

The power of tests for bioequivalence in feed experiments with poultry¹

W. B. Roush*² and P. R. Tozer†

*USDA-ARS Poultry Research Unit, Mississippi State, MS 39762 and

†Department of Dairy and Animal Science, Pennsylvania State University, State College 16802

ABSTRACT: Several studies have compared the feeding of genetically modified (GM) grains and conventional grains to poultry. The general conclusion has been that there were no significant differences detected in the biological performance of the birds (i.e., the grains were bioequivalent). However, the question has been posed whether the experimental designs used in the studies had sufficient statistical power to detect treatment differences. The power of tests can be used to determine the ability of an experimental design to detect treatment differences. The definition of statistical power is the probability of rejecting the null hypothesis when it is false and should be rejected. The complement of statistical power is the Type II error (β). That is, accepting the null hypothesis that there is no difference in treatments when there is one. *A priori* power analysis can indicate the probability at which the sampling regi-

men or experiment can actually detect an effect if a difference exists. Post hoc power analysis indicates the sufficiency or the sample size needed for an experiment that has already been conducted. In the current study, the power of tests for experiments published in the literature where significant and nonsignificant differences were reported between control birds and birds fed new feed grains was examined. With some exceptions, the power of tests is rarely formally considered or mentioned in poultry research. The results of the survey of the literature showed, in general, low power of statistical tests for feeding experiments involving non-GM grains or in those cases when GM and non-GM grains were compared in poultry feeding experiments. These results suggest that care needs to be taken when designing experiments for bioequivalence of grains fed to poultry.

Key Words: Bioequivalence, Genetically Modified Grains, Poultry, Power Tests

©2004 American Society of Animal Science. All rights reserved. J. Anim. Sci. 2004. 82(E. Suppl.):E110–E118

Introduction

Several studies have been conducted involving the comparison of feeding genetically modified (GM) grains with conventional grains to poultry. The general conclusion has been that there were no significant differences in the biological performance of the birds, implying that the GM grains are bioequivalent with the non-GM grains (Table 1). However, the question has been proposed whether the statistical designs used in the studies had sufficient statistical power to detect biologically important differences.

Researchers typically design experiments to determine if differences between experimental treatments exist. The statistical difference between treatments is

tested with the null hypothesis that there is no difference between treatments. The researchers test the null hypothesis and typically set a significance level for the test (α) at a predetermined level, usually 0.05. The significance level of the test measures the probability of making a Type I error. A Type I error is the rejection of a true null hypothesis (Steel et al., 1997).

Poultry, compared with other agricultural and companion species, have an advantage as experimental subjects. Generally, poultry are small in size, relatively docile, available in large numbers, and they have a short production cycle. Despite these advantages, sampling for poultry studies is still subject to the availability of facilities, availability of animals, ethical considerations, limited budgets, convenience, and tradition. With some exceptions (e.g., Hammond et al., 1996; Hall et al., 2003), researchers conducting poultry studies have not mentioned the consideration of the analysis of power of statistical research designs. Statistical power analysis shows whether an experiment is capable of detecting differences in the treatment mean values for given level of α . Formally, statistical power is defined as the probability of rejecting the null hypothesis when it is false and should be rejected

¹This article was presented at the 2003 ADSA-ASAS-AMPA meeting as part of the Contemporary Issues symposium "Designing Animal Experiments for Power."

²Correspondence: P.O. Box 5367 (phone: 662-320-7480; fax: 662-320-7589; e-mail: broush@ars.usda.gov).

Received July 9, 2003.

Accepted August 22, 2003.

Table 1. Reported effects of genetically modified grains in comparison with conventional feeds fed to broilers and laying hens (modified from OECD, 2003)

Poultry	Grain ^a	Effects noted
Broilers and laying hens ^b	<i>Bt</i> -maize	No significant effects on production measurements or ME content
Laying hens ^c	<i>Bt</i> -maize	No significant effects on nutrient digestibility or AME _N . ^k Egg production not examined
Broilers ^d	<i>Bt</i> -maize	No significant differences in live weight gain or survival. Feed:gain significantly improved for <i>Bt</i>
Broilers ^e	<i>Bt</i> -maize	No significant differences in performance measurements compared with near isogenic parent lines
Broilers ^f	<i>Bt</i> -maize	No significant differences in live weight gain, DMI or feed:gain ratio
Broilers ^g	Herbicide tolerant maize, <i>Bt</i> maize	No significant differences live weight gain, DMI, feed:gain ratio, or survival
Broilers ^h	<i>Bt</i> -soybean meal	No significant differences in growth, feed:gain ratio, carcass weight or breast meat
Broilers ⁱ	<i>Bt</i> -maize	No significant differences in weight gain or feed:gain ratio and true ME values
Broilers ^j	Herbicide tolerant maize	No significant differences in live weight gain, feed:gain and fat pad weights

^a*Bt* = *Bacillus thuringiensis*.

^bAeschbacher et al., 2001.

^cAulrich et al., 1998.

^dBrake and Vlachos, 1998.

^eGaines et al., 2001.

^fHalle et al., 1999.

^gHammond et al., 1996.

^hKan et al., 2001.

ⁱMireles et al., 2000.

^jSidhu et al., 2000.

^kAME_N = Nitrogen-corrected apparent ME.

(Table 2). The power of a statistical test is calculated as $1-\beta$, where β is the probability of making a Type II error. A Type II error is accepting a null hypothesis that is false (Steel et al., 1997).

