TrendPowerTool: A lookup tool for estimating the statistical power of a monitoring program to detect population trends

Emily L. Weiser1 | Jay E. Diffendorfer2 | Laura Lopez-Hoffman3 | Darius Semmens2 | Wayne E. Thogmartin1

1U.S. Geological Survey, Upper Midwest Environmental Sciences Center, La Crosse, Wisconsin
2U.S. Geological Survey, Geosciences and Environmental Change Science Center, Denver, Colorado
3The University of Arizona, School of Natural Resources and the Environment and Udall Center for Studies in Public Policy, Tucson, Arizona

Abstract

A simulation-based power analysis can be used to estimate the sample sizes needed for a successful monitoring program, but requires technical expertise and sometimes extensive computing resources. We developed a web-based lookup app, called TrendPowerTool (https://www.usgs.gov/apps/TrendPowerTool/), to provide guidance for ecological monitoring programs when resources are not available for a simulation-based power analysis. TrendPowerTool is implemented through the shiny package in R, but is accessible through a webpage without the need for users to install any software. By drawing on results of 1.4 million scenarios that we simulated on a supercomputer, TrendPowerTool quickly and easily provides an estimate of the statistical power to detect a population trend of a particular magnitude with a planned monitoring program, based on user-specified parameters for the monitoring design and population of interest. TrendPowerTool provides a user-friendly interface that retrieves results instantaneously, facilitating the important step of conducting a power analysis when designing monitoring programs.

KEYWORDS
monitoring design, population monitoring, population trend, power analysis, status and trend, study design

1 | INTRODUCTION

Success of any scientific study or monitoring program hinges on the ability of the study to accurately quantify the attribute or relationship of interest. The majority of long-term monitoring programs are designed to detect change over time, which often involves measuring a population trend (Marsh & Trenham, 2008). Without sufficient statistical power, a monitoring program could fail to achieve its goals, and the resources expended would therefore be wasted (Legg & Nagy, 2006). If the monitoring program detects a trend when none is present (type I error), fails to detect a real population trend (type II error), or estimates a trend that is opposite to the one present (type III error; Condon, 1986), any management decisions based on the apparent population trend would be counterproductive to the management objective(s).
Scientists and statisticians have long recommended conducting a power analysis to inform robust study design (Gerrodette, 1987; Reynolds, 2012; Urquhart, 2012). For simple study systems and questions, a power analysis is straightforward. Lookup tables are available for some simple tests, such as t-tests. For more complex situations, such as a study that repeatedly monitors multiple sampling units ("sites"), a simulation-based approach to power analysis is more flexible and accurate (Gibbs & Melvin, 1997). However, performing computer simulations requires technical expertise, and evaluating a large number of scenarios can require a high-performance computing cluster. Designers of monitoring programs might not have the necessary resources to conduct a simulation-based power analysis. As a result, power analyses are often neglected for monitoring programs and other scientific studies.

We conducted a comprehensive simulation-based power analysis to provide generalized guidance for monitoring programs that will repeatedly monitor a resource at a fixed set of study sites. We assessed a range of values across 1.4 million scenarios, so our results can be applied to many monitoring programs. The results of our simulations are publicly available via a web-based application (TrendPowerTool) that allows users to look up the expected statistical power for a proposed monitoring program. TrendPowerTool is easy to use and can be used to quickly compare study designs to identify the most appropriate design for a new monitoring program.

## METHODS

### 2.1 Model description

For our simulation-based power analysis, design parameters included the sample sizes (number of study sites, \(N\), and number of years or other time steps, \(Y\)) and the overall trend that the study aims to detect (\(\tau\)) as a proportion of the overall mean value of the attribute of interest (e.g., population density or abundance; \(\mu\)). Population parameters included \(\mu\), variation among sites in the mean value (\(SD_\mu\)), variation among sites in the local population trend (\(SD_\tau\)), interannual variation around the expected mean in each time step for each site (\(CV_Y\)), and observation error around the true value (\(SD_{obs}\); Table 1). For simplicity, we assumed that observations would follow a normal distribution. When defining population parameters from pilot data or example datasets, we started by

