Carfentanil and Overwinter Survival in Bison: The Alternative Hypothesis

Authors: Halverson, Timothy G., and Teare, J. Andrew

Source: Journal of Wildlife Diseases, 25(3) : 448-450

Published By: Wildlife Disease Association

URL: https://doi.org/10.7589/0090-3558-25.3.448

Journal of Wildlife Diseases, 25(3), 1989, pp. 448–450

LETTER TO THE EDITOR . . .

Carfentanil and Overwinter Survival in Bison: The Alternative Hypothesis

A recent paper by Berger and Kock (1988) on overwinter survival of bison (*Bison bison*) following immobilization with carfentanil attracted the attention of one of us (JAT) because of a fundamental error in the statistical inference which led to the conclusion presented in the paper. Further consideration of the paper uncovered additional flaws in statistical methodology. We felt compelled to bring these problems to the attention of the editor and readers of the *Journal of Wildlife Diseases* in part because these errors are indicative of a widespread misunderstanding of some basic principles of statistical hypothesis testing, and in part because there is a danger that the unfounded conclusion will misdirect the decisions of wildlife managers working in the field.

The conclusion that ". . . carfentanil had no long-term negative effect on the survival of bull bison . . ." was inferred from the lack of statistical significance of a Chi-square test applied to re-sighting records of 21 immobilized bulls and 30 non-immobilized bulls. In fact, Berger and Kock (1988) have no basis for making this statement and have made the elementary error of accepting the null hypothesis when they merely failed to reject the null hypothesis. In statistical terminology they have failed to protect themselves sufficiently against committing a Type II error (the error of accepting a false null hypothesis). In general Type II errors are ignored in biological research and the emphasis is placed on Type I errors (the error of rejecting a true null hypothesis). This singular emphasis on Type I error rates is acceptable in studies which demonstrate differences among treatment groups (most published work) but it is entirely inappropriate for studies such as that by Berger and Kock (1988) which reports equivalence between experimental treatments. We will illustrate this point below with a careful examination of the data presented by Berger and Kock (1988).

Our first discovery upon examination of the data was that we were unable to arrive at the Chi Square value of 0.17 which Berger and Kock (1988) presented in their paper. Nowhere did they state the method they used to calculate their Chi-square value but we assume they used a contingency table to produce expected values and then calculated a continuity adjusted Chi-square value. (Their value of 0.17 is close to our calculation of 0.178 using this method.) In actuality, however, the Chi-square test is not valid for these data since more than 20% of the contingency table cells have expected values less than 5 (Sokal and Rohlf, 1973).

A more appropriate test for these data is Fisher's Exact Test which calculates exact probabilities of observed frequencies under the null hypothesis that the probability of re-sighting individuals is the same for the two study groups; or see Rice (1988) for a more powerful alternative to Fisher's Exact Test. Since the alternative hypothesis is a lower re-sighting probability for immobilized individuals (i.e., we are not interested in the possibility that immobilization increases survival) we calculated a one-tailed Fisher's Exact Test and arrived at a probability of $P = 0.33$. Thus, using the customary $\alpha = 0.05$ Type I error rate we failed to reject the null hypothesis as did Berger and Kock (1988) who reported "$0.90 > P > 0.75$." However, rather than simply accepting the null hypothesis at this point, it must be demonstrated that our statistical test was sensitive enough to find non-trivial differences if they existed. This

was the crucial step that Berger and Kock (1988) failed to take and it involves examination of β, the Type II error rate.

Type II error rates are a function of the true differences between groups being compared. We will fail to reject a false null hypothesis frequently if the differences between groups are small but we will make this error less frequently if the differences are large. Obviously we do not know the true difference in survival rates of carfentanil-immobilized and control bison or there would have been no need for the Berger and Kock (1988) study. Nevertheless, it is instructive to examine β assuming that the true probabilities of overwinter loss are accurately estimated by Berger and Kock's samples. Thus if the true probability of overwinter loss are $q = 0.067$ (two of 30 animals) for the control group and $q = 0.143$ (three of 21 animals) for the carfentanil group, then following the method of Sokal and Rohlf (1973) for binomial frequency distributions the Type II error rate for a sample of 21 animals is $\beta = 0.65$. In other words, if the probability of overwinter loss of bison following the administration of carfentanil increases from 0.067 to 0.143, then a sample of 21 animals will fail to demonstrate a significant difference 65% of the time.

