IMPACT OF IMPLANTS ON PERFORMANCE AND CARCASS VALUE OF BEEF CATTLE: STATISTICAL PROBLEMS

J. W. Oltjen
Department of Animal Science
University of California
Davis, CA 95616

ABSTRACT

Implant trials present several statistical concerns. The power of the statistical test depends on the ratio of the difference between the two population means we want to detect and the standard deviation of the populations from which the observations are drawn. For uniform feedlot steers, an implant effect of .25 lb/d and a SD of .5 lb/d, about 60 animals per implant treatment will be needed to have a .67 probability of finding a significant difference. For a power of .95, we need about 100 animals. When comparing results over a number of independent trials, the use of the difference between treatments provides a more powerful estimate than using the absolute means. In planning trials with a limited numbers of large pens, designs with more than one implant treatment per pen can provide statistically valid results for individual animal variables. Choice of experimental endpoint may affect treatment differences so that statistical significance is a function of that choice. Feedlot simulations of implant treatments for yearling steers were conducted for days on feed, body weight, and quality grade endpoints. As values chosen for all endpoints increase, the mean average daily gains converge with increasing time on feed, resulting in larger differences between treatments earlier in the feeding period. In contrast, quality grade differences between treatments diverged with increasing days on feed while they converged with an increasing body weight endpoint. Experimental results using body weight endpoints may show fewer differences between implant treatments than those using days on feed. The number of animals per treatment needed to detect gain differences observed for the various endpoints can be estimated. Over 100 steers are needed to have two chances out of three for detecting differences between implant protocols simulated; 200 or more steers are needed for detecting differences at most body weight endpoints at normal slaughter weights.

INTRODUCTION

When planning or reviewing results from implant comparison trials, several statistical concerns should be considered. In particular, the statistical power (probability) of the trial to find biological or economical differences provides a convenient starting point, but power often is overlooked. For trials where the number of pens limited, and implant treatments must be applied within pens, there are valid protocols for experiments, but the ability to make meaningful inferences on intake or efficiency effects is limited. The choice of experimental endpoint may inflate or contract experimental differences; hence statistical significance is a function of that choice. For each of these considerations, problems for interpretation of results occurs. It is imperative for those concerned to determine if the experimental conditions are appropriate for their particular interest.

Statistical Concerns

Power of Tests in Experimental Design In a typical statistical comparison of implant treatments, we have a null hypothesis (H_0: the treatment means are equal) and an alternate hypothesis (H_1: the treatment means are different). An experiment is planned or conducted to gather evidence to reject H_0 and accept H_1 usually by developing some statistic (F, t) with a known statistical distribution to test against. For example, if the calculated t statistic from an experiment (the difference between the means divided by the standard error of the difference) is larger than the tabular (expected) value of t based on its known distribution when the null hypothesis is true, then we have evidence to reject H_0. The statistical error, or probability, we often report (α) is that of rejecting the null hypothesis when it is really true (Table 1). However, and more importantly, when planning an experiment, we ought to be more concerned with the power of the test ($1-\beta$), i.e., the probability of finding a statistically significant result when the null hypothesis is false (reject H_0 because there is a real difference).

Table 1. Statistical tests and the probability (P) of error for the null hypothesis (H_0).

Decision	H_0 is true	H_0 is false
Accept H_0	Correct (P = 1-α)	Type II error (P = β)
Reject H_0	Type I error (P = α)	Correct (P = 1-β)

When we have an experiment which does not show a significant difference, it would ideally be because there is no difference between implants, or the difference is too small to be important. We control this with the power of the test.

Power (1-β) depends on the ratio between the difference between two population means we want to detect and the standard deviation of the populations from which observations are drawn. It also depends on the Type I error rate α (Steele and Torrie, 1980). The number of observations per treatment (n) to detect a difference (D) is:

$$n = (Z_{\alpha/2} + Z_{\beta})^2 \, 2 \, \sigma^2 / D^2$$

where σ is the population SD and Z is the standard normal probability. For example, rather uniform feedlot steers have a standard deviation for average daily gain of about .5 lb/d. If an implant effect of .25 lb/d or more is important (and we would like to confirm it experimentally), then for the ratio of .5 (.25/.5) and α=.05, we need 60 animals per implant treatment to have a .67 probability (power) to find a significant difference (Figure 1). For power of .95 we need about 100 animals. If we were interested in only a .05 lb/d difference (ratio of .1), 2% of 2.5 lb/d (a typical difference between similar types of implants), over 1,000 animals per treatment are needed for a power of only .67. Clearly university trials with 8-50 animals per treatment are of little value in consistently determining small but real differences between implants.

