribution. Journal of the American Statistical Association 62:626-633.

Westfall, P.H., et al. 1999. Multiple comparisons and multiple tests. SAS Institute Inc., Cary, NC, USA.

Westfall, P.H. and Young, S.S. 1993. Resampling-Based Multiple Testing: Examples and Methods for P-value Adjustment. New York: John Wiley & Sons.

Table 1. P-values for Overall Treatment Differences and Pairwise Treatment Comparison

Lice Species	Overall Treatment	Pairwise Treatment Comparison [1]			
		A vs E	B vs E	C vs E	D vs E
Bovicola bovis	<.0001	<.0001	<.0001	<.0001	<.0001
Haematopinus eurysternus	0.5873	0.2154	0.2746	0.1311	0.2167
Linognathus vituli	<.0001	<.0001	<.0001	<.0001	<.0001
Solenopotes capillatus	<.0001	<.0001	0.0001	<.0001	0.0001

[1] A = Spinosad 400 ppm spray, B = Spinosad 2 mg/kg pour-on, C = CoRal 5.8% spray, D = Cylence 1% pour-on, E = Untreated control.

Table 2. P-value of Testing Simple Treatment Effects at Each Week

Lice Species	Treatment Week							
	1	2	3	4	5	6	7	8
Bovicola bovis	<.0001	<.0001	<.0001	0.0002	0.0006	0.0208	0.0222	0.1036
Haematopinus eurysternus	0.1133	0.4687	0.8966	0.9774	0.9870	0.8333	0.9991	1.0000
Linognathus vituli	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
Solenopotes capillatus	0.0345	0.0491	0.0677	0.0972	0.1493	0.1406	0.1430	0.1598

Table 3. P-values for Pairwise Treatment Differences at Each Week for the Three Major Lice Species.
Non-Adjusted Raw P-value (71 out of 96)

Lice Species	Treatment Week							
Treatment [1]	1	2	3	4	5	6	7	8
Bovicola bovis								
A vs E	<.0001	0.0001	<.0001	0.0001	0.0002	0.0057	0.0062	0.0179
B vs E	<.0001	0.0001	<.0001	0.0001	0.0003	0.0074	0.0058	0.0272
C vs E	<.0001	0.0001	0.0001	0.0016	0.0168	0.1297	0.0772	0.0949
D vs E	<.0001	0.0001	<.0001	0.0001	0.0004	0.0061	0.0058	0.0245
Linognathus vituli								
A vs E	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0001
B vs E	<.0001	<.0001	<.0001	<.0001	<.0001	0.0003	0.0001	0.4233
C vs E	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0001	0.0041
D vs E	0.0005	0.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0004
Solenopotes capillatus								
A vs E	0.0167	0.0221	0.0266	0.0365	0.0538	0.0563	0.0493	0.0533
B vs E	0.0306	0.0416	0.0891	0.1149	0.1319	0.1299	0.1180	0.1344
C vs E	0.0216	0.0285	0.0342	0.0464	0.0676	0.0706	0.0996	0.0999
D vs E	0.0889	0.0773	0.0646	0.0801	0.1177	0.0811	0.0747	0.0880

[1] A = Spinosad 400 ppm spray, B = Spinosad 2 mg/kg pour-on, C = CoRal 5.8% spray, D = Cylence 1% pour-on, E = Untreated control.

Table 4. P-values for Pairwise Treatment Differences at Each Week for the Three Major Lice Species. Bonferroni (48 out of 96)