| Symbol | Definition | Values tested |
|--------|------------|---------------|
| \(X\)  | Response variable, such as abundance or density. Transformed to approximate a normal distribution, and centered on a mean of 1 by dividing each observation by the overall mean. | Count per unit effort (time, area, or distance sampled) |
| \(Y\)  | Number of time steps (e.g., years) over which survey is conducted. | 5, 10, 15, 20, 25, 30, 40, 50 |
| \(N\)  | Number of sites sampled. | 10, 20, 30, 50, 100, 200, 300, 400, 500, 750, 1,000, 2000 |
| \(\tau\) | Overall population trend to detect: Proportional change in \(\mu\) per time step. Determined as per management goals. | ±0.03, 0.04, 0.05, 0.075, 0.10, 0.165 |
| \(\mu\) | Mean abundance across sites, calculated by averaging \(X\) across years within each site \(s\) to estimate the site-specific mean \(\mu_s\), then averaging across sites. Near 1 in most datasets unless data are skewed. | 1 |
| \(SD_\mu\) | Standard deviation of \(\mu\) among sites. | 0, 0.5, 1.0 |
| \(SD_\tau\) | \(SD\) across sites in the site-specific population trend, where site-specific trend is expressed as the annual proportional change in \(\mu_s\). | 0, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50 |
| \(CV_Y\) | Interannual variation, expressed as the detrended coefficient of variation among years in the site-specific population size (residuals/\(\mu_s\)), averaged across sites. | 0, 0.5, 1.0, 4.0, 8.0 |
| \(SD_{CV_Y}\) | \(SD\) across sites in the site-specific values of \(CV_Y\). | 0, 0.25, 0.50, 1.00, 3.00 |
| \(SD_{obs}\) | Observation error. Appropriate calculation may vary among study systems. | 0.10, 1.00, 2.00 |

*Calculated following Gibbs (2000).

*Because the population metric is centered to a mean of 1, results of the power analysis are the same regardless of whether trend is positive or negative.
transforming the data to approximate normality, if necessary. Then, we centered the observations to a mean of 1 by dividing the (transformed) observations by the mean. This step ensures that the magnitude or measurement units of the mean would not affect statistical power, which is helpful for producing results that can be generalized across systems. In addition, a given study design will have the same statistical power to detect a trend regardless of the direction of the population trend (positive or negative). Further details on how to calculate the input values are provided in TrendPowerTool (“Input Values” tab).

Our framework assumes that a specific set of study sites will be sampled repeatedly, once per time step. For many projects, the time step will be annual, but any other regular interval could be used. The framework does not include designs where the sampling interval varies among sites. We also assumed that a linear model was appropriate for describing population trends, following transformation of the data to normality. Our model did not explicitly account for imperfect detection or observation error other than random noise. If a study system violates these assumptions, TrendPowerTool might not produce accurate estimates of statistical power.

We simulated a range of expected values of each parameter to define statistical power across the parameter space that is relevant for long-term monitoring programs for wild animals or plants (Table 1). We simulated 1,425,600 scenarios to cover all possible combinations of the input values. For each scenario, we first simulated data based on the input values, then fit a linear model to estimate the trend in the simulated dataset (including a random effect of site on the trend estimate), and finally evaluated the proportion of 100 replicates in which the true trend was detected (i.e., the estimated trend was significantly different from zero, based on 95% CIs, and in the same direction as the true trend). Details of the structure and function of our simulation-based power analysis are provided by Weiser et al. (2019) and the computer code is archived online (Weiser, 2018). The simulations took a few weeks to run in parallel on a supercomputing cluster.

Ideally, we would provide predictions for statistical power given all possible input values, even those intermediate to the values that we simulated. However, preliminary tests indicated that if we fit a generalized linear model fitted to the results of our simulations, with input parameters as predictors and power as the response, the model produced predictions of statistical power that were unbiased but imprecise. Other methods of prediction such as regression trees provided better precision, but would not be able to make predictions from intermediate values. We therefore did not attempt to predict power for input values that we did not simulate. For parameters that have a strong effect on statistical power, such as the magnitude or among-site variation in the trend (Weiser et al., 2019), we simulated input values in small increments to provide the best predictive ability (Table 1). We evaluate the practical implications of choosing from a finite list of discrete values in the “Validation” section below.