If the test shows that a significant difference at the predetermined level of α is present, then usually there is no need to further consider the statistical power of the study (Sheppard, 1999). However, if no significant differences are detected, then a post hoc statistical power analysis may be conducted. It could mean that there is actually no difference or that the sample size was insufficient to detect a treatment difference. Researchers are occasionally tempted to report numerical (but statistically nonsignificant treatment) differences, a practice that is discouraged in journals such as *Poultry Science*. The question then arises as to whether detection of significant differences in the means would have been made had there been enough replications in the experiment.

Traditionally, scientists avoid committing Type I (α) rather than Type II (β) statistical errors. That is, they want to avoid the conclusion that there is a difference in treatments, when a treatment difference does not exist (Type I error). With some studies, however, it

may be important to avoid errors at the other extreme. That is, a change is made in the perspective of the hypothesis from a no-difference null hypothesis ($H_0: \mu_1 = \mu_2$) to a nonequivalence null hypothesis ($H_0: \mu_1 \neq \mu_2$; Hoenig and Heisey, 2001). For example, when feeding animals GM grains, it may be more desirable to have no differences in biological responses. Type II statistical errors may occur when sample sizes are too small to show an effect. An *a priori* power analysis can be used to show the sample size and probability at which the statistical design would show a difference. It should be pointed out that, ideally, the appropriate use of power analysis is made during the planning stage of the experiment. It is inappropriate to use post hoc power analysis on observations in an effort to justify conclusions about the responses (Lenth, 2001). As an alternative, Hoenig and Heisey (2001) recommend the use of confidence intervals for interpreting inconclusive experiments. In this paper, the power analysis was used to gain insight into the capability of published research on grain comparisons to detect treatment differences.

A key component of power analysis is the effect size. Thomas and Krebs (1997) point out that effect size is the difference between the null and alternative hypotheses, and can be measured by using raw or standardized values. An example of a raw measure is the difference between means. Alternatively, standardized measures as shown in Table 3 are dimensionless and incorporate the sampling variance.

Cohen (1962) was the first to conduct a systematic literature survey of statistical power analysis in the behavioral sciences. For survey purposes, Cohen

Table 2. Probability errors and conclusions

Item	Reality	Conclusion
Confidence level ($1-\alpha$)	No difference	No difference
Type I error (α)	No difference	Difference
Type II error (β)	Difference	No difference
Power ($1-\beta$)	Difference	Difference

Table 3. Population parameters defining the levels of size of effect for various statistical tests (Cohen, 1962; Fox and Mathers, 1997)

Test	Formula ^a
<i>t</i> -test	$(\mu_a - \mu_b) / \sigma$
<i>F</i> -test	σ of means/pooled σ
Correlation (Pearson)	<i>r</i>
Multiple regression	$1 / (R^2 / 1 - R^2)$

^a μ_a = mean of treatment a; μ_b = mean of treatment b; σ = standard deviation.

(1962) suggested using standardized (unit-free) effect size calculations for various statistical tests (Table 3). He suggested calculating and categorizing the effects into small, medium, and large effects based on treatment number, total sample size, and probability level. Since Cohen's (1962) initial survey, several other literature surveys, including Rossi (1990) and Fox and Mathers (1997), have been conducted in the behavioral and medical sciences to determine the adequacy of the statistical power to detect differences in treatments. The surveys have shown that the power of statistics in research is, generally, insufficient to comfortably accept the null hypothesis.

It is conventional to set 80% as the target value for statistical power (Fox and Mathers, 1997). That is, unless the statistical design of a study has a power of at least 80%, and if the treatments are not significant at the $\alpha = 0.05$ level, then the results must be considered inconclusive (Lalouel and Rohrwasser, 2002).

This paper describes a survey of the power of tests in poultry-feeding research, including investigations with birds that were fed GM and non-GM feed grains. The intention of this paper is to suggest that statistical power analysis is a planning tool that may aid poultry researchers to enhance the credibility of feed grain comparisons and GM bioequivalence studies.

Methods and Procedures

Examination was made of the statistical power of recent poultry-feeding experiments that compared bird response to different non-GM feed grains, or the comparison of bird responses to GM and non-GM feed grains in poultry feeding experiments.

Survey Procedure

Examination was made of research comparing feed ingredients, including GM grains, to poultry. The objective was to estimate the power of experiments to detect effect sizes of 0.1, 0.4, and 0.8. A sample of 37 papers from 1995 to 2003 was analyzed for the relative power of tests using previous surveys as a guideline (Table 4). The treatments were determined for each experiment or trial. Only the tests (i.e., *t*-test, *F*-test or regression) designated by the authors were used for

Table 4. Distribution of the power of 37 poultry ingredient feeding studies according to population effect size calculated with G*Power ($\alpha = 0.05$)^a

Power	Effect size		
	0.1	0.4	0.8
0.99			15
0.95–0.98			3
0.90–0.94		1	6
0.80–0.89		2	3
0.70–0.79		3	5
0.60–0.69		4	3
0.50–0.59		5	0
0.40–0.49		5	1
0.30–0.39	1	7	1
0.20–0.29		6	
0.10–0.19	7	4	
0.05–0.09	29		
<i>n</i>	37	37	37
Mean	0.08	0.47	0.88
Median	0.07	0.47	0.96
Mode	0.08	0.69	1.00
σ	0.04	0.22	0.16
Q1 ^b	0.06	0.30	0.82
Q3 ^b	0.08	0.69	1.00

