OPTIMIZING LIMITED MONITORING CAPACITY: APPLYING POWER ANALYSIS AS A MANAGEMENT TOOL FOR THE ECOLOGICAL INTEGRITY PROGRAM AT POINT PELEE NATIONAL PARK

Paul Zorn[1], Tom Linke[2], Matthew Smith[2], Valerie Blazeski[2]

[1] Parks Canada. Ontario Service Centre. Ecosystem Conservation Section. Ottawa, ON. K1A 0M5. tel: Email: paul.zorn@pc.gc.ca
[2] Point Pelee National Park. Leamington, ON N8H 3V4 Email: tom.linke@pc.gc.ca, matthew.smith@pc.gc.ca, valerie.blazeski@pc.gc.ca

SUMMARY

Two common questions every manager has of their monitoring program is "How much data do I need and how much will it cost?" At Point Pelee National Park power analysis was systematically applied to indicators within their ecological integrity monitoring program to address these two questions. Through power analysis, sample sizes required to detect monitoring trends of certain sizes, with specific probabilities of making an error, given the variability associated with each indicator, were estimated. Ranges of sample sizes associated with a corresponding range of risk of error was calculated for each indicator. The cost to implement the monitoring protocols for these indicators were also estimated and translated into a cost per unit within the sample. With this information scenarios were developed that identified the current risk of making an error, the sample sizes needed to sufficiently reduce this risk, and the associated cost. This power analysis approach to determining appropriate monitoring effort should be of keen interest to managers as it provides a tool to integrate the science of monitoring with the business of monitoring. Results provide managers with specific cost estimates need to implement a successful monitoring program which can be used to forecast budgetary needs, staffing levels and work plans.

1. INTRODUCTION

A common question that every manager has of their monitoring program is "how much data do I need?". A manager usually has two major considerations when asking this question. The first is that the manager wants to ensure that monitoring will provide the information needed to determine if goals and objectives are being met. The second is that the manager needs to be fiscally responsible and so the monitoring program must provide the right kind of information in a timely fashion while being as cost-effective as possible.

The answer to this question, of course, is "it depends". Specifically, it depends on three factors: 1. the minimum trend that the manager wants to be able to detect through monitoring, 2. the amount of variability or "noise" in the thing that is monitored, and 3. the probability or level of risk the manager wants to accept in the monitoring program being wrong. The process by which these three factors become quantified in order to answer the question "how much data do I need" is called power analysis.

Power analysis is a group of statistical methods that is related to confidence level testing. It quantifies the relationship between the probability of making a Type I error (confidence level – this is the probability of concluding that a significant trend has occurred when it hasn't – a false alarm), the probability of making a Type II error (sample power – this is the probability of concluding that no significant trend has occurred when it has – for monitoring this is the most serious kind of error since the monitoring program has failed to detect the kind of trend that it was designed for), the variability in the measure that is monitored, the minimum detectable trend, and sample size (see Table 1 Tables). Through power analysis any part of this relationship can be solved for if the other parts are known.

To address the question "how much data do I need?" sample size is unknown. Therefore, to solve for sample size, estimates are needed for Type I error, Type II error, variability and minimum detectable trend. In addition to these estimates the method of statistical analysis (e.g., linear regression, ANOVA, chi square) appropriate for the monitoring program's objectives must also be known since the method of power analysis differs with statistical method.

Estimates for confidence level and sample power are determined by the manager and are generally determined by how risk adverse the manager wants to be. A 95% confidence level is thought of as a standard for scientific publication but is generally thought of as being inappropriate for management (1). This is true for two reasons. Large sample sizes are usually required to reach a 95% confidence level. These large sample sizes are typically expensive and require a monitoring effort that is beyond most organizations to sustain over the long term. Second, if a great deal of time is spent through monitoring to collect enough data for a manager to be 95% confident that a problem actually does exist the opportunity to do anything about it in a proactive manner may be lost. A publication standard for sample power is typically 80% (although most scientific journals only require that confidence levels, and not sample power, be included with published results and so sample power is usually less well known). Since the ramifications of a Type II error are more severe for management the sample power of a monitoring program should be of keen interest to all managers.

