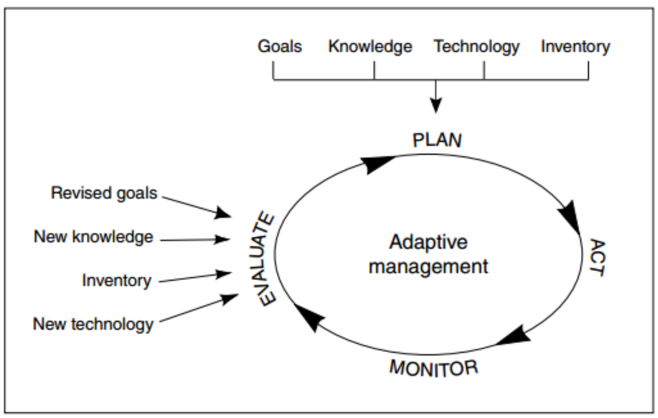
Monitoring concepts for managers: What, Why, When, How and Where

WHAT is Monitoring?

Monitoring is the process of tracking change or condition. In land management the change event is typically considered a “disturbance” that can be human caused, such as timber harvest, prescribed fire, mowing, herbicide application (often called “treatments”); or natural disturbance such as wildfire, a wind event or an ice storm.

Monitoring can take place at any scale. The Forest Service maintains the Forest Inventory and Analysis program to sample resources at the national scale. But land managers typically monitor at the project scale, which may be a few acres or thousands of acres depending on the goals of the project. The discussion that follows is most applicable to project scale monitoring.

Monitoring has a role in adaptive management. The Forest Service Handbook defines the adaptive management as - *A system of management practices based on clearly identified intended outcomes and monitoring to determine if management actions are meeting those outcomes; and, if not, to facilitate management changes that will best ensure that those outcomes are met or re-evaluated. Adaptive management stems from the recognition that knowledge about natural resource systems is sometimes uncertain.*



WHY Monitor?

Basically, there are three reasons: law, policy and professional knowledge. Numerous Federal laws require monitoring. The most commonly referenced is the National Environmental Policy Act (NEPA).

Here is the Council on Environmental Quality guidance for departments and agencies of the Federal government on the mitigation and monitoring of activities undertaken in a National Environmental Policy Act (NEPA) process (Feb. 18, 2010):

- *Under NEPA, a federal agency has a continuing duty to gather and evaluate new information relevant to the environmental impact of its actions.*
- *Agencies have the discretion to select the form and method for monitoring but should be sure to identify the monitoring area and establish the appropriate monitoring system.*
- *Implementation monitoring is designed to ensure that the mitigation measures are being performed as described in the NEPA documents and related decision documents.*

Also, monitoring is required by agency policy such as the Interagency Prescribed Fire Planning and Implementation Procedures Guide, Planning Rule and Cohesive Strategy.

And, hopefully, you will undertake monitoring because you want to learn more about the systems you are managing.

WHAT to monitor

Goals are broad descriptions of the desired outcome. Objectives should tier to the goals so that when objectives are met then goals are met. Goals and objectives will come from NEPA and other decision documents and are used to develop treatments (thinning, burning, herbicide…) – sort of answering the question, "What management tool(s) can I use to meet the goals of the project".

Use SMART Objectives: *Specific, Measurable, Achievable or Agreed, Realistic, Time-based*.

For example, assume there is a project goal to restore scrub jay habitat. A SMART objective may be: Increase the cover of live scrub oak trees at least 50% two years after prescribed fire. Many additional objectives will likely be used to assess whether the project goal was met.

Well thought out objectives are a critical component of a successful monitoring program.

HOW and WHEN to Monitor

The objectives will define the specific attributes to monitor (e.g., tree density and shrub cover) and will determine the field methods used to monitor them (fixed-area tree plot, line intercept). Method descriptions, recommended plot sizes, transect lengths, etc. can be found in sampling handbooks (see the References section below). The objectives will also help determine when to monitor and the time interval(s) between monitoring visits.