### 2.2 | User interface

We used the shiny package (Chang et al., 2017) in R (R Core Team, 2019) to develop TrendPowerTool (https://www.usgs.gov/apps/TrendPowerTool/) to provide a web-based interface to the results of the 1.4 million scenarios we simulated. We designed the tool to allow users to look up scenarios of interest to estimate the statistical power of potential study designs, using only a web

| Dataset           | Source | Input values^b | SD_p | CV_Y | SD_CV_Y | SD_\tau |
|-------------------|--------|--------------|------|------|---------|---------|
| Steppe plants     | 1      | 27 (25)      | 0.28 (0.5) | 0.38 (0.5) | 0.12 (0) | 0.06 (0.05) |
| Kelp forest fish  | 2      | 17 (15)      | 0.12 (0)   | 0.29 (0.5) | 0.11 (0) | 0.05 (0.05) |
| Dabbling ducks    | 3      | 61 (50)      | 0.32 (0.5) | 0.48 (0.5) | 0.72 (0.5) | 0.01 (0) |
| Small mammals     | 4      | 32 (30)      | 0.54 (0.5) | 3.95 (4.0) | 1.03 (1.0) | 0.08 (0.05) |

^bDatasets were accessed via the Global Population Dynamics Database (NERC Centre for Population Biology at Imperial College, 2010). (Source) 1. Zachmann, Moffet, and Adler (2010), 2. Reed (2016), 3. U.S. Fish and Wildlife Service (2017), 4. Kaufman (2016).

^bThe values here were averaged across species, sites, and years in each dataset. The expected power for any given species within each dataset would depend on the species-specific values of variation. For these datasets, all scenarios assumed an observation error of SD_obs = 1.00, as information on observation error was not available.
browser and an internet connection (R and shiny do not need to be installed on the user’s computer).

2.3 Validation

We assessed the consequences of having only a limited set of input values available in TrendPowerTool using four publicly available datasets (Table 2). The datasets were selected based on data accessibility (public), availability of abundance data for >10 sites over >10 years, and taxonomic representation: density of sagebrush steppe plants in Idaho, USA (Zachmann et al., 2010), abundance of fish in kelp forests in California, USA (Reed, 2016), abundance of breeding dabbling ducks in North America (U.S. Fish and Wildlife Service, 2017), and abundance of small mammals in Kansas, USA (Kaufman, 2016). We used simulations to predict the statistical power of each dataset to detect trends that correspond to IUCN Red List criteria for the

**FIGURE 1** Screenshot of the web-based user interface for TrendPowerTool. The user can choose up to six scenarios to display simultaneously, with the input values recorded to the right of each plot. A full results table can be downloaded. Other pages of the tool provide guidance for evaluating whether TrendPowerTool is appropriate for a specific program and for calculating input values.
following reductions in population size over 10 years or 3 generations, whichever is longer, where the causes of the declines have not been mitigated (IUCN, 2012): Vulnerable = −30% (−4% per year), Endangered = −50% (−7.5% per year), Critically Endangered = −80%, (−16.5% per year). Here, we consider the simulated estimates to be the “true” estimates of statistical power. We then compared these estimates to the estimates available in TrendPowerTool, using the closest values available in the lookup menus to predict the statistical power to detect each trend. By comparing the TrendPowerTool estimates to the estimates simulated using the exact input values, we evaluated whether the options available in TrendPowerTool allowed accurate prediction of statistical power.

3 | RESULTS

3.1 | User interface

TrendPowerTool is available at http://www.usgs.gov/apps/TrendPowerTool (Figure 1). When a user of TrendPowerTool selects their input values, the app provides an estimate of statistical power with an associated standard error. The app can plot statistical power across a range of values of one parameter (e.g., number of sites), and additional scenarios can be added in additional plots. The app also provides an option to download a results table for the specified scenario(s).