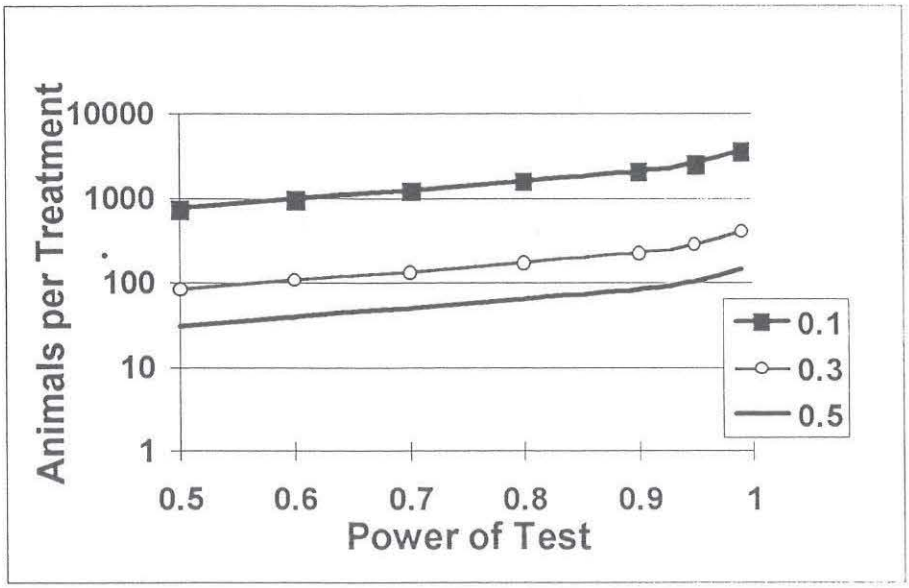

Figure 1. Animals needed per treatment to detect treatment differences at different ratios of the difference to the population SD (.1, .3, and .5) versus statistical power of the test (1-β).

Table 2. Summarizing literature data and the use of treatment means or treatment differences.

Trial:	A	B	C	Mean	SE	$SE_{difference}$
Control	2.0	3.0	3.4	2.8	.72	
						.61
Treatment	2.1	3.2	3.7	3.0	.75	
Difference	0.1	0.2	0.3	0.2		.10

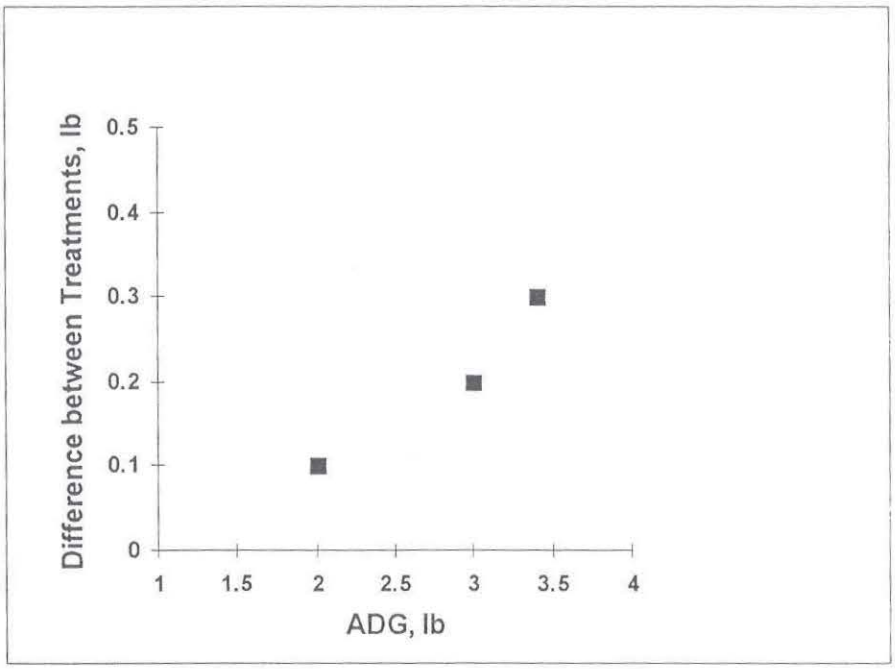

Figure 2. Plot of relationship between control average daily gain (ADG) and treatment differences for data from Table 2.