Lice Species	Treatment Week							
Treatment [1]	1	2	3	4	5	6	7	8
Bovicola bovis								
A vs E	0.0013	0.0082	0.0023	0.0107	0.0199	0.5435	0.5920	1.0000
B vs E	0.0012	0.0088	0.0024	0.0114	0.0301	0.7131	0.5525	1.0000
C vs E	0.0010	0.0078	0.0060	0.1521	1.0000	1.0000	1.0000	1.0000
D vs E	0.0012	0.0089	0.0025	0.0116	0.0392	0.5861	0.5593	1.0000
Linognathus vituli								
A vs E	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0007	0.0135
B vs E	<.0001	<.0001	0.0008	0.0017	0.0014	0.0256	0.0114	1.0000
C vs E	<.0001	<.0001	<.0001	0.0005	0.0012	0.0047	0.0085	0.3951
D vs E	0.0460	0.0080	0.0004	0.0006	0.0020	0.0003	0.0002	0.0425
Solenopotes capillatus								
A vs E	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
B vs E	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
C vs E	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
D vs E	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

[1] A = Spinosad 400 ppm spray, B = Spinosad 2 mg/kg pour-on, C = CoRal 5.8% spray, D = Cylence 1% pour-on, E = Untreated control.

Table 5. P-values for Pairwise Treatment Differences at Each Week for the Three Major Lice Species. Sidak (48 out of 96)

Lice Species	Treatment Week							
Treatment [1]	1	2	3	4	5	6	7	8
Bovicola bovis								
A vs E	0.0013	0.0081	0.0023	0.0106	0.0197	0.4202	0.4478	0.8242
B vs E	0.0012	0.0087	0.0024	0.0114	0.0297	0.5112	0.4254	0.9295
C vs E	0.0010	0.0078	0.0060	0.1412	0.8040	1.0000	0.9996	0.9999
D vs E	0.0012	0.0088	0.0025	0.0115	0.0384	0.4445	0.4293	0.9073
Linognathus vituli								
A vs E	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0007	0.0134
B vs E	<.0001	<.0001	0.0008	0.0017	0.0014	0.0253	0.0113	1.0000
C vs E	<.0001	<.0001	<.0001	0.0005	0.0012	0.0047	0.0084	0.3269
D vs E	0.0450	0.0080	0.0004	0.0006	0.0020	0.0003	0.0002	0.0417
Solenopotes capillatus								
A vs E	0.8006	0.8829	0.9251	0.9719	0.9950	0.9962	0.9922	0.9948
B vs E	0.9492	0.9831	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000
C vs E	0.8771	0.9374	0.9644	0.9896	0.9988	0.9991	1.0000	1.0000
D vs E	0.9999	0.9996	0.9984	0.9997	1.0000	0.9997	0.9994	0.9999

[1] A = Spinosad 400 ppm spray, B = Spinosad 2 mg/kg pour-on, C = CoRal 5.8% spray, D = Cylence 1% pour-on, E = Untreated control.

Table 6. P-values for Pairwise Treatment Differences at Each Week for the Three Major Lice Species. Holm's step-down Bonferroni (48 out of 96)

Lice Species	Treatment Week							
Treatment [1]	1	2	3	4	5	6	7	8
Bovicola bovis								
A vs E	0.0010	0.0054	0.0017	0.0066	0.0112	0.2604	0.2625	0.6819
B vs E	0.0009	0.0056	0.0018	0.0069	0.0163	0.3045	0.2604	0.9057
C vs E	0.0008	0.0053	0.0041	0.0761	0.6662	1.0000	1.0000	1.0000
D vs E	0.0009	0.0056	0.0018	0.0069	0.0208	0.2625	0.2604	0.8567
Linognathus vituli								
A vs E	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0006	0.0078
B vs E	<.0001	<.0001	0.0007	0.0013	0.0010	0.0142	0.0069	1.0000
C vs E	<.0001	<.0001	<.0001	0.0004	0.0009	0.0033	0.0055	0.1934
D vs E	0.0235	0.0053	0.0004	0.0005	0.0015	0.0002	0.0002	0.0222
Solenopotes capillatus								
A vs E	0.6662	0.7993	0.9057	1.0000	1.0000	1.0000	1.0000	1.0000
B vs E	0.9477	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
C vs E	0.7993	0.9107	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
D vs E	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

[1] A = Spinosad 400 ppm spray, B = Spinosad 2 mg/kg pour-on, C = CoRal 5.8% spray, D = Cylence 1% pour-on, E = Untreated control.