There has been some recent debate as to the appropriate monitoring standards for confidence level and sample power. Several authors set a goal of 80% confidence level (instead of the usual 95%) and 90% sample power (instead of the usual 80%) (2,3). This is not a rule. For many managers a 70% probability, for example, that a problem may exist may be sufficient to warrant action. Depending on what is being monitored and the consequences of not detecting a negative trend when it occurs, the manager may chose to accept varying levels of confidence and power. As a general pattern, however, managers should demand a higher sample power over confidence level due to the higher consequences of making a Type II error relative to a Type I error.

Estimates for variability and minimum detectable trend are usually based on at least one of the following sources of information: baseline monitoring data, a pilot project, published studies, or expert opinion. Often the first five years of a monitoring program are viewed as the "pilot" stage. Results from this pilot stage can be used as the main source of information from which estimates of variability are derived. The minimum size of trend is another measure for which there is no set standard. The minimum size of trend is generally determined by the goals of the monitoring program and the ecology of indicator being monitored. As a rule of thumb the minimum detectable trend should be no larger than the average amount of variation across sampling periods (i.e., one standard deviation about the mean calculated from baseline data).

Once estimates (or a range of estimates) are determined for the parameters listed in table 1, and once the method of statistical analysis needed to address the monitoring question is selected, then the process of power analysis can continue. The output that power analysis provides can be expressed as a "power curve" (see Figure 1). A power curve shows the relationship between the sample size needed to achieve the desired level of sample power given the parameters.

Managers should be concerned about distinguishing between three zones within the power curve. A monitoring program's indicator that has a sample size that lies within "Zone A" of the curve means that this indicator has low power and a high probability of making a Type II error. In these situations there is a low return on the investment made into monitoring because the data generated is likely to be insufficient to determine if goals and objectives are being met. Indicators that lie within "Zone B" represents the most efficient use of monitoring resources. Indicators within this zone have sufficient sample sizes to detect trends with high power but are not over sampled. Indicators within "Zone C" are effective in that they have high power but they are not efficient because the same trends can be detected with smaller sample sizes. The additional cost required to increase sample sizes within this zone do not translate into increased power. This is a problem given that the inefficiently used resources from Zone C can be better used for indicators that lie within Zone A.