The available input values cover a wide range of values likely to be relevant to monitoring programs (Table 1). When an intermediate value would be most appropriate for a given monitoring program, the user can elect to use the nearest available value, round up (for values of variance) or down (for sample size options) in a conservative approach, or test the values above and below the best estimate. Similarly, if some of the input values are not well known for the study system of interest, the user can quickly try a range of values by changing one input parameter at a time, and use the resulting range of expected power to select the best design for the monitoring program. The app provides detailed guidance on calculating the input values from pilot data, and users are referred to this paper for further details on the framework of the power analysis.

3.2 | Validation

Even when the exact input values for an example dataset were not available in TrendPowerTool, the app reported estimates of statistical power that closely matched the simulated results for each of the four example datasets (Figure 2). High interannual variation (small mammals) and short monitoring duration (kelp forest fish) likely prevented achieving the desired power in detecting all but the trends of largest magnitude in some of the example scenarios. Both the simulation approach and the TrendPowerTool approach assumed the same underlying structure for the model describing the system; i.e., this comparison does not evaluate whether the assumptions of TrendPowerTool are valid for the example datasets.

4 | DISCUSSION

A power analysis is a vital part of planning a robust study or monitoring program, but is often neglected (Legg & Nagy, 2006). TrendPowerTool provides a quick and easy way for planners to predict the statistical power of a proposed monitoring program. With no technical expertise required, TrendPowerTool will make these estimates available to conservation practitioners to help ensure the success of monitoring programs.

As with any power analysis, the estimates from TrendPowerTool will be valid only if the underlying assumptions are met. In particular, we assumed that a fixed set of sites would be sampled repeatedly during each time step, that any observation error (including imperfect detection) can be accurately represented by a constant term, that observed counts are normally distributed, and that a linear model is appropriate to describe...
the population trend. In situations that severely violate these assumptions, a robust power analysis would require additional data simulations rather than a quick estimate from TrendPowerTool. Even when assumptions are met, results from this tool should be considered approximate, especially when input parameters are not well defined by existing data. As monitoring commences and further data become available, a power analysis tailored specifically to the study system would provide a more accurate assessment of the necessary sample sizes to achieve the monitoring goal. In some cases, the R code used for our simulations (Weiser, 2018) could be modified as needed for a particular monitoring system. Other software such as the simr (Green & Macleod, 2016) or TrendNPS (Starcevich, 2018) R packages can also facilitate running full power analyses for users with some technical expertise.

If the underlying assumptions are met, TrendPowerTool provides a quick, user-friendly method for estimating statistical power for programs that aim to monitor population trend. TrendPowerTool thus makes power analysis a more accessible step in the process of designing monitoring programs.

ACKNOWLEDGMENTS
We thank Joel Putnam and Aaron Fox for developing the platform to host the shiny app for TrendPowerTool; T. Wilson, B. Verheijen, J. Lamb, B. Ross, and L. Rosen for comments on the app interface; and B. Gray and two anonymous reviewers for comments on a previous draft of this manuscript. This research used resources provided by the Core Science Analytics, Synthesis, & Libraries (CSASL) Advanced Research Computing (ARC) group at the U.S. Geological Survey. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

CONFLICT OF INTEREST
The authors declare no potential conflict of interest.

AUTHOR CONTRIBUTIONS
Emily L. Weiser and Wayne E. Thogmartin conceived the ideas; Emily L. Weiser developed TrendPowerTool, scripted and ran the underlying simulations, and led the writing of the manuscript; all authors contributed critically to the drafts of the tool and manuscript and gave final approval for publication.

ETHICS STATEMENT
This was a simulation study and did not require ethics approvals.

DATA AVAILABILITY STATEMENT
TrendPowerTool is available at http://www.usgs.gov/apps/TrendPowerTool. The computer code used to conduct the simulation-based power analysis is publicly archived (Weiser, 2018). The example datasets were accessed through the Global Population Dynamics Database (NERC Centre for Population Biology at Imperial College, 2010).

