

NOTE: This document is an earlier, expanded Web version of a manuscript modified and submitted for journal publication.

The Statistical Basis for the DecAID Decayed Wood Management Advisor

Bruce G. Marcot
Janet L. Ohmann
Kim Mellen

Version: 23 July 2002

ABSTRACT

The DecAID Decayed Wood Management Advisor presents data on wildlife use of snags and down wood, and summaries of forest inventories depicting range of natural (unharvested) and current (unharvested and harvested) dead wood conditions. We analyzed the wildlife data by combining, across studies, reported values on wildlife use of, and selection for, snag diameter, snag density, down wood diameter, and down wood percent cover. Because the wildlife data conformed to normal distributions, we combined the wildlife data across studies by using parametric, distribution-based approaches. We calculated composite means by weighting studies by their sample size n , and calculated composite standard deviations by weighting studies by their degrees of freedom. These composite parameters were then used to calculate tolerance intervals, which represent the percent of observations of each wildlife species that correspond to particular sizes or amounts of snags and down wood. Tolerance intervals were calculated by using table values corrected for composite n , at a confidence level of $p = 0.90$. Results for the wildlife data are presented as 30%, 50%, and 80% tolerance intervals for each species, which the user can interpret as increasing levels of “assurance” of providing for species. For each tolerance limit, we rank-ordered all wildlife species into cumulative species curves for each snag and down wood parameter.

For the inventory summaries, we combined field plot data on snags and down wood from three regional forest inventories. Because the sample of dead wood did not conform to normal distributions, we summarized the inventory data by using distribution-free calculations of the same tolerance limits as used with the wildlife data. Inventory sampling intensity (plot spacing) varied with land ownership, geographic location, and wilderness status. Therefore, in constructing relative frequency distributions from the plot data we applied varying plot weights. For calculating tolerance limits, we subsampled from the component datasets to achieve equal sampling intensity, and then calculated quantiles associated with the tolerance limits based on the binomial distribution adjusted for n

Equations and code used in statistical analysis programs for all wildlife and inventory calculations are presented or available upon request.

INTRODUCTION

DecAID is an advisory system for use in managing snags, down wood, and other wood decay elements in forests of Oregon and Washington (Marcot et al., submitted). DecAID is based largely on a statistical summary and synthesis of empirical data on wildlife and forest vegetation. This paper presents the statistical methods used in DecAID.

DecAID provides information, for each of 27 vegetation conditions, combined from existing studies on:

- Cumulative number of wildlife species and specific species composition (the cumulative species curves)
- Percent of the statistical (~biological) population, of each species, that is observed to use specific sizes and amounts of snags and down wood
- Amounts of snags and down wood in unharvested and all current conditions sampled on regional forest inventory plots.

One also can use the statistics provided in DecAID on snag and down wood levels, along with the cumulative species curves, to determine the sizes and amounts of snags and down wood needed to meet stated target species goals (such as a specified number of species, or needs for a given species). This approach would allow the analysis of alternate targets, and would help in the evaluation of tradeoffs between (1) costs associated with providing snags and down wood, and (2) risks to species from not meeting their habitat needs.

There are two major components to the DecAID Advisor: summaries of empirical wildlife study data, and summaries of plot-level data from regional forest inventories. In this paper, we discuss each component separately, as the statistical basis and methods for each differed.

STATISTICAL ANALYSIS OF WILDLIFE DATA

A New Analysis Approach for Wildlife

DecAID represents a new approach to analyzing snags, down wood, and other wood decay elements as habitat for wildlife. It departs from previous efforts to model wildlife use of snags and down wood in Washington and Oregon, such as modeling population potential of wildlife species (e.g., Thomas et al. 1979, Neitro et al. 1985), in several important ways:

- DecAID is based entirely on a statistical synthesis of empirical data from field studies. Previous approaches were based on calculating numbers of snags used based on home range size and modeling snag use per breeding pair of cavity-excavator wildlife species.

DecAID does not rely on untested calculations of species' use of, and requirements for, snag sizes (diameters) or amounts as previous models had done. DecAID is based on observed patterns of use, and mostly on studies that have demonstrated species' selection for specific types, densities, or sizes of snags and down wood. Previous models were based on generalized and assumed functional associations. DecAID explicitly addresses wildlife use of down wood diameters and percent cover, whereas previous models of wildlife use addressed down wood poorly or not at all.

DecAID provides information expressed as probabilities -- proportions of populations using various sizes (diameters) or amounts of snags and down wood -- instead of single, deterministic levels as in previous models. This lends better to risk analysis and risk management approaches.