Summarizing Literature Data When conducting an analysis of implant treatments from previous experiments, the use of a treatment effect (differences) provides a more powerful statistic than does comparison of treatment means when using each trial as an observation (Table 2). The standard error of the difference between treatments is much smaller (.10 vs. .61). Also, it usually is instructive to plot the data and look for other relationships, as well as shown in Figure 2.

Within versus Across Pen Comparisons When implant trials are designed, all cattle in one pen usually are treated similarly with the same implant (assuming there are more pens than implant treatments). However, in many feedlots the number of pens for use in trials is limited even though pen size is large. In this case, statistical significance may be achieved by assigning multiple implants within each pen and treating each animal as an experimental unit, as long as there are enough animals. Of course, feed intake and efficiency data cannot be compared in such a trial.

Endpoint Effects Perhaps the most interesting statistical problems arising in implant experiments is the choice of trial endpoint. Does choice of endpoint affect overall animal performance, and thus statistical results? That is, do treatment differences depend on the endpoint chosen, or does experimental design depend on endpoint choice? In this paper, a simulation was chosen to study the effects of choosing either 1) a constant days on feed, 2) constant body weight, or 3) constant marbling endpoint. Medium

frame yearling steers with an initial weight of 700 lb and 50 lb SD were fed a high energy ration of .94 Mcal/kg DM NE_m and .62 Mcal/kg DM NE_g. Feed intake equations of Thornton et al. (1985) were used in the growth and composition model of Oltjen et al. (1986). Monte Carlo simulations were run, with proportional changes in maintenance (P_{maint}) and protein synthesis (P_{ps}) so that the coefficients of variation in the model were 33% for maintenance and 7% for protein synthesis, based on analysis of University of California research data (unpublished). Protein degradation was not made stochastic, so variation in protein accretion is solely due to the stochastic generation of P_{ps}. The large CV for maintenance is the sum of the variation in maintenance and fat deposition; they are not independently estimated by the growth model. Proportional change of feed intake (P_{DMI}) was adjusted as follows:

$$P_{DMI} = .2\ P_{maint} + .05\ P_{ps} + \varepsilon$$

where ε is normally distributed with mean zero and SD .1. For 130 d simulations, this results in SD of average daily gain and dry matter intake of .8 and 2.7 lb/d, respectively. Implant treatments (Figure 3) included none (CONTROL); protein synthesis and DMI increased by 4% and 10%, respectively, at 50 d, then linearly reduced to no effect by 100 d (ONE); protein synthesis and DMI increased by 4% and 10%, respectively, at 100 d, then linearly reduced to no effect by 150 d (TWO); protein synthesis and DMI increased by 6% and 15%, respectively, for 100 d, then linearly reduced to no effect by 150 d (TWO+). Five hundred animals were simulated for each treatment for each run; when body weight or quality grade endpoints were chosen, all 500 steers were slaughtered when the pen mean body weight or quality grade was achieved. Quality grade is an empirical estimate based on empty body fat in the model.

For a constant days of feed endpoint, the mean body weights began to converge with increasing time on feed, resulting in larger differences between treatments earlier in the feeding period for treatment average daily gains (Figure 4a). Conversely, quality grade differences between treatments diverged with days on feed, with the implant treatments becoming different after 140 days on feed (Figure 4b)

Figure 3. Effect of implant treatments (see text for description) on increase in protein synthesis for simulations of steer performance.