ORCID
Emily L. Weiser https://orcid.org/0000-0003-1598-659X

REFERENCES
Chang W., Cheng J., Allaire J. J., Xie Y., McPherson J. 2017. shiny: Web Application Framework for R. R package version 1.0.5. https://CRAN.R-project.org/package=shiny.
Condon, R. E. (1986). Type III error. Archives of Surgery, 121, 877–878.
Gerrodette, T. (1987). A power analysis for detecting trends. Ecology, 68, 1364–1372.
Gibbs, J. P. (2000). Monitoring populations. In L. Boitani & T. K. Fuller (Eds.), Research technique in animal ecology—Controversies and consequences (pp. 213–252). New York, NY: Columbia University Press.
Gibbs, J. P., & Melvin, S. M. (1997). Power to detect trends in waterbird abundance with call-response surveys. Journal of Wildlife Management, 61, 1262–1267.
Green, P., & Macleod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. Methods in Ecology and Evolution, 7, 493–498.
IUCN. 2012. IUCN Red List Categories and Criteria, version 3.1. www.iucnredlist.org.
Kaufman, D. W. (2016). CSM01 seasonal summary of numbers of small mammals on 14 LTER Traplines in prairie habitats, ongoing since 1981. Environmental Data Initiative. https://doi.org/10.6073/pasta/08e9c0ebf3727711dbac749416a60645b
Legg, C. J., & Nagy, L. (2006). Why most conservation monitoring is, but need not be, a waste of time. Journal of Environmental Management, 78, 194–199.
Marsh, D. M., & Trenham, P. C. (2008). Current trends in plant and animal population monitoring. Conservation Biology, 22, 647–655.
NERC Centre for Population Biology at Imperial College. 2010. The global population dynamics database version 2. http://www.sw.ic.ac.uk/cpb/cpb/gpdd.html.
R Core Team. (2019). R: A language and environment for statistical computing. Version 3.6.2. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/
Reed, D. C. (2016). SBC LTER: Reef: Kelp Forest Community dynamics: Fish abundance. Environmental Data Initiative. https://doi.org/10.6073/pasta/c3977b8fff17b302f12eb8b32642808a4
Reynolds, J. H. (2012). An overview of statistical considerations in long-term monitoring. In R. A. Gitzen, J. J. Millsbaugh, A. B. Cooper, & D. S. Licht (Eds.), Design and analysis of long-term ecological monitoring studies (pp. 23–53). Cambridge, UK: Cambridge University Press.
Starcevich, L. A., & Mitchell, J. (2018). TrendNPS: NPS trend analyses for complex survey designs. R package version 0.0.5.4.
U.S. Fish and Wildlife Service. 2017. Waterfowl breeding population and habitat survey: Traditional survey.

Urquhart, N. S. (2012). The role of monitoring design in detecting trend in long-term ecological monitoring studies. In R. A. Gitzen, J. J. Millsbaugh, A. B. Cooper, & D. S. Licht (Eds.), Design and analysis of long-term ecological monitoring studies (pp. 151–173). Cambridge, UK: Cambridge University Press.

Weiser, E. (2018). Simulation-based power analysis code. U.S. Geological Survey code release. https://doi.org/10.5066/P9SFUH2K

Weiser, E. L., Diffendorfer, J. E., López-Hoffman, L., Semmens, D., & Thogmartin, W. E. (2019). Consequences of ignoring spatial variation in population trend when conducting a power analysis. *Ecography, 42*, 836–844.

Zachmann, L., Moffet, C. A., & Adler, P. (2010). Mapped quadrats in sagebrush steppe: Long-term data for analyzing demographic rates and plant–plant interactions. *Ecology, 91*, 3427.

---

How to cite this article: Weiser EL, Diffendorfer JE, Lopez-Hoffman L, Semmens D, Thogmartin WE. TrendPowerTool: A lookup tool for estimating the statistical power of a monitoring program to detect population trends. *Conservation Science and Practice*. 2021;3:e445. https://doi.org/10.1111/csp2.445