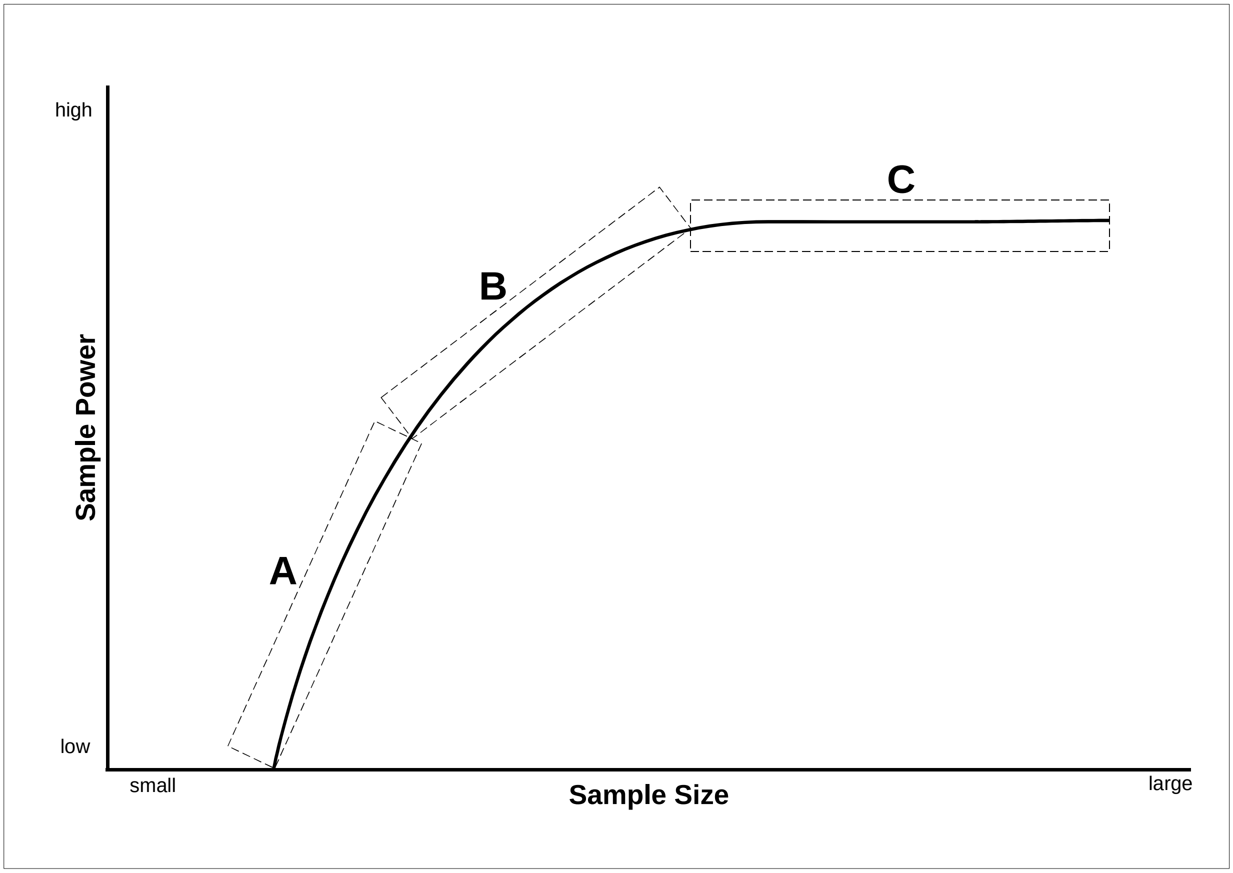

Figure 1. Generic power curve that depicts the relationship between sample size and statistical power. Zone A indicates a situation where a monitoring program has low power due to small sample size. Zone B indicates an optimal situation where a monitoring program has high power and efficient use of resources. Zone C indicates a situation where a monitoring program has high power but resources are being used inefficiently.

For every monitoring program a sample size can be potentially reached where the sample accurately reflects the true trend in the indicator. If sample sizes are increased beyond this point the true trend does not change and so power is not improved. Also for every indicator there will be a range of sample sizes that are too small to accurately reflect trends and so will have low power. Given this, every indicator within a monitoring program will have a power curve with a shape similar to that in figure 1 and each curve will have a distinct Zone A, B and C. Where ever possible managers should strive to invest in their monitoring programs such that indicators have sample sizes that lie within Zone B.

2. ECOLOGICAL INTEGRITY MONITORING AT POINT PELEE NATIONAL PARK

Power analysis was conducted for the ecological integrity monitoring program at Point Pelee National Park. Individual indicators were selected from the program based on the availability of baseline data (see Table 2 Tables). For these indicators power analysis was conducted in order to address the following questions: 1. Are data currently collected for these indicators sufficient to detect trends?, and 2. How should the park adjust its level of investment in its monitoring program to make these indicators more effective?

3. METHODS

Monitoring data for each indicator was consolidated and formatted using Microsoft Excel and then brought into SPSS for Windows 10.0. Descriptive statistics required for power analysis parameters were then calculated (see Table 3 for forest birds, Table 4 for marsh birds and Table 5 for all other indicators Tables). The protocols for each indicator were also reviewed and annual costs required for implementation were estimated. Costs included capital costs and staff time required for logistics, data collection and data input. These costs were then divided by the average annual sample size per indicator in order to derive an estimated "average annual cost per unit monitored". This measure was later used to estimate the required additional investment needed to improve the quality of the monitoring program for each indicator (see Results).