WHERE to Monitor

Random sampling is an important prerequisite for collecting data in a statistically valid design. A random sample simply means you are collecting data without bias. A *stratified random sample* is a sample design that divides a project area into smaller units (or polygons or strata) based on some attribute of interest like cover type or disturbance history. While a statistically valid sample is often preferred it is not required for monitoring. Some managers place plots with bias (where they want them) so they can monitor specific conditions or attributes.

Randomized sample designs for monitoring and research are similar but, typically, a research study has more observations (plots) per sample. Sample designs with many observations (i.e., sufficient to capture the variability of an attribute being tested) often allow investigators to note smaller statistically significant changes with higher confidence than when testing samples with a smaller number of observations. Small sample sizes (<30 observations) don't preclude statistical testing, but they will make it more difficult to note small statistically significant changes with high confidence. See the *Statistical Inference and Sample Size* section at the end if this handout for more information.

Include control plots, when possible, to help ensure changes are not due to non-treatment effects. For example, you might be looking at the influence of a treatment in reducing the cover of some species at two time periods. If you note a cover reduction on the treated plots, then you should also compare the change in cover of the same species on the control plots. Doing so will help you determine if the decrease in understory cover was due to a specific treatment effect (i.e., cover reduced only on the treated plots) or some other effect, like a dry growing season (i.e., cover reduced by the same amount on the treated plots and the control plots).

The top left image in the figure below shows a hypothetical project area divided into three units based on some characteristic such as cover type, treatment type, fuel loading, shrub cover, or distribution of beetle killed trees. The five maps show how plots could be distributed across the project area. Three examples show plot locations selected randomly and two show plot locations selected with bias. Note that each sample design illustration also includes control plots outside the project boundary.

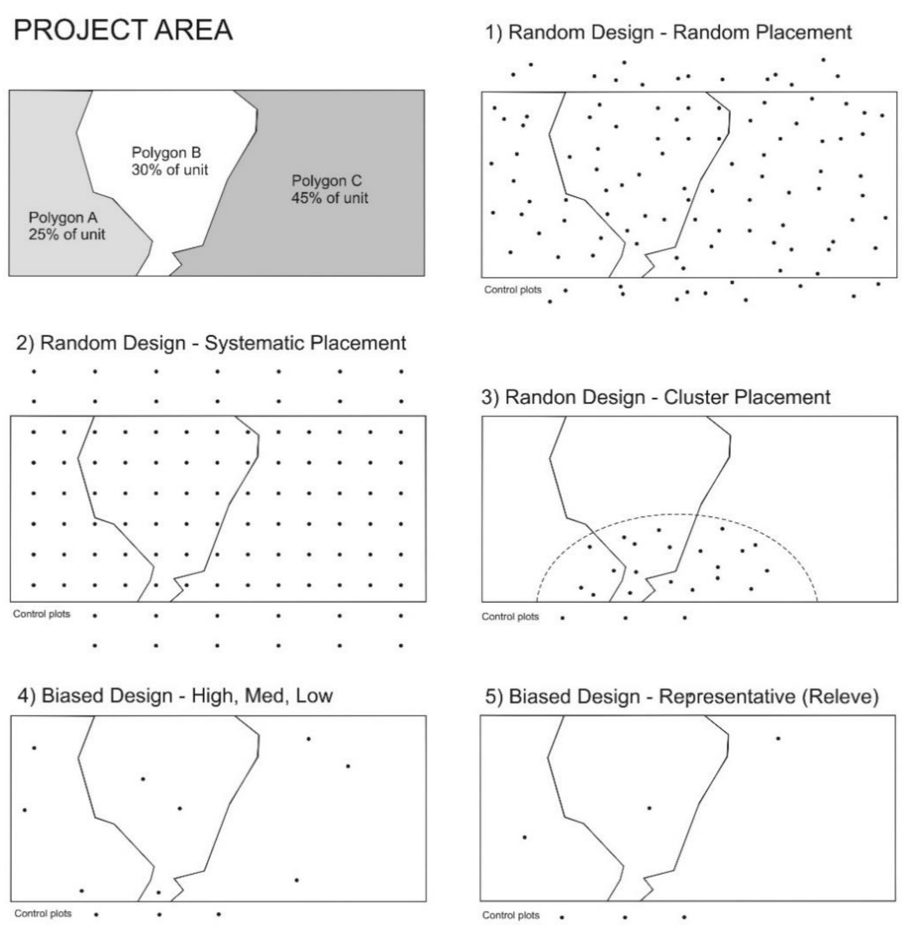
1) Random – All plots randomly located, for example, using GIS. Sample size is calculated using standard deviation, margin of error and significance level. Plot count is usually proportional to polygon size. Plots near edges may need to be moved.