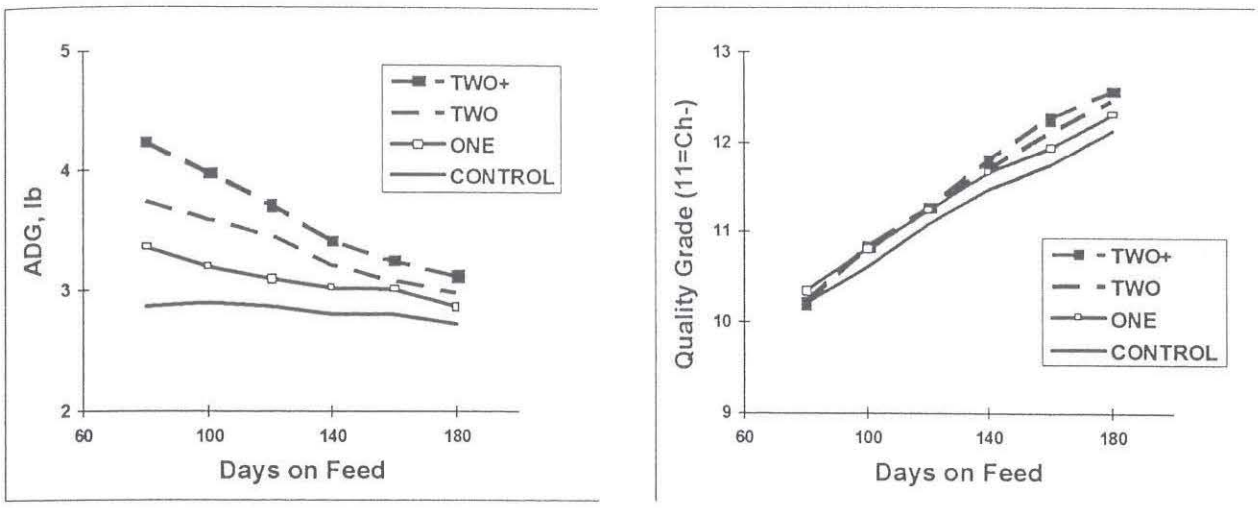

Figure 4. Mean average daily gain and quality grade of 500 steers simulated to a given days on feed for implant treatments (see text).

When pen mean body weight was the endpoint (Figure 5a), differences between treatment gains decreased with heavier endpoints, just as with longer feeding periods above. However, differences between quality grades narrowed with heavier endpoints, unlike the larger differences with increasing days on feed (Figure 5b). Thus, composition tends to reach a common point at a given body weight, if cattle are fed long enough to reach it. Thus experimental results using body weight endpoints may show fewer differences between implant treatments than those using days on feed. In a production sense that is fine, because cattle may be fed to a given body weight (as long as it is heavy enough) regardless of implant treatment with little effect on quality grade (and composition). However, some compromise in gain may be experienced with large body weight (or days on feed) endpoints.

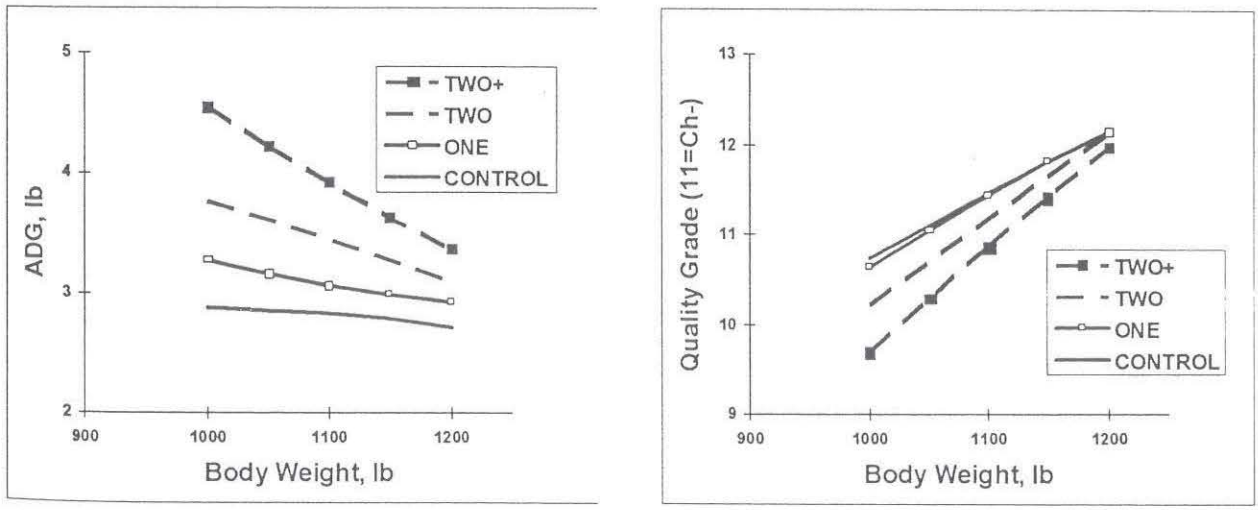

Figure 5. Mean average daily gain and quality grade of 500 steers simulated to a given mean body weight for implant treatments (see text).