During the protocol reviews the appropriate method for statistical analysis was determined according to the stated monitoring question and the data types for each indicator. For situations where the dependent variable was a count, the independent variable was time and trends were analyzed using linear regression, Monitor 7.0 (4) was used to calculate power. For all other situations PASS 2000 (5) was used for power analysis. In total, power analysis was conducted for 68 separate species and the park's ability to detect significant trends through its ecological integrity monitoring program was assessed.

4. RESULTS

4.1 Identifying Sample Power Under the Current Monitoring Regime

The park's ability to detect population declines (10% decline over a 5 year period) in individual forest bird species is shown in Figure 2. Along this power curve, 32 species are in Zone A (low power), 10 species are in Zone B (high power, efficient use of resources. Species = AMCR, COGR, GCFL, OROR, EAKI, BHCO, MODO, GRCA, YBCU, BLJA) and 6 species are in Zone C (high power, inefficient use of resources. Species = INBU, REVI, YWAR, BAOR, EAWP, NOCA).

In terms of the park's ability to detect population declines (10% decline over a 5 year period) in marsh birds (see Figure 3) only one species, BARS, is in Zone B. All others are in Zone A.

Figure 4 shows the relative position of the other indicators along the power curve. Colonial waterbirds, eastern mole, and five-lined skink are in Zone A. Forest bird diversity (richness, Shannon index) are in Zone B and anuran richness is in Zone C (the minimum detectable trend for each of these indicators was set to ∀1 standard deviation about the mean calculated from available baseline data). Note the specific sample power (Y axis) for each indicator in figures 2, 3 and 4.