Table 7. P-values for Pairwise Treatment Differences at Each Week for the Three Major Lice Species. Hochberg's step-up Bonferroni (48 out of 96)

Lice Species	Treatment Week							
Treatment [1]	1	2	3	4	5	6	7	8
Bovicola bovis								
A vs E	0.0010	0.0054	0.0017	0.0066	0.0112	0.2563	0.2590	0.2687
B vs E	0.0009	0.0055	0.0018	0.0068	0.0163	0.2687	0.2563	0.2687
C vs E	0.0008	0.0053	0.0041	0.0761	0.2687	0.2687	0.2687	0.2687
D vs E	0.0009	0.0055	0.0018	0.0068	0.0208	0.2590	0.2563	0.2687
Linognathus vituli								
A vs E	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0006	0.0078
B vs E	<.0001	<.0001	0.0007	0.0013	0.0010	0.0142	0.0068	0.4233
C vs E	<.0001	<.0001	<.0001	0.0004	0.0009	0.0033	0.0055	0.1934
D vs E	0.0235	0.0053	0.0004	0.0005	0.0015	0.0002	0.0002	0.0222
Solenopotes capillatus								
A vs E	0.2687	0.2687	0.2687	0.2687	0.2687	0.2687	0.2687	0.2687
B vs E	0.2687	0.2687	0.2687	0.2687	0.2687	0.2687	0.2687	0.2687
C vs E	0.2687	0.2687	0.2687	0.2687	0.2687	0.2687	0.2687	0.2687
D vs E	0.2687	0.2687	0.2687	0.2687	0.2687	0.2687	0.2687	0.2687

[1] A = Spinosad 400 ppm spray, B = Spinosad 2 mg/kg pour-on, C = CoRal 5.8% spray, D = Cylence 1% pour-on, E = Untreated control.

Table 8. P-values for Pairwise Treatment Differences at Each Week for the Three Major Lice Species. False discovery rate (FDR) (67 out of 96)

Lice Species	Treatment Week							
Treatment [1]	1	2	3	4	5	6	7	8
Bovicola bovis								
A vs E	0.0001	0.0002	0.0001	0.0003	0.0005	0.0106	0.0108	0.0292
B vs E	0.0001	0.0002	0.0001	0.0003	0.0007	0.0127	0.0106	0.0409
C vs E	0.0001	0.0002	0.0002	0.0031	0.0279	0.1340	0.0927	0.1060
D vs E	0.0001	0.0002	0.0001	0.0003	0.0009	0.0108	0.0106	0.0379
Linognathus vituli								
A vs E	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0003
B vs E	<.0001	<.0001	<.0001	0.0001	0.0001	0.0006	0.0003	0.4233
C vs E	<.0001	<.0001	<.0001	<.0001	0.0001	0.0002	0.0002	0.0079
D vs E	0.0010	0.0002	<.0001	<.0001	0.0001	<.0001	<.0001	0.0009
Solenopotes capillatus								
A vs E	0.0279	0.0348	0.0406	0.0516	0.0707	0.0730	0.0667	0.0707
B vs E	0.0445	0.0579	0.1006	0.1240	0.1347	0.1340	0.1245	0.1358
C vs E	0.0346	0.0420	0.0489	0.0637	0.0854	0.0881	0.1090	0.1090
D vs E	0.1006	0.0927	0.0827	0.0949	0.1245	0.0950	0.0919	0.1006

[1] A = Spinosad 400 ppm spray, B = Spinosad 2 mg/kg pour-on, C = CoRal 5.8% spray, D = Cylence 1% pour-on, E = Untreated control.