If quality grade is used as the pen endpoint (Figure 6), body weight and average daily gain converge at higher grades, or increased body fatness. This is expected based on the above discussion of body weight endpoint, where compositions converged with increasing body weight.

The above results have important implications for experimental design of implant trials. Using the formula (Steele and Torrie, 1980) to estimate the number of observations per treatment to detect the average daily gain differences observed for the various endpoints, α of .05 and 1-β (power) of .67, animals per implant can be estimated (Figure 7). Unless short feeding periods to light body weights and quality grades are used, over 100 steers are needed to have two chances out of three (power of .67) to find significant effects for the differences simulated above. If body weight is the endpoint, the comparison of control and one implant requires nearly 200 or more animals at normal slaughter weight endpoints.

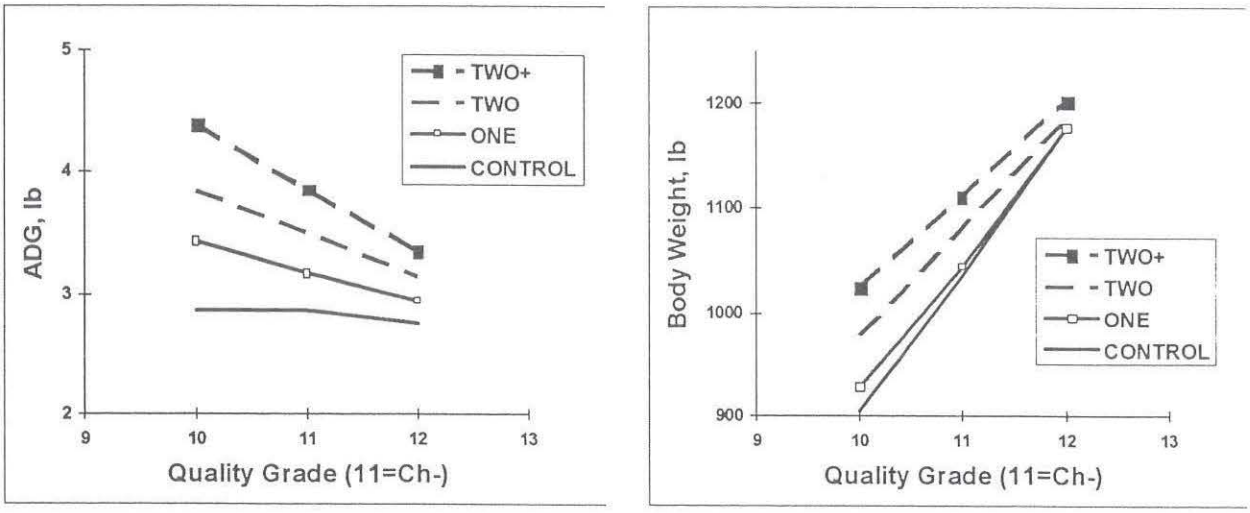

Figure 6. Mean average daily gain and body weight of 500 steers simulated to a given mean quality grade for implant treatments (see text).