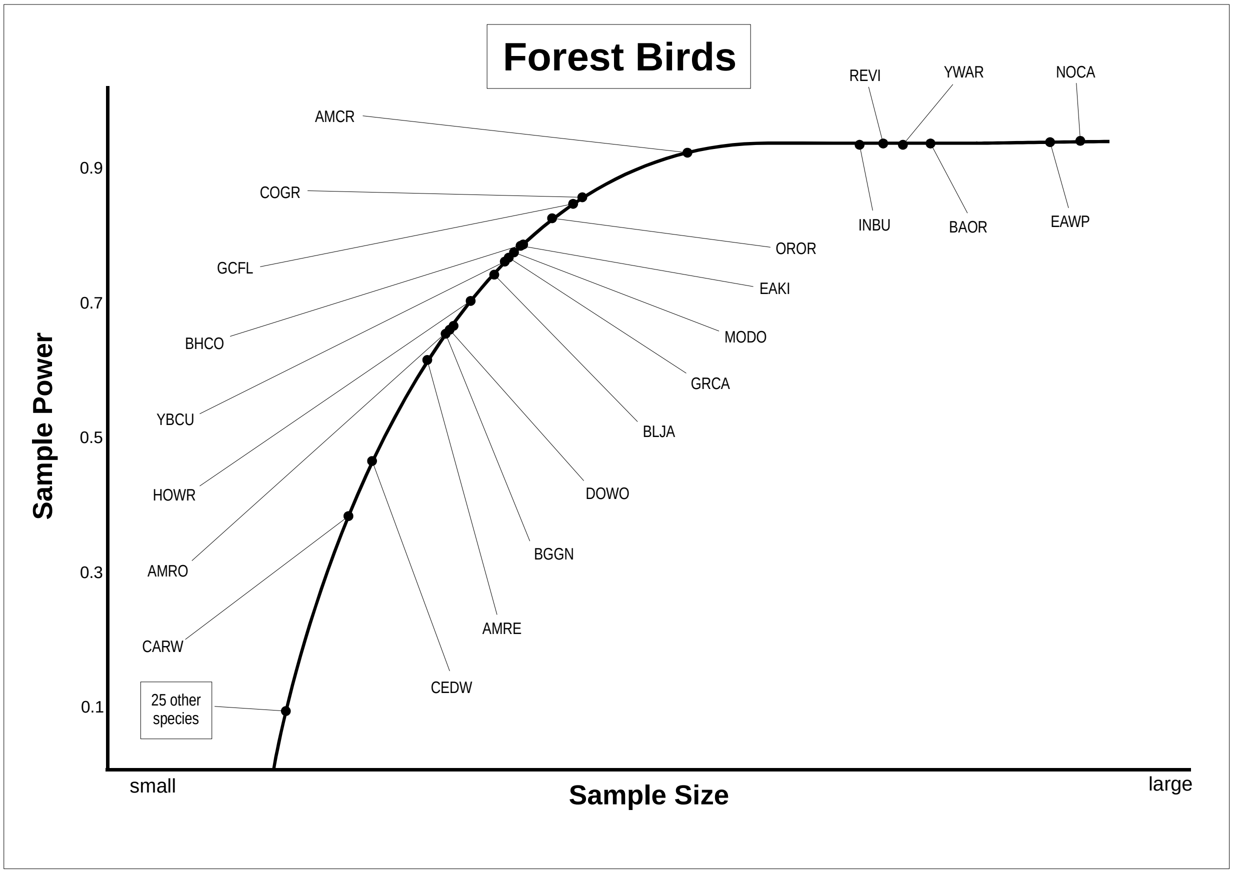

Figure 2. Power to detect a 10% population decline over 5 years among species monitored through Point Pelee National Park's current forest bird monitoring program]