^aBased on the following studies: Ali and Leeson (1995); Hammond et al. (1996); Ravindran et al. (1996a,b); Parsons et al. (1997); Rosenfeld et al. (1997); Wang and Parsons (1997); Brake and Vlachos (1998); Leeson (1998); Scott et al. (1998); Benitez et al. (1999); Del Carmen et al. (1999); Donkoh et al. (1999); Douglas et al. (1999); Fernandez et al. (1999); Slominski et al. (1999); Boling et al. (2000); Douglas and Parsons (2000); Gonzalez-Esquerra and Leeson (2000); Li et al. (2000); Mireles et al. (2000); Perez et al. (2000); Sidhu et al. (2000); Taylor et al. (2000a,b; 2003a,b); Aulrich et al. (2001); Gaines et al. (2001); Lee et al. (2001); Novak and Scheideler (2001); Piva et al. (2001); Bennett et al. (2002); Humphrey et al. (2002); Sterling et al. (2002); Svihus and Gullord (2002); Batal and Parsons (2003); Brake et al. (2003); and Dagher et al. (2003).

^bQ1 and Q3 represent the 1st and 3rd quartiles of the frequency distribution, respectively.

the analysis. Total observations or samples depending on the test were determined. The total sample size in this survey is defined as the number of observations summed over all of the treatment groups of the design. Statistical power was calculated based on the type of test. The power reported is an average for each paper. This allowed each paper to count equally in the survey. Descriptive statistics for the power survey, such as means, modes, and medians, were calculated in an MS Excel spreadsheet based on the collected data. A subset of 15 of the 37 representing GM and non-GM feed grain comparison experiments was examined for their power according to the three levels of effect size (Table 5).

Standard Conditions

The $\alpha = 0.05$ Type I error was used as a standard condition throughout the survey. That is, the null hypothesis in the studies is considered a no-difference null hypothesis ($H_0: \mu_1 = \mu_2$). Following the reasoning of Cohen (1962), the nondirectional version of the null hypothesis (two-tailed distribution) was used. The cal-

Table 5. Distribution of the power of 15 poultry studies related to the feeding of genetically modified grains according to population effect size calculated with G*Power ($\alpha = 0.5$)^a

Power	Effect size		
	0.1	0.4	0.8
0.99–			9
0.95–0.98			1
0.90–0.94			1
0.80–0.89		1	0
0.70–0.79		1	4
0.60–0.69		4	
0.50–0.59		3	
0.40–0.49		1	
0.30–0.39		1	
0.20–0.29		4	
0.10–0.19	1		
0.05–0.09	14		
n	15	15	15
Mean	0.07	0.52	0.92
Median	0.07	0.56	1.00
Mode	0.08	0.69	1.00
σ	0.01	0.21	0.12
Q1 ^b	0.06	0.30	0.82
Q3 ^b	0.08	0.69	1.00

^a Hammond et al. (1996); Brake and Vlachos (1998); Leeson (1998); Slominski et al. (1999); Mireles et al. (2000); Sidhu et al. (2000); Taylor et al. (2000a,b; 2003a,b); Aulrich et al. (2001); Gaines et al. (2001); Piva et al. (2001); Humphrey et al. (2002); and Brake et al. (2003).

^bQ1 and Q3 represent the 1st and 3rd quartiles of the frequency distribution, respectively.

ulation of the power of tests was conducted with G*Power, a freeware program. Buchner et al. (1997) provide a downloadable version of the G*Power program with supporting documents.

Effect Size

Cohen (1962) suggested an arbitrary classification of tests according to small, medium, and large effects. Sheppard (1999) pointed out that the meaning of small, medium, and large are likely to differ between disciplines and may not directly correspond to a perceived degree of importance for the variable(s) of interest. The levels determined by Cohen (1988) for social sciences may not easily correspond to a contaminant level for example. With this in mind, the calculated power levels in this paper are reported for effect sizes of 0.1, 0.4, and 0.8. These effect sizes might be roughly thought of as small, medium and large. The effect sizes were calculated according to the statistical test (i.e., *t*-test, *F*-test or regression) employed in the respective papers. Table 3 shows the calculation of the effect size for various statistical tests according to Cohen (1962).

Results and Discussion

The descriptive statistics and dispersal of the papers for the power values according to the effect sizes for

the non-GM grain and GM grain studies are shown in Tables 4 and 5, respectively. On average, the surveyed studies had less than one chance in 12 or 13 of detecting a small effect (0.1). None of the studies had as much as a 50% chance of detecting a slight effect. For a medium effect size (0.4), the studies averaged less than a 50% chance of successfully rejecting a null hypothesis for ingredient studies, and approximately a 50% probability for the subset of GM grain feeding studies. Large size effects (0.8) needed fewer samples for detection than small effects. When the effect size was large, the power of the studies met and exceeded the minimum of 80% power that is considered a standard.