This difference is not trivial. It represents more than a two-fold increase in the risk of death. Yet it is too small to be detected reliably given samples of 21 animals. We may now ask what would be required in order to be reasonably certain of finding statistical significance if a real difference between the two groups exists. We present two illustrative answers to this question. If the increase in risk of overwinter loss due to immobilization with carfentanil follows the calculations given above, then we would need control and carfentanil groups of 342 animals each to be 95% sure of finding that difference (i.e., $\beta = 0.05$). Alternatively, given a sample size of 21 animals, we can be 95% sure of detecting a difference only if the true risk of overwinter loss following immobilization with carfentanil is around 0.30 (almost a five-fold increase). Thus Berger and Kock (1988) would have had to have lost six animals from their carfentanil group before they could have claimed a significant difference in survival using Fisher's Exact Test, or seven animals using the inappropriate continuity adjusted Chi-square.

In short, the data provided by Berger and Kock (1988) do not allow us to accept or reject the null hypothesis with any degree of certainty. The final conclusion has to be that statistical hypothesis testing cannot resolve the question until more data are available.

An alternative approach to hypothesis testing, and one preferred by many statisticians, is the estimation of parameters and their confidence limits. This method usually provides more biologically meaningful information than the statistical asterisk. The data of Berger and Kock (1988) are well suited to the estimation of relative risk ratios, a technique commonly used in epidemiological studies to evaluate risks associated with various environmental or genetic factors for the onset of specific diseases. Using the FREQ procedure available in SAS (SAS Institute, Inc., 1985) we estimated the relative risk of overwinter loss among animals immobilized with carfentanil to be 2.14 with 95% confidence limits of 0.39 to 11.73. The extreme width of this confidence interval is an indication of unreliability of the estimate and it reflects the ambiguity of the earlier attempts at hypothesis testing. Nevertheless, in the absence of additional information, the best estimate of the effects of immobilization with carfentanil is a slightly more than two-fold increase in the risk of overwinter loss among male bison. In addition we can conclude with 95% certainty that the true risk ratio is between 0.39 and 11.73. While this does not provide definitive conclusions on the benefits or dangers of carfentanil immobilization, it does give wildlife managers unbiased estimates from which to work.

Berger and Kock (1988) have presented

some valuable and unique data on the immobilization of bison with carfentanil. It is unfortunate that errors in statistical methodology and interpretation were not recognized during the review and editorial process. The unsupported statement that ". . . carfentanil had no long-term negative effect on survival of bull bison . . ." has now appeared in a reputable journal where it may have an impact on the decisions and actions of wildlife managers in excess of its true value. Authors, reviewers, and editors need to be more aware of the implications of statistical hypothesis testing in order to avoid such unwarranted conclusions. We encourage the inclusion of statisticians during the design of studies, the analysis of data, and the review of manuscripts in order to help eliminate such errors prior to publication.

LITERATURE CITED

BERGER, J., AND C. KOCK. 1988. Overwinter survival of carfentanil-immobilized male bison. Journal of Wildlife Diseases 24: 555–556.

RICE, W. R. 1988. A new probability model for determining exact p-values for 2 × 2 contingency tables when comparing binomial proportions. Biometrics 44: 1–22.

SAS INSTITUTE, INC. 1985. SAS user's guide: statistics, version 5 edition. SAS Institute, Inc., Cary, North Carolina, 956 pp.

SOKAL, R., AND F. ROHLF. 1973. Introduction to biostatistics. W. H. Freeman and Company, San Francisco, California, 368 pp.

Timothy G. Halverson[1] **and J. Andrew Teare,**[2] [1] Smith Kline and French Laboratories, Research and Development, P.O. Box 1539, King of Prussia, Pennsylvania 19406, USA; [2] Milwaukee County Zoo, 10001 W. Bluemound Rd., Milwaukee, Wisconsin 53226, USA.