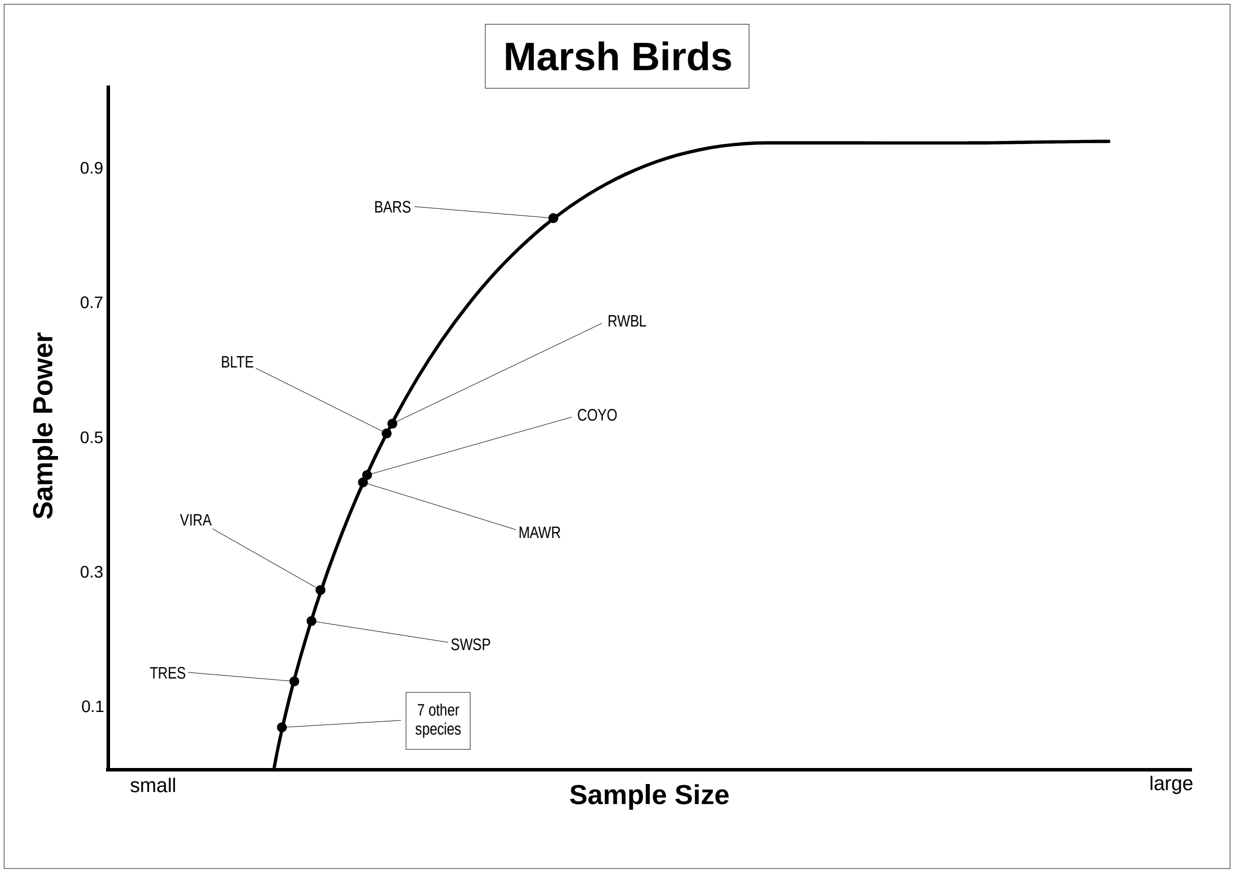

Figure 3. Power to detect a 10% population decline over 5 years among species monitored through Point Pelee National Park's current marsh bird monitoring program.]