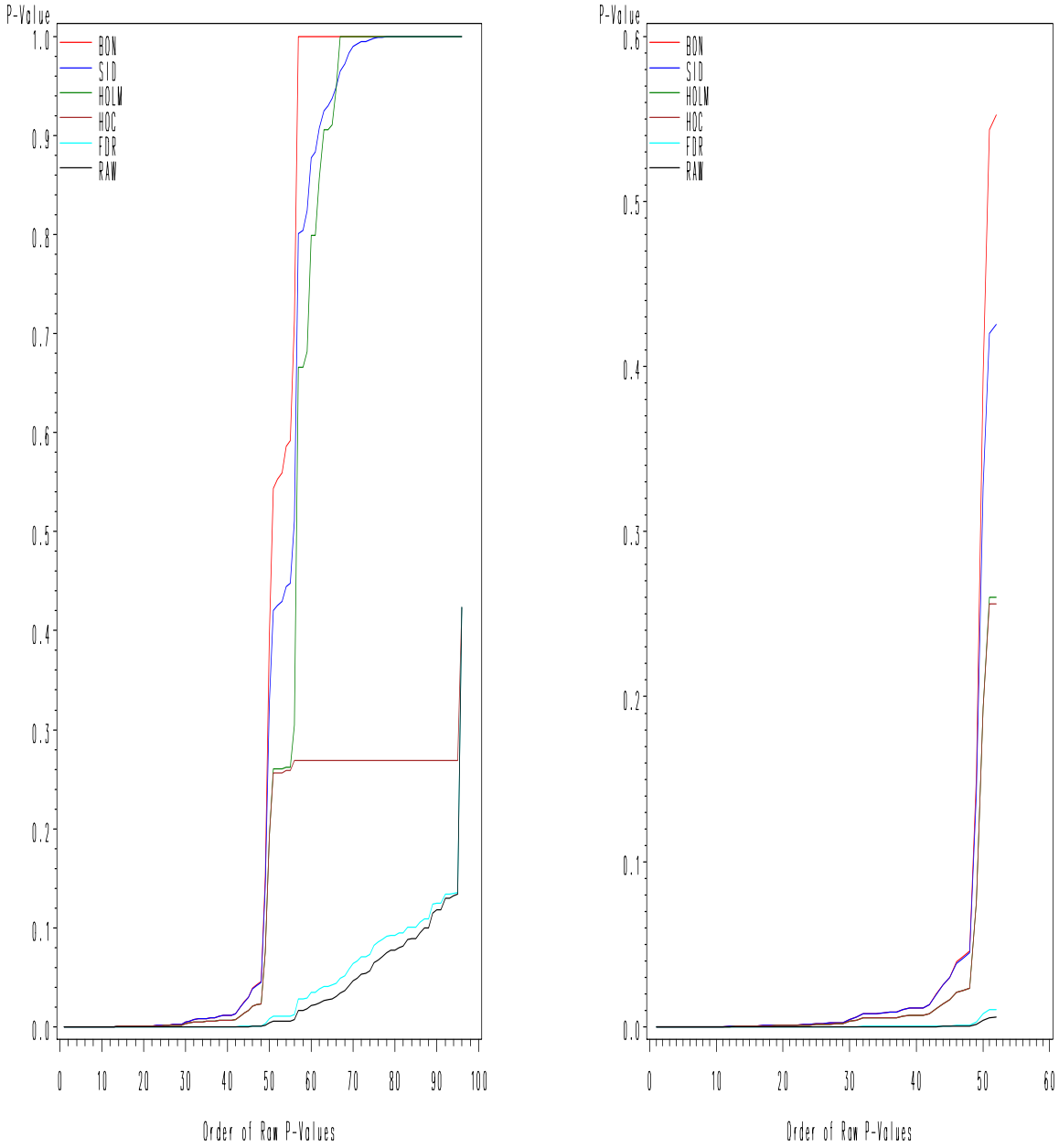

Figure 1. Comparison of the magnitude of the non-adjusted raw p-values and the adjusted p-values from Bonferroni, Sidak, Holm's step-down Bonferroni, Hochberg's step-up Bonferroni, and false discovery rate methods. Panel on the left shows the 96 raw p-values and their corresponding adjusted p-values. Panel on the right shows the first 52 smallest raw p-values and their corresponding adjusted p-values in a closer look.