FISH PASSAGE CENTER

1827 NE 44th Ave., Suite 240, Portland, OR 97213
Phone: (503) 230-4099 Fax: (503) 230-7559
http://www.fpc.org/
e-mail us at fpcstaff@fpc.org

<u>MEMORANDUM</u>

TO: Fish Passage Advisory Committee

Michele DeHart

FROM: Michele DeHart

DATE: June 8, 2011

RE: Performance Standard Testing, Hydroelectric project configuration testing, comparing operations

Recently the Fish Passage Center has been asked to review study results that are being used to inform important operations and fish passage management decisions. In these reviews, we have found that the statistical power of these tests were inadequate to detect expected differences, indicating a high probability of committing a type II error (failing to detect an effect). Most of the fish passage management questions center around which operation is better for juvenile fish passage, and these deliberations frequently focus on fish survival and whether or not there is a "significant" difference in fish survival between operations or conditions. Managers are often provided with study results that conclude that there is "no significant difference" or "no compelling difference" between operations or conditions. In our reviews of these studies and results and their management application, there is the serious and reoccurring problem that **the statistical power of the test is frequently inadequate to detect the expected effect and therefore does not provide sufficient support for the management decision. As a result, there is a high probability that an erroneous management decision will be made. Specifically, management decisions risk being based upon a conclusion of "no significant difference" when indeed an important difference exists.**

We offer the following points for your consideration to address and avoid this problem in the future. In particular, power analysis should be included *both* as part of study proposals and as part of study results. We have also provided some specific examples and discussion to further illustrate the technical issue.

- Study proposals should always include a power analyses as a basis for calculating the appropriate sample size and to determine what magnitude difference a study can test for across treatments. If this difference is particularly large, then conducting the study may not be useful.
- The power of the test should be considered and established **BEFORE** the study is conducted not after the results are completed. This avoids wasting an investment on research that cannot be utilized. If the sample size is not adequate, the study should not be conducted, because the results will be controversial and will not be applicable to a management decision.
- Power analysis should be done on all studies that show no difference between treatment conditions, such as contrasting spill levels or configurations
- The statistical power of the test should be evaluated and discussed in the study results and discussion. This should include reporting the minimum effect that could be detected by the test.
- The recent studies evaluating fish passage, project passage operations, some which are required by the Biological Opinion, fail to report β, the probability of making a type II error, or statistical power (1-β). In other words, the probability of not finding a statistical difference due to small sample size or large sampling variability, when a difference actually exists.
- In developing study proposals and study designs, the minimum and maximum expected effect or difference should be identified based upon past research.
- When results are provided to managers as a basis of management decisions, the power of the test to detect an effect should be identified to protect managers from erroneous conclusions and decisions. The minimum detectable effect of the test should be clearly identified in the results.
- The limits of the application of the study results should be clearly stated in study reports.

Power Analysis

Power analysis is important in designing an experiment and allows for a meaningful interpretation of non-significant results when the analysis is completed. (Thomas and Juanes 1996). Statistical power analysis can help interpret past results and improve designs of future experiments, impact assessment and management regulations (Peterman,1990). The power of a statistical test is the probability that the test will reject the null hypothesis when the null hypothesis is false. That means that the higher the power of a test the less chance there is that a Type II error or a false negative decision will occur. As the power of a test increases, the chances of making a Type II error decrease. The probability of a Type II error occurring is referred to as the false negative rate (β). Therefore power is equal to 1 − β, which is also known as the sensitivity. Power analysis is be used to calculate the minimum sample size required so that one can be reasonably likely to detect an effect of a given size. Power analysis can also be used to calculate the minimum effect size that is likely to be detected in a study using a given sample size. Statistical power is a function of sample variance, sample size, the level of statistical significance, and the effect size of interest. Calculating statistical power after a failure to reject the null hypothesis is an important second step in the analytical process surrounding tests of hypotheses (Parnell,2008).

Low power studies

Within the Columbia Basin, the problem of inadequate statistical power and the subsequent application of those results has been an issue in several studies: John Day Dam acoustic tag testing of 30% versus 40% spill proportions (2008, 2009, 2010), Lower Monumental testing of bulk versus flat spill configurations (2009), and McNary turbine operation range tests (2003). In these cases researchers and managers concluded that there was no statistical difference between treatments, when large sampling variability or small sample size could have precluded detection of differences. From Lower Monumental (2009) studies of fish survival and passage timing, the managers concluded that there was “no compelling” evidence that one spill configuration was better than the other. A power analyses conducted after the study was concluded showed that the study could only detect a difference of 5% or more (FPC memo, 4/25/2011). Historic research on the Columbia and Snake Rivers indicate that spill passage mortality is less than 2%. The study conducted at Lower Monumental, as designed could not detect a meaningful difference among the two configurations, leading managers to conclude that there was no difference when there may have been a difference. In 2003, McNary turbine efficiency operating ranges were a subject of research and management discussions. Increasing flow through turbine units, would result in turbine operations outside the efficiency range. A study was proposed that would detect a plus or minus 3% difference in fish survival at the two test conditions. However, the expected difference in survival was 1%. The proposed study would not be able to detect a difference because of small sample size. At John Day Dam acoustic tag studies were conducted to compare juvenile fish survival at two spill levels that were only 10% different. The issue of statistical power will also arise in evaluations of turbine operation range at Bonneville Dam, identified in the recent white paper, discussed at FPOM.

Peterman, R. M. 1990. Statistical power analysis can improve fisheries research and management. Canadian Journal of Fisheries and Aquatic Sciences 47, no. 1: 2–15.

Parnell, Ian. Statistical power analysis and Components of statistical power. Scientific Inquiry; suite 101.com, Nov.8,2008.

Thomas, L., and F. Juanes. 1996. The importance of statistical power analysis: an example from Animal Behavior. Animal Behavior 52, no. 4: 856–859.