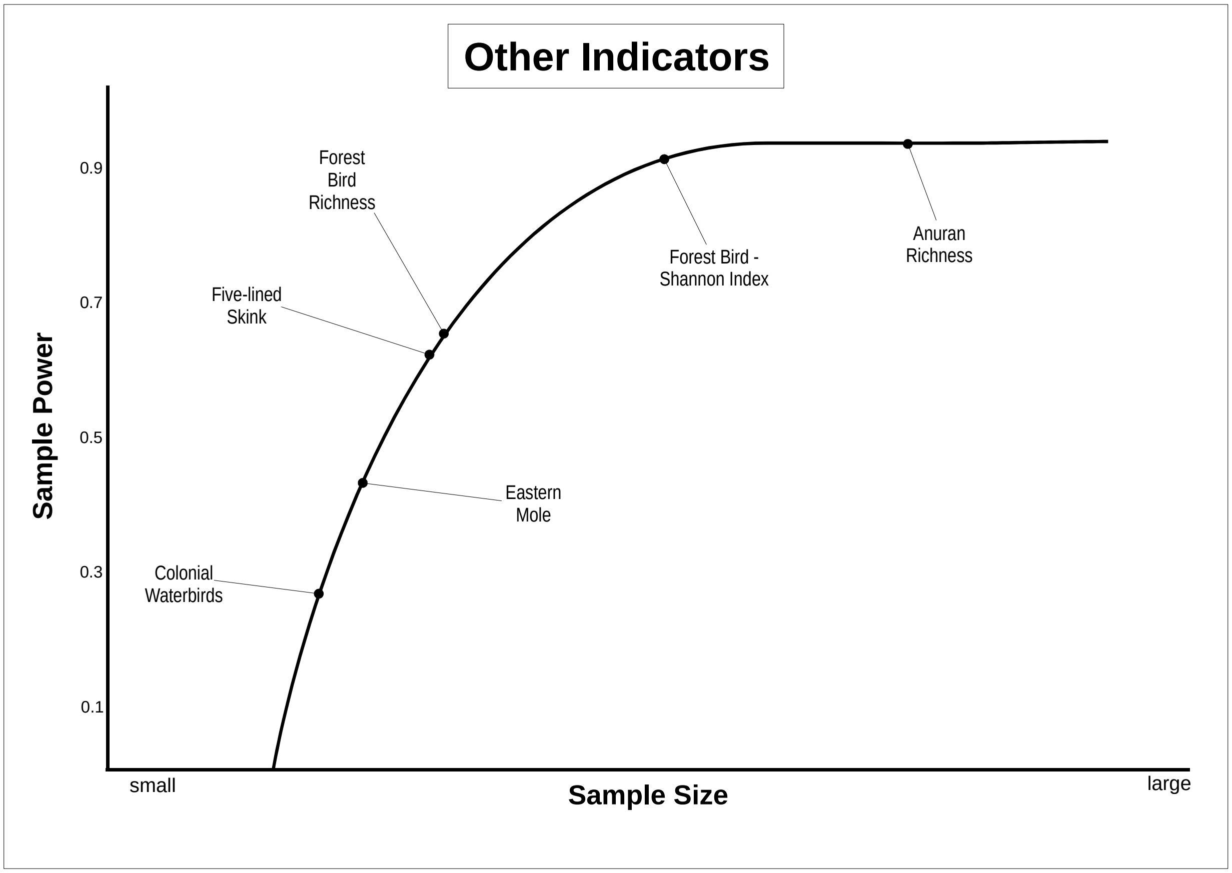

Figure 4. Power to detect trends among other indicators in Point Pelee National Park's ecological integrity monitoring program. The minimum detectable trend among these indicators vary depending on the standard deviation about the mean value per indicator calculated from baseline data.]

4.2 Determining the Costs for Improving Monitoring Program Effectiveness

For indicators that lie within Zone A or C on the power curve, various power analysis scenarios were generated to determine the sample size required to get that indicator into Zone B. The required sample size was then converted into dollars through each indicators' "average annual cost per unit monitored" value (see Table 6 Tables).

Sample sizes in anuran diversity, currently in Zone C, were scaled back through power analysis scenarios and it was estimated that high power in this indicator can still be achieved with 2/3 of the current monitoring effort, potentially saving the park $2000 annually.

Colonial waterbirds, in Zone A, could not be bumped up into Zone B under any power analysis scenario. The annual variability in this indicator (from baseline data) is extremely high such that low power is still only achieved with 10 times the sampling effort (which would cost $50,000 a year). Given this indicator's low power the monitoring information this program generates is suspect in terms of its ability to indicate population trends in colonial waterbirds. The annual cost of $5000 in this indicator can probably be more effectively spent on another indicator within the park's ecological integrity monitoring program.

The power to detect trends in eastern mole is estimated to be 45.2% based on the current sampling design of 6 sites monitored twice annually. Results from power analysis scenarios suggest that the addition of 14 more sites and one more survey occasion per year would provide 90% power. This assumes that the data generated from these 14 additional sites would be similar to the existing 6 sites located throughout the park. The additional annual cost of adding these monitoring sites would be approximately $6000.

The current monitoring regime for five-lined skinks, 9 monitoring sites surveyed once a year, provided 61.6% power. It is estimated that the addition of 7 more sites and one more survey occasion per year would improve the power of this indicator to 90%. The estimated annual cost to expand this program to this level is $2200.

For forest birds the current sampling design provided high power for all measures of species diversity (e.g., bootstrapped species richness, Shannon diversity index, Fisher's alpha, abundance coverage estimator). In terms of detecting population trends within individual species, however, the current monitoring program is currently recording enough occurrences to have 90% power for only 7 species representing 4 guilds. If just one more route was added results from power analysis indicates that 90% power could be achieved in 16 species representing 5 guilds. The annual cost to add this additional route is $1000.

For marsh birds, currently monitored from 2 survey routes, high power to detect population trends is achieved in only one species. From power analysis scenarios it is estimated that high power can be achieved in 5 species (representing 3 guilds) through the addition of 4 more routes. The estimated annual cost for this change is $6000.