Table 6 shows post hoc power test calculations of selected production and carcass measurements from the literature for laying hens and broilers. The standardized effect sizes were calculated from the standard deviation of treatment means and the pooled standard deviation. For 17 of the 19 variables measured, the power of the test was considerably below the conventional power level of 0.8. Exceptions are noted for the grain energy level evaluations for both broilers and layers. The average effect size for variables, not including the energy values, was 0.28 ± 0.14 , and with the energy values, the average effect size was 0.40 ± 0.39 . The average power for the variables measured, not including energy values, was 0.24 ± 0.15 . When the energy power values were included, the average power size was 0.31 ± 0.26 . These effect sizes and powers provide the researcher a sense of the magnitude for these values found in the literature. The generally small power values for these effect sizes suggest the sample sizes are low. (In many cases, and in the discussion that follows, sample size and number of experimental units can be used interchangeably.) Subsequently, an *a priori* test was conducted to examine the sample size that would be required to produce a power of 80% (Table 7). The results illustrate the dilemma in determining sample size. For some variables, such as digestibility of CP and energy levels, the required sample size might be considered reasonable. For other physiological variables, the sample size would probably be considered quite unreasonable (e.g., 1,890 and 2,480 total samples for feed conversion and thigh percent, respectively).

As an alternative, a “compromise” option is available in G*Power (Erdfelder, 1984), which is based on a balance between Type I (α) and Type II (β) errors. In the *compromise* approach, the researcher specifies the size of effect to be detected, the sample size that seems reasonable, and the error ratio that defines the relative seriousness of both types of error. Given this, an optimal critical value for the test statistic and the associated α and β values are computed. An illustration of the *compromise* option is shown in Table 8. The calculation of power was based on an arbitrary sample size of 120 (60 observations for each grain treatment) and an equal ratio between α and β (i.e., ratio = 1). The

Table 6. Post hoc power test calculation of selected measurements from the literature for laying hens and broilers (*F*-test, $\alpha = 0.05$). Power was calculated with G*Power^a

Measurement	σ of means	Pooled σ	Effect size ^a	Power
Laying hens				
Hen day egg production, % ^c	1.26	4.68	0.27	0.12
Feed intake, g·bird ⁻¹ ·d ^{-1c}	0.90	2.58	0.35	0.17
Feed conversion, g of feed/g of egg ^b	0.02	0.31	0.06	0.05
Digestibility of CP, % ^b	0.40	1.05	0.38	0.22
AME _N , MJ/kg of DM ^{be}	0.22	0.13	1.76	1.00
Egg weight, g ^c	0.68	1.54	0.44	0.25
Haugh units ^c	1.94	4.87	0.40	0.21
Yolk color (Roche) ^c	0.12	0.35	0.35	0.17
Broilers				
Body weight, g ^d	41.1	260.8	0.16	0.15
Feed conversion, g/g ^b	0.01	0.05	0.20	0.10
Digestibility of CP, % ^b	0.95	1.50	0.63	0.50
AME _N , MJ/kg of DM ²	0.26	0.24	1.06	0.91
Dressed carcass, % ^d	0.35	1.44	0.24	0.31
Fat pad, % ^d	0.19	0.55	0.35	0.61
Drums, % ^d	0.12	0.67	0.18	0.19
Thighs, % ^d	0.06	0.89	0.07	0.07
Wings, % ^d	0.11	0.38	0.28	0.43
Pectoralis major, % ^d	0.26	1.15	0.22	0.30
Pectoralis minor, % ^d	0.05	0.26	0.18	0.19
Mean \pm SD, values without energy calculations			0.28 \pm 0.14	0.24 \pm 0.15
Mean \pm SD, values with energy calculations			0.40 \pm 0.39	0.31 \pm 0.26

^aEffect size = σ of means/pooled σ . Calculated with G*Power (available at www.psych.uni-duesseldorf.de/aap/projects/gpower/).

^bNumber of treatments = 2; total number of samples = 12.

^cNumber of treatments = 3; total number of samples = 15.

^dNumber of treatments = 4; total number of samples = 64.

^eAME_N = nitrogen-corrected apparent ME.

Table 7. *A priori* power test calculation of selected measurements from the literature for laying hens and broilers (*F*-test, $\alpha = 0.05$). Sample size was calculated with G*Power^a

Measurement	Effect size ^a	Power	Calculated sample size
Laying hens			
Hen day egg production, % ^c	0.27	0.8	135
Feed intake, g·bird ⁻¹ ·d ^{-1c}	0.35	0.8	84
Feed conversion, g of feed/g of egg ^b	0.06	0.8	1,890
Digestibility of CP, % ^b	0.38	0.8	58
AME _N , MJ/kg of DM ^{be}	1.76	0.8	6
Egg weight, g ^c	0.44	0.8	54
Haugh units ^c	0.40	0.8	66
Yolk color (Roche) ^c	0.35	0.8	84
Broilers			
Body weight, g ^d	0.16	0.8	444
Feed conversion, g/g ^b	0.20	0.8	200
Digestibility of CP, % ^b	0.63	0.8	22
AME _N , MJ/kg of DM ^b	1.06	0.8	10
Dressed carcass, % ^d	0.24	0.8	196
Fat pad, % ^d	0.35	0.8	96
Drums, % ^d	0.18	0.8	332
Thighs, % ^d	0.07	0.8	2,480
Wings, % ^d	0.28	0.8	140
Pectoralis major, % ^d	0.22	0.8	224
Pectoralis minor, % ^d	0.18	0.8	332

^aEffect size = σ of means/pooled σ . Calculated with G*Power (available at www.psych.uni-duesseldorf.de/aap/projects/gpower/).

^bNumber of treatments = 2; total number of samples = 12.