2) Systematic – Randomly located first plot then subsequent plots placed systematically. Sample size is calculated using standard deviation, margin of error and significance level. Plot count is usually proportional to polygon size. Plots placed throughout the unit(s). Plots near edges may need to be moved. Easier to relocate plots in subsequent sample visits because they are on a grid.

3) Cluster – Randomly located plots limited to a specific area around an easily accessed location. Try to make plot count proportional to the area of the unit. Assumes the attributes that define each unit are consistent across the unit. May be considered a biased sampling approach because some parts of the project area not sampled. The figure displays a smaller sample size than shown in the random or systematic layout, which may limit inference.

4) High, med, low – Not statistical. Three plots placed in each unit with bias to sample high, medium and low levels of the attribute(s) of interest.

5) Representative (Releve') – Not statistical. Place one plot with bias in the most representative part of the unit.



References: Field Sampling Handbooks

Fire Monitoring Handbook: https://www.nps.gov/orgs/1965/upload/fire-effects-monitoring-handbook.pdf

FIREMON: www.frames.gov/firemon

Measuring and Monitoring Plant Populations:
http://msuinvasiveplants.org/documents/archives_cism/BLM_Measuring_and_monitoring.pdf

Statistical Inference and Sample Size

Statistical inference is the process of using a subsample of a population to make some judgement about a population as a whole. The judgement is based on a statistical test. For example, you might test for a statistically significant decrease in mean tree density after a thinning treatment. The result of a basic statistical test is the function of three inputs: 1) the variability of the attribute within the population (standard deviation), 2) the magnitude of the change you want to identify and 3) the certainty with which you want to be sure you are "right" about the answer (the p-value).

Determining sample size is the first step when undertaking a statistical study. Sample size is determined using the variability of the attribute within the population (standard deviation), the certainty of the test result (the confidence level) and margin of error. The goal of determining the appropriate sample size is calculating the number of observations needed to capture the variability of the attribute. In other words, you need to collect enough observations such that the variability in the sample is similar to the variability of the population. Once that is accomplished there is no statistical reason to collect more data. Note that projects typically use multiple objectives, each based on an attribute of interest (e.g., tree density or fuel load). Attributes typically have independent variability, which suggests there will be a different sample size required for each objective. The simplest way to handle this is sample the number of plots needed to capture the variability of the most variable attribute but that can lead to oversampling for other attributes. Sample size needs to be carefully balanced between capturing the variability of the attributes with the desired certainty, and the resources available for sampling.

For managers, making the effort to determine sample size is often moot. Managers rarely know the variability of each attribute of interest (e.g., tree density, fuel loading, species veg cover), which is an important input in the sample size calculation. Also, we have limited resources (i.e., time, people, money, equipment, expertise), so collecting large samples isn't practical, which means the opportunity to collect enough data to note a small, statistically significant change with high confidence typically isn't realistic.

If you are not able to collect enough data to precisely estimate the variability of the population you may still note statistically significant change, but the test will require a larger margin of error or lower confidence or both, compared to using an appropriate sample size. Given the same sample size, you can trade off the confidence and the margin of error - you can increase the confidence by expanding your margin of error or you can decrease the confidence and narrow the margin of error.

Confidence Intervals are calculated using standard error of the mean (not the same as the standard deviation), which decreases as sample size increases. Thus, given a constant level of significance, the confidence interval will decrease as sample size increases.