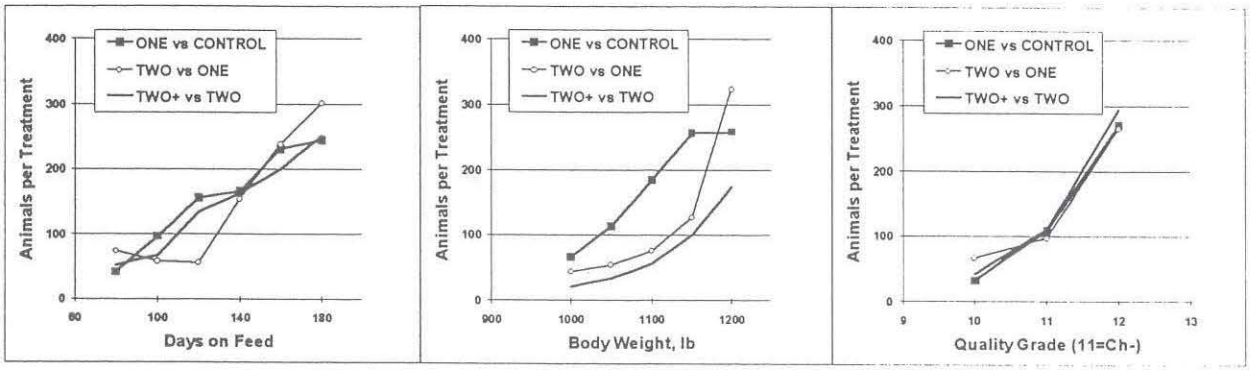

Figure 7. Number of animals needed to find significant treatment differences simulated between implant treatments (see text) for days on feed, body weight, or quality grade endpoints and α of .05 and 1-β (power) of .67.

Table 3. Number of animals needed to detect a significant treatment difference in ADG between implant treatments (see text) for days on feed, body weight, or quality grade endpoints and α of .05 and 1-β (power) of .67.

Endpoint	ONE vs CONTROL	TWO vs ONE	TWO+ vs TWO
Days on feed, 120 d	156	55	134
Mean body weight, 1,100 lb	185	76	57
Mean quality grade, low choice	109	97	110

These results apply only for treatments with the parameters and coefficients of variation described previously. Nevertheless, the general trends are likely to be valid regardless of how precise the estimates of treatment effects used are. For arbitrary endpoints of 120 days on feed, 1,100 lb body weight, or low choice quality grade, animals needed per treatment are least for the comparison between TWO and ONE (55), TWO+ and TWO (57), or ONE and CONTROL (109), respectively (Table 3). That is, steers are most different in average daily gain relative to the SD of ADG due to the treatment comparison. Also, in sequentially collected body weight (gain) data, longevity of implant response diminishes with time.

Implications

In planning or interpreting implant trials, the power of the statistical test should be considered; adequate numbers of animals per treatment should be used to detect the most important difference relative to the standard deviation for the animals. Often more than 100 animals per treatment are needed to detect subtle differences between implant treatments. When summarizing multiple trials, analysis of treatment effects (differences or proportional changes), not absolute values, are more sensitive. Treatment effects should be related to other variables to look for new relationships which may increase understanding. By simulation, choice of endpoint, (both the variable chosen and its value) probably affects the treatment effect and the number of animals needed to determine a significant effect. In general, differences between treatments are larger for shorter trials, or for animals fed to lighter weights. The greater the variability in animal management, the greater the number of animals needed to detect significant treatment effects.

LITERATURE CITED

Oltjen, J.W., A. C. Bywater, R. L. Baldwin and W. N. Garrett. 1986. Development of a dynamic model of beef cattle growth and composition. J. Anim. Sci. 62:86.

Steele, R.G.D., and J. H. Torrie. 1980. Principals and Procedures of Statistics: A Biometrical Approach (2nd Ed.). McGraw-Hill Publishing, New York.

Thornton, J.H., F. N. Owens and D. R. Gill. 1985. Feed intake by feedlot beef steers: Influence of initial weight and time on feed. Okla. Agric. Exp. Sta. Misc. Publ. MP117:320-331.

QUESTIONS & ANSWERS

Q: Why does it take more animals per treatment to achieve significance when your experiencing control versus weighting time? What I am talking about is magnitude or response will be bigger as between two different implants.

A: It does for days on feed and it does not for body weight. As we look at average daily gain the differences between average daily gain confers with increasing body weight. The differences here are greater at this point than they are down here. I think what your asking is why does it take more animals for these treatments. The main reason why you need more animals here is because the effect of the implants here runs about goes 50 days and then it starts to run out and it goes down to 0 in 100 days so that the animals tend to come together in

terms of their performance and in terms of their average daily gain numbers. That's why I went back to the previous slide to show you the average daily gain numbers as we get out further into those implants. More so faster effect treatment than it does with these others because , particularly between these two, where the implants last longer. And again the relative difference is larger in all cases but basically the quality grades of those two different cures come pretty close together. Average daily gains come pretty close together and a different quality grain but it takes a while to get there.