^cNumber of treatments = 3; total number of samples = 15.

^dNumber of treatments = 4; total number of samples = 64.

^eAME_N = Nitrogen-corrected apparent ME.

Table 8. Compromise power calculation of selected measurements from the literature for laying hens and broilers (*F*-test) where total possible sample size is held at 120 and the β/α ratio = 1. The $\alpha = \beta$ value and power were calculated with G*Power^a

Measurement	Total sample size	Effect Size ^a	$\alpha = \beta$	Power
Laying hens				
Hen day egg production, % ^c	120	0.27	0.13	0.87
Feed intake, g·bird ⁻¹ ·d ^{-1c}	120	0.35	—	—
Feed conversion, g of feed/g of egg ^b	120	0.06	0.45	0.55
Digestibility of CP, % ^b	120	0.38	—	—
AME _N , MJ/kg of DM ^{be}	120	1.76	—	—
Egg weight, g ^c	120	0.44	—	—
Haugh units ^c	120	0.34	—	—
Yolk color (Roche) ^c	120	0.35	—	—
Broilers				
Body weight, g ^d	120	0.16	0.33	0.67
Feed conversion, g/g ^b	120	0.20	0.27	0.73
Digestibility of CP, % ^b	120	0.63	—	—
AME _N , MJ/kg of DM ^b	120	1.06	—	—
Dressed carcass, % ^c	120	0.24	0.19	0.81
Fat pad, % ^d	120	0.35	—	—
Drums, % ^d	120	0.18	0.28	0.72
Thighs, % ^d	120	0.07	0.46	0.54
Wings, % ^d	120	0.28	0.14	0.86
Pectoralis major, % ^c	120	0.22	0.22	0.78
Pectoralis minor, % ^d	120	0.18	0.28	0.72

^aEffect size = σ of means/pooled σ . Calculated with G*Power (available at www.psych.uni-duesseldorf.de/aap/projects/gpower/).

^bNumber of treatments = 2; total number of samples = 12.

^cNumber of treatments = 3; total number of samples = 15.

^dNumber of treatments = 4; total number of samples = 64.

^eAME_N = nitrogen-corrected apparent ME.

challenge to the researcher is that the *compromise* results may give nonstandard levels for α and β , as indicted in the example. That is, probability values other than 0.05 or 0.01.

Yet another approach is shown in Table 9. In this case, the researcher would like to maintain a balance of $\alpha = 0.05$ and $\beta = 0.20$ (power = 0.80). These sample sizes were calculated using the *a priori* option of G*Power according to different effect sizes. Again, the detection of small effect sizes requires large sample sizes.

Power is a function of three variables: 1) sample size (Cohen, 1962, Fox and Mathers, 1997); 2) the effect size (the degree of departure in the population from

the null hypothesis); and 3) the level of significance (α or Type I error). The following are some thoughts on how these variables, and power itself, might be adjusted to improve the ability of the researcher to make statistical conclusions.

Sample Size

Berndtson (1991) provides equations for determining numbers of replicates and tables for determining replication at power levels of 0.8, 0.90, and 0.95. Software such as G*Power gives some flexibility in determining sample sizes. Sampling strategies have been noted using a *compromise* approach to balance Type I and Type II errors. An *a priori* approach was suggested to maintain a relationship between Type I and Type II errors (e.g., $\alpha = 0.05$ and $\beta = 0.20$). However, there is no absolute size of a sample that is best. Declaration of a critical number of samples as being absolutely necessary is tempting. The interaction of strain of bird, grain varieties, environmental conditions etc. makes it impossible to make concrete declarations of sample size or levels of significance. The reference section for the G*Power program (Buchner, 1997) illustrates approaches to determining sample sizes for interactions in multifactor designs.

Generally, the cost of large sample sizes makes it uneconomical in terms of money and time. Reductions

Table 9. *A priori* sample sizes for a balance between $\alpha = 0.05$ and $\beta = 0.20$ (power = 0.80) for small (0.1), medium (0.4), and large (0.8) effect sizes. Sample sizes were calculated with G*Power^a

Treatment size	Effect size		
	Small (0.1)	Medium (0.4)	Large (0.8)
2	788	52	16
3	969	66	21
4	1096	76	24

^aG*Power available at www.psych.uni-duesseldorf.de/aap/projects/gpower/.

of sample sizes can come at a cost in power and may result in probabilities other than the conventional 0.05 and 0.01 levels.

The sample size is related to the amount of variance in the experiment. Accordingly, the variance can be affected by relationship between the experimental unit and the subsamples making up the experimental unit. The treatment variance decreases as the number of replicates increase, at the expense of the number of subsamples in the treatment replicate (Steel et al., 1997). In other words, as the replications of pens increase and the bird number per pen decreases, there is a decrease in treatment variation.

Effect Size

An observation can be made that the larger the effect size, the lower the number of replicates is needed for sufficient statistical power. However, Lenth (2001) cautions that the standardized effect size (as a ratio of the pooled standard deviation of the means to the error standard deviation) has no relevance, in itself, as a criterion for determining sample size. Lenth (2001) argues that the numerator and denominator of the effect size ratio for the *t*-test and *F*-test (Table 3) need to be looked at separately. For example, the effect size can be improved by lowering the standard deviation value of the denominator. The type of measurement tools used can be effective in more accurately determining measurements and reducing variance. Another approach is to reduce experimental variation by using sensitive research designs (e.g., blocking, covariates, etc.).