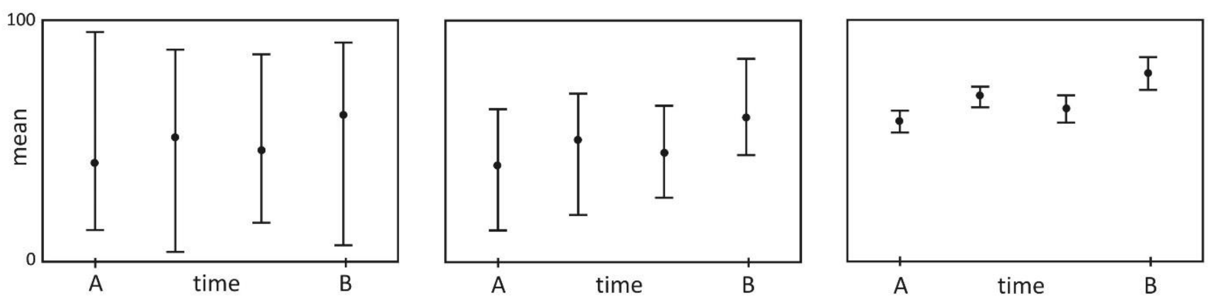
When the confidence intervals of two samples overlap there may or may not by statistically significant differences in the sample means. If the confidence intervals don't overlap, then there is significant difference in sample means.

The confidence intervals below show the same sample means (circles) from three samples of the same population with small (left) to large (right) sample sizes. Focusing on just the two labeled points (A and B) you can see the sample mean is higher at time period B than at time period A but the questions to ask yourself is, is it a statistically significant increase?

Left graph - The confidence intervals broadly overlap so there is little or no evidence of a statistically significant increase in the mean between time periods A and B.

Center graph - The confidence intervals of samples A and B overlap more narrowly than in the left graph but may or may not indicate a significant increase in the mean. However, encouragingly, there is also a positive *trend* – the means tend to be increasing and the confidence intervals are of similar length – providing some evidence of an attribute increase. Managers noting an intuitive change in an attribute, especially one supported by research (e.g., increased cover of a desirable species due to a specific treatment), may use trends as evidence of treatment effectiveness with moderate confidence.

Right graph - The confidence intervals for time periods A and B do not overlap, which is evidence there is a statistically significant difference in the means at the two time periods, at the confidence level used to calculate the confidence intervals (e.g., 95%). Note that, even though the test indicates a statistically significant increase, there is the possibility the increase is 'by chance' and the significance test is in error.



A chance increase in the mean with little or no indication of significant different in the means at time periods A and B

An encouraging, but statistically uncertain, positive trend in the sample means at time periods A and B

A statistical increase in sample means at time periods A and B

Example Using Real Data

The data displayed in the graph and table below were collected to test for a significant reduction in tree density before and after treatment. The bar graph displays means (blue = pretreatment and gray = posttreatment) and confidence intervals (capped lines within the bars) for randomly selected subsets of the total sample with sample sizes ranging from *n*=5 to *n*=35. This example uses a 95% confidence level to test for significantly different means, pre- and post-treatment. The *Prob* column displays the probability of the *t*-statistic. If the probability in the table is less than 0.05 (i.e., 1 – confidence level) then the pre- and post-treatment means for that sample are significantly different at the 95% confidence level.

Observations to note:

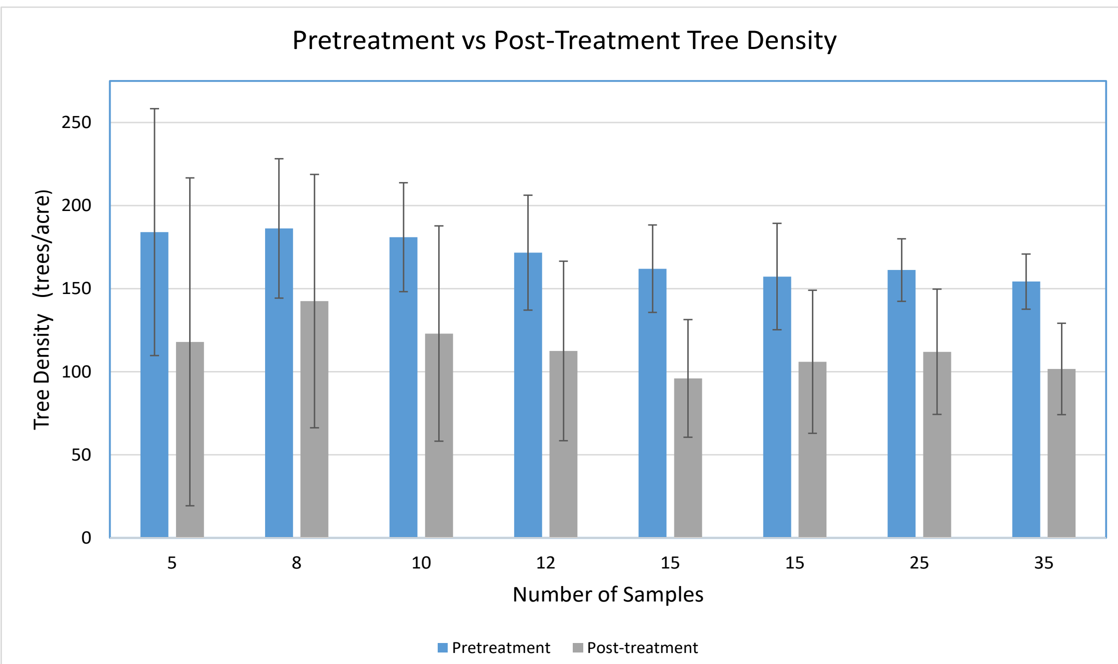
1) As sample size increases the confidence intervals tend to get shorter making it easier to detect a smaller significant change. For example, the difference in mean tree density (Diff) when *n*=5 is 66 trees per acre and the means are not significantly different (*P*(*t*)=0.176). The difference in mean tree density when *n*=35 is about 53 trees per acre and the means are significantly different (*P*(*t*)=0.001). In other words, even though the difference in pre- and post-treatment tree density is 20% smaller in the 35-sample data set vs. the 5-sample dataset, the means in the 35-sample dataset are significantly different but the means in the 5-

sample dataset are not. This demonstrates the general rule: as the sample size increases you will likely be able to note smaller differences in the means in statistical tests.

2) Compare the amount of overlap in the confidence intervals for each sample in the graph and the *Prob* value in the table for the same sample. This will help you visualize the amount of overlap needed to note: a) non-significance, b) significance and c) the amount of overlap where you will need to do a t-test to identify if there is a significance difference or not.

3) There are two 15-sample datasets. The means of the first 15-sample dataset are significantly different ($P(t)$=0.003). The means of the second sample are very close to the critical level (P(t)=0.049). These two samples demonstrate how sample selection (i.e., 'chance') may influence statistical test results.

4) We are testing at the 95% confidence level; that means there is a 1 in 20 chance a statistical test result will be incorrect due to a rare event. We don't know (and can't find out) if any test result is 'wrong' but with eight different comparisons shown in this example there may very well be an error in one of the test results. That is why it is important to include the confidence level when stating you test result – you are never 100% sure a statistical test is correct. If interested, you can search Wikipedia for *Type I Error* and *Bonferroni Correction* for more information.



N	Pre Tree Den	Post Tree Den	Diff	Prob	Pre CI Lower	Pre CI Upper	Post CI Lower	Post CI Upper
5	184	118	66	0.18	109.7	258.3	19.3	216.7
8	186.3	142.5	43.8	0.25	144.3	228.2	66.3	218.7
10	181	123	58	0.09	148.2	213.8	58.2	187.8
12	171.7	112.5	59.2	0.06	137.1	206.2	58.2	166.3
15	162	96	66	0.00	135.7	188.3	60.6	131.4
15	157.3	106	51.3	0.05	125.3	189.3	63	149
25	161.2	112	49.2	0.02	142.4	180	74.3	149.7
35	154.3	101.7	52.6	0.00	137.7	170.9	74.2	129.2