5. DISCUSSION

Power analysis can be a very effective tool in integrating the science of monitoring with the business of monitoring. Through power analysis the level of risk a manager is willing to accept (expressed through the confidence level and sample power) and the minimum detectable trend must be explicitly stated. Something that is rare in most monitoring programs. This is useful information in its own right and can be communicated to stakeholders. Once power analysis parameters are estimated, usually from existing baseline data, multiple sampling scenarios can be quickly developed. These scenarios can address questions like: "how much data do I need to collect to meet my monitoring goals?", "what level of sample power can I achieve with 2, 4 or 6 survey routes?", "what sample size do I need for this indicator to achieve 90% power?". Once sample size estimates are determined the required monitoring effort can easily be converted into dollars and staff time. This information is vital to managers in forecasting budgets, staffing requirements and work plans.

Point Pelee National Park currently spends approximately $23,800 annually on monitoring the indicators discussed in this report. For this $23,800 annual investment only 2 indicators (anuran diversity and forest bird diversity) have high power. All other indicators are in Zone A on the power curve. Through power analysis it is estimated that an additional $8200 a year (see Table 6) could move all these indicators into Zone B (except for colonial waterbirds which would be discontinued). This seems to be a wise investment since limited value is currently being generated from the existing $23,800 monitoring program because of low power. Due to this low power it is unlikely that a manager would feel confident in using this existing monitoring information to base decisions upon because the risk of making an error (both Type I and Type II) is too high.

If an additional $8200 per year is beyond the park's capacity to afford over the long term a manager could use other supplementary criteria to guide the investment in monitoring. For example, forest birds, in additional to being selected as an ecological integrity indicator for the park, are of high social-political importance because of the economic opportunities bird watching generates in the area. Five-lined skinks are listed by COSEWIC as a species of special concern due to their rarity. If the cost savings from anuran diversity and colonial waterbirds (see Table 6) were realized the forest bird and five-lined skink indicators could be moved into Zone B on the power curve for an estimated $3200/year and the park would still save approximately $3800 annually.

Point Pelee National Park is currently one of the few national parks in Canada that is systematically applying power analysis in an evaluation of their entire ecological integrity monitoring program. This evaluation continues and is scheduled to address other indicators in the park's monitoring program. Through the application of power analysis Point Pelee National Park is realizing the following benefits:

- Past investments in collecting baseline monitoring data are being applied to influence management decisions.
- The level of risk (confidence level and sample power) and minimum detectable trend associated with each indicator in the monitoring program are made explicit.
- The method of statistical analysis is identified (which is not always the case in many existing monitoring protocols).
- Appropriate sample sizes and levels of monitoring effort are identified that are needed to generate information with a high level of utility for managers.
- The ongoing cost for this level of monitoring effort can be determined and used to set financial budgets, staffing levels and work plans.

Managers are responsible for fulfilling their organization's goals and objectives and they are often trying to do so with limited staff and budgets. In an ecological monitoring context, therefore, managers should become very interested in power analysis as a primary means for optimizing their limited capacity. Power analysis should become a central tool in any manager's monitoring toolbox as it provides them the opportunity to merge the science of monitoring with the business of monitoring.

REFERENCES

1. Johnson, D. H. 1999. The Insignificance of Statistical Significance Testing. Journal of Wildlife Management 63(3):763-772.
2. Droege, S., L. Monti, and D. Lantz. 1997. The Terrestrial Salamander Monitoring Program. Patuxent Wildlife Research Center. U.S. Geological Survey. http://www.mp2-pwrc.usgs.gov/sally/
3. Zorn, P., V. Blazeski and B. Craig. 2002. Joint EMAN / Parks Canada National Monitoring Protocol for Plethodontid Salamanders. Draft. Parks Canada. Ontario Service Centre. Ottawa, ON.
4. Gibbs, J. B. 1995. Monitor 7.0. Software for Estimating the Statistical Power of Population Monitoring Programs. U.S. Geological Survey. http://www.mbr-pwrc.usgs.gov/software/monitor.html
5. NCSS Inc. 2000. PASS 2000. Power Analysis and Sample Size Software. http://www.ncss.com/pass.html