Significance (α) Level

In science, the tendency is to think in terms of absolutes. This orientation is inherited, no doubt, from the hard disciplines of mathematics, physics, and chemistry, where accuracy and precision are hallmarks. In the biological sciences, variability is ubiquitous. Statistics in research is commonly driven by the almost magical values of 0.05 and 0.01 (Sheppard, 1999). As Sheppard (1999) points out, it must be kept in mind that these values are convenient cut-off values. Gill (1981) referred to the strict adherence to the 0.05 level as the cult of $P < 0.05$ or perish. That is, if P is 0.06, then the treatments are not considered significantly different by some researchers. One approach to put experimental results in perspective is to report the probability level at which the treatments are significant. These values are available in most statistical package printouts.

Power ($1-\beta$)

Improvement of statistical power can be made by the use of proper experimental designs. For example, quantitative experiments are occasionally analyzed by ANOVA. The analysis of a quantitative experiment as

a regression will increase the power of the test. It should be kept in mind that for quantitative experiments, where the regression is significant, the differences between treatment values are significant.

The ability to conduct power tests has been enhanced. The website of R. V. Lenth (www.stat.uiowa.edu/rlenth/Power/) provides links to several power analysis calculators. In addition, Thomas and Krebs (1997) have reviewed a number software programs, available to researchers for power analysis and sample size determination.

Implications

With regard to safety (perceived or otherwise) and in an effort to preserve the credibility of the research, it is important that careful consideration be given to the experimental design for the detection of differences or lack of differences. With this in mind, Buhl-Mortensen (1996) pointed out that the awareness of how methods affect results in science is crucial if we want to appraise objectively what an investigation does or does not say about the actual extent of a given risk and will help to keep science as objective as possible. Accordingly, power analysis of experimental designs for planned research is encouraged to accomplish meaningful bioequivalent feed studies for poultry.

Literature Cited

- Aeschbacher, K., L. Meile, R. Messikommer, and C. Wenk. 2001. Genetically modified maize in diets for chickens and laying hens: Influence on performance and product quality. Pages 41–42 in Proc. Int. Symp. Genetically Modified Crops and Co-Products as Feeds for Livestock, Nitra, Slovak Republic.
- Ali, M. A., and S. Leeson. 1995. The nutritive value of some indigenous Asian poultry feed ingredients. *Anim. Feed Sci. Technol.* 55:227–237.
- Aulrich, K., H. Böhme, R. Daenicke, I. Halle, and G. Flachowsky. 2001. Genetically modified feeds in animal nutrition 1st communication: *Bacillus thuringiensis* (*Bt*) corn in poultry, pig and ruminant nutrition. *Arch. Anim. Nutr.* 54:183–195.
- Aulrich, K., I. Halle, and G. Flachowsky. 1998. Inhaltsstoffe und Verdaulichkeit von Maiskornen der Sorte Cesar und der gentechnisch veränderten *Bt*-hybride bei Legenhennen. Pages 465–468 in Proc. Einfluss von Erzeugung und Verarbeitung auf die Qualität landwirtschaftlicher Produkte (VDLUFÄ) Kongressband. VDLUFÄ-Kongress, Giessen, Germany.
- Batal, A. B., and C. M. Parsons. 2003. Utilization of different soyproducts as affected by age in chicks. *Poult. Sci.* 82:454–462.
- Bennett, C. D., H. L. Classen, and C. Riddell. 2002. Feeding broiler chickens wheat and barley diets containing whole, ground and pelleted grain. *Poult. Sci.* 81:995–1003.
- Benitez, J. A., A. G. Gernat, J. G. Murillo, and M. Araba. 1999. The use of high oil corn in broiler diets. *Poult. Sci.* 78:861–865.
- Berndtson, W. E. 1991. A simple, rapid and reliable method for selecting or assessing the number of replicates for animal experiments. *J. Anim. Sci.* 69:67–76.
- Boling, S. D., M. W. Douglas, J. L. Snow, C. M. Parsons, and D. H. Baker. 2000. Citric acid does not improve phosphorus utilization in laying hens fed a corn-soybean meal diet. *Poult. Sci.* 79:1335–1337.
- Brake, J., and D. Vlachos. 1998. Evaluation of transgenic event 176 “*Bt*” corn in broiler chickens. *Poult. Sci.* 77:648–653.

- Brake, J., M. A. Faust, and J. Stein. 2003. Evaluation of transgenic event *Bt11* hybrid corn in broiler chickens. *Poult. Sci.* 82:551–559.
- Buchner, A., E. Erdfelder, and F. Faul. 1997. How to Use G*Power. Available: http://www.psych.uni-duesseldorf.de/aap/projects/gpower/how_to_use_gpower.html. Accessed Aug. 12, 2003.
- Buhl-Mortensen, L. 1996. Type-II statistical errors in environmental science and the precautionary principle. *Mar. Pollut. Bull.* 32:528–531.
- Cohen, J. 1962. The statistical power of abnormal-social psychological research: A review. *J. Abnorm. Soc. Psychol.* 65:145–153.
- Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Lawrence Erlbaum Associates, Hillsdale, N.J.
- Daghir, N. J., M. T. Farran, G. W. Barbour, and M. M. Beck. 2003. Nutritive value of high-oil corn grown under semi-arid conditions and its impact on broiler performance and carcass composition. *Poult. Sci.* 82:267–271.
- Del Carmen, J., A. G. Gernat, R. Myhrman, and L. B. Carew. 1999. Evaluation of raw and heated velvet beans (*Mucuna pruriens*) as feed ingredients for broilers. *Poult. Sci.* 78:866–872.
- Donkoh, A., C. C. Atuahene, Y. B. Poku-Prempeh, and I. G. Twum. 1999. The nutritive value of chaya meal (*Cnidoscolus aconitifolius* (Mill.) Johnston): studies with broiler chickens. *Anim. Feed Sci. Technol.* 77:163–172.
- Douglas, M. W., and C. M. Parsons. 2000. Effect of presolvent extraction processing method on the nutritional value of soybean meal for chicks. *Poult. Sci.* 79:1623–1626.
- Douglas, M. W., C. M. Parsons, and T. Hymowitz. 1999. Nutritional evaluation of lectin-free soybeans for poultry. *Poult. Sci.* 78:91–95.
- Erdfelder, E. 1984. Zur Bedeutung und Kontrolle des β -Fehlers bei der inferenzstatistischen Prüfung log-linearer Modelle. [On significance and control of the β error in statistical tests of log-linear models]. *Z. Sozialpsychol.* 15:18–32.
- Fernandez, J. I. M., F. R. Lima, C. X. Mendonca, I. Mabe, R. Albuquerque, and P. M. Leal. 1999. Relative bioavailability of phosphorus in feed and agricultural phosphates for poultry. *Poult. Sci.* 78:1729–1736.
- Fox, N., and N. Mathers. 1997. Empowering research: Statistical power in general practice research. *Fam. Pract.* 14:324–329.
- Gaines, A. M., G. L. Allee, and B. W. Ratliff. 2001. Nutritional evaluation of *Bt* (MON810) and Roundup Ready® corn compared with commercial hybrids in broilers. *Poult. Sci.* 80(Suppl. 1):51. (Abstr.)
- Gill, J. L. 1981. Evolution of statistical design and analysis of experiments. *J. Dairy Sci.* 64:1494–1519.
- Gonzalez-Esquerria, R., and S. Leeson. 2000. Studies on the metabolizable energy content of ground full-fat flaxseed fed in mash, pellet, and crumbled diets assayed with birds of different ages. *Poult. Sci.* 79:1603–1607.
- Hall, L. E., R. B. Shirley, R. I. Bakalli, S. E. Aggrey, G. M. Pesti and H. M. Edwards, Jr. 2003. Power of two methods for the estimation of bone ash of broilers. *Poult. Sci.* 82:414–418.
- Halle, I., K. Aulrich, and G. Flachowsky. 1999. Einsatz von Maiskörnern der Sorte Cesar und des gentechnisch veränderten *Bt*-Hybriden in der Broiler mast. Pages 265–267 in Proc. 5. Tagung, Schweine- und Geflügelernährung, Wittenberg, Deutschland.
- Hammond, B. G., J. L. Vicini, G. F. Hartnell, M. W. Naylor, C. D. Knight, E. H. Robinson, R. L. Fuchs and S. R. Padgett. 1996. The feeding value of soybeans fed to rats, chickens, catfish and dairy cattle is not altered by genetic incorporation of glyphosate tolerance. *J. Nutr.* 126:717–727.
- Hoening, J. M., and D. M. Heisey. 2001. The abuse of power: The pervasive fallacy of power calculations for data analysis. *Amer. Stat.* 55:19–24.
- Humphrey, B. D., N. Huang and K. C. Klasing. 2002. Rice expressing lactoferrin and lysozyme has antibiotic-like properties when fed to chicks. *J. Nutr.* 132:1214–1218.
- Kan, C. A., H. A. J. Versteegh, T. G. Uijtengogarrt, H. G. M. Reimert, and G. F. Hartnell. 2001. Comparison of broiler performance when fed *Bt* parental/isogenic control or commercial varieties of dehulled soybean meal. Pages 19–22 in Proc. Int. Symp. on Genetically Modified Crops and Co-Products as Feeds for Livestock, Nitra, Slovak Republic.
- Lalouel, J., and A. Rohrwasser. 2002. Power and replication in case-control studies. *Am. J. Hypertens.* 15:201–205.
- Lee, B. D., D. J. Kim, and S. J. Lee. 2001. Nutritive and economic values of high oil corn in layer diet. *Poult. Sci.* 80:1527–1534.
- Leeson, S. 1998. The effect of corn hybrid CBH35 on the growth of male broiler chickens. Dept. of Animal and Poultry Science, Arkel Research Farms, Univ. of Guelph, Guelph, ON, Canada. Lab project ID No. C–2–98. April 20, 1998.
- Lenth, R. V. 2001. Some practical guidelines for effective sample-size determination. *Am. Stat.* 55:187–193.
- Li, Y. C., D. R. Ledoux, T. L. Veum, V. Raboy, and D. S. Ertl. 2000. Effects of low phytic acid corn on phosphorus utilization, performance, and bone mineralization in broiler chicks. *Poult. Sci.* 79:1444–1450.
- Mireles, A., S. Kim, R. Thompson and B. Amundsen. 2000. GMO (*Bt*) corn is similar in composition and nutrient availability to broilers as non-GMO corn. *Poult. Sci.* 79(Suppl. 1):65. (Abstr.)
- Novak, C., and S. E. Scheideler. 2001. Long-term effects of feeding flaxseed-based diets. 1. Egg production measurements, components, and eggshell quality in two strains of laying hens. *Poult. Sci.* 80:1480–1489.
- OECD. 2003. Considerations for the safety assessment of animal feedstuffs derived from genetically modified plants. Series on the Safety of Novel Foods and Feeds. No. 9. Org. for Econ. Coop. and Dev., Paris, France.
- Parsons, C. M., F. Castanon, and Y. Han. 1997. Protein and amino acid quality of meat and bone meal. *Poult. Sci.* 76:361–368.
- Perez, J. F., A. G. Gernat, and J. G. Murillo. 2000. The effect of different levels of palm kernel meal in layer diets. *Poult. Sci.* 79:77–79.
- Piva, G., M. Morlacchini, A. Pietri, F. Rossi, and A. Prandini. 2001. Growth performance of broilers fed insect-protected (MON 810) or near isogenic control corn. *Poult. Sci.* 80(Suppl. 1):320. (Abstr.)
- Ravindran, V., R. L. Hood, R. J. Gill, C. R. Kneale, and W. L. Bryden. 1996a. Nutritional evaluation of grain amaranth (*Amaranthus hypochondriacus*) in broiler diets. *Anim. Feed Sci. Technol.* 63:323–331.
- Ravindran, V., R. Sivakanesan, and H. W. Cyril. 1996b. Nutritive value of raw and processed colocasia (*Colocasia esculenta*) corn meal for poultry. *Anim. Feed Sci. Technol.* 57:335–345.
- Rosenfeld, D. J., A. G. Gernat, J. D. Marcano, J. G. Murillo, G. H. Lopez, and J. A. Flores. 1997. The effect of using different levels of shrimp meal in broiler diets. *Poult. Sci.* 76:581–587.
- Rossi, J. S. 1990. Statistical power of psychological research: What have we gained in 20 years? *J. Consult. Clin. Psychol.* 58:646–656.
- Scott, T. A., F. G. Silversides, H. L. Classen, M. L. Swift, M. R. Bedford, and J. W. Hall. 1998. Broiler chick bioassay for measuring the feeding value of wheat and barley in complete diets. *Poult. Sci.* 77:449–455.
- Sheppard, C.R. 1999. How large should my sample be? Some quick guides to sample size and the power of tests. *Mar. Pollut. Bull.* 38:439–447.
- Sidhu, R. S., B. G. Hammond, R. L. Fuchs, J.-N. Mutz, L. R. Holden, B. George, and T. Olson. 2000. Glyphosate-tolerant corn: The composition and feeding value of grain from glyphosate-tolerant corn is equivalent to that of conventional corn (*Zea mays* L.). *J. Agric. Food Chem.* 48:2305–2312.
- Slominski, B. A., J. Simbaya, L. D. Campbell, G. Rakow, and W. Guenter. 1999. Nutritive value for broilers of meals derived from newly developed varieties of yellow-seeded canola. *Anim. Feed Sci. Technol.* 78:249–262.
- Steel, R. G. D., J. H. Torrie, and D. A. Dickey. 1997. *Principles and Procedures of Statistics. A Biometrical Approach*. 3rd ed. McGraw-Hill Book Co., New York.

- Sterling, K. G., E. F. Costa, M. H. Henry, G. M. Pesti, and R. I. Bakalli. 2002. Responses of broiler chickens to cottonseed- and soybean meal-based diets at several protein levels. *Poult. Sci.* 81:217–226.
- Svihus, B., and M. Gullord, 2002. Effect of chemical content and physical characteristics on nutritional value of wheat, barley and oats for poultry. *Anim. Feed Sci. Technol.* 102:71–92.
- Taylor, M. L., G. F. Hartnell, M. A. Nemeth, B. George and J. D. Astwood, 2001a. Comparison of broiler performance when fed diets containing Roundup Ready corn event NK603, parental line or commercial corn. *Poult. Sci.* 80(Suppl. 1):320. (Abstr.)
- Taylor, M. L., G. F. Hartnell, M. A. Nemeth, B. George, and J. D. Astwood. 2001b. Comparison of broiler performance when fed diets containing YieldGard corn, YieldGard and Roundup Ready® corn, parental lines, or commercial corn. *Poult. Sci.* 80(Suppl. 1):319. (Abstr.)
- Taylor, M. L., G. F. Hartnell, S. G., Riordan, M. A., Nemeth, K. Karunanandaa, B. George, and J. D. Astwood. 2003a. Comparison of broiler performance when fed diets containing grain from Roundup Ready (NK603), YieldGard × Roundup Ready (MON810 × NK603), non-transgenic control, or commercial corn. *Poult. Sci.* 82:443–453.
- Taylor, M. L., G. F. Hartnell, S. G. Roirdan, M. A. Nemeth, K. Karunanandaa, B. George, and J. D. Astwood. 2003b. Comparison of broiler performance when fed diets containing grain from YieldGard (MON810), YieldGard × Roundup Ready (GA21), nontransgenic control, or commercial corn. *Poult. Sci.* 82:823–830.
- Thomas, L., and C. J. Krebs. 1997. A review of statistical power analysis software. *Bull. Ecol. Soc. Am.* 78:126–139
- Wang, X., and C. M. Parsons. 1997. Effect of processing systems on protein quality of feather meals and hog hair meals. *Poult. Sci.* 76:491–496.