DecAID provides information according to specified combinations of forest type and forest structural condition. Previous models of wildlife use of snags and down wood in Oregon and Washington, at best, stratified patterns by west or east sides of the Cascade Mountains.

DecAID provides information on all wildlife showing associations with or selection for wood decay elements, whereas previous models usually focused solely on primary cavity-excavating birds (woodpeckers, nuthatches, and chickadees).

Using probabilities or statistical intervals is an important and useful way to represent variation in data and model outcomes in ecology and wildlife management (e.g., Bender et al. 1996, Cherry 1996, Nigh 1998). As well, using parametric approaches to estimating statistical intervals is more desirable than using nonparametric (distribution-free) approaches if the data are continuous and appropriately conform to assumptions of the analyses. Otherwise, distribution-free statistics are preferable (Potvin and Roff 1993).

Combining Information

The basis for the wildlife component of DecAID is a compilation of empirical data on wildlife use of wood decay elements. Such compilation of data across multiple studies is called "combining information" in statistics (e.g., Dominici et al. 1997, Draper et al. 1992, Pena 1997). Combining information entails aggregating data so that composite means and variances (or standard deviations) are calculated correctly. Combining information has been used in a variety of other ecological analyses (e.g., Akçakaya 2000, Brook et al. 2000).

How to combine data?-- When combining data across studies, an early decision is whether to use parametric or distribution-free methods. As discussed below, the wildlife data conformed adequately to assumptions of normality, so we chose parametric methods.

The first analysis decision to make then is how to calculate composite estimators--in particular, a composite mean. Two choices are to (1) calculate simple averages, which

assumes all studies are the same in sample size and other factors, or (2) weight by studies, which assumes the studies differ. The studies we combined data from clearly differed in many ways, including sample size, so we chose the latter approach. Weighting is often the preferred approach with combining data from different studies (Gurevitch and Hedges 1999).

How to weight?-- The next analysis decision pertains to how to weight the different studies. Again, two choices present themselves:

(1) Weight by some estimate of variability, such as standard error (SE), which accounts for variation in subpopulations of data and for different sample sizes among studies. This approach assumes that you can partition variation due to n, from that due to subpopulation effects. This approach was not used for DecAID because this assumption was not true.

(2) Weight by n. This assumes all studies are independent and from the same subpopulation. The studies from which we combined data were independent, but not necessarily from the same data subpopulation. To help meet this second requirement, we stratified the data by forest vegetation type, forest structural condition class, and, in some cases, also by geographic subregion. Then, we followed the second approach, weighting by n.

How to calculate the composite means and variations?-- The next analysis decision is how to calculate a **composite mean** across studies. For K studies, each with n_i sample size, the composite mean \bar{y} over \bar{x}_i study means is:

$$\bar{y} = \frac{\sum_{i=1}^K (\bar{x}_i n_i)}{\sum_{i=1}^K n_i} \quad [\text{Eq. 1}]$$

Next, there are two ways to estimate **composite variation** among studies.

First, if SD from each population is available, and if the purpose is to estimate variation across all elemental observations, then weight SD by n (or by df, which = [n-1]). This provides the estimate of $SD^2(n-1) = SS$ (sums of squares). Sum this over all K studies and divide by the composite DF, or:

$$V_1 = \frac{\sum_{i=1}^K [(SD_i)^2 (n_i - 1)]}{\sum_{i=1}^K (n_i - 1)} \quad [\text{Eq. 2}]$$

(Note that

$$\frac{SS}{N_i - 1} = SD^2$$

for each study i.)

That is, for each combination of species, forest vegetation type, and forest structural condition class, the composite variance is estimated by equation [2] over all K pertinent studies.

Studies reported variation as standard deviation (SD), standard error (SE) or confidence interval (CI). To calculate combined tolerance limits, we converted SE and CI estimates to SD. CI was first converted to SE using $SE = CI / t_{df,\alpha}$. SE was converted to SD by using $SD = SE \cdot \sqrt{n}$.

A second way to estimate composite variance is to calculate a simple average of SD across all K studies, as:

$$V_2 = \frac{\sum_{i=1}^K SD_i}{K} \quad [\text{Eq. 3}]$$

but this does not weight by different sample sizes across studies. Since sample sizes definitely varied across studies, we chose to use estimator V_1 , equation [2] above.

To demonstrate that equation [2] (estimator V_1) is the appropriate estimator to use for calculating a composite variance among wildlife studies, we plotted SD as a function of n across all studies for combinations of species, forest vegetation type, and forest structural condition class, for which data were available. Only 6 results out of 42 tests showed statistically significant correlations ($P < 0.05$) between SD and n (Table 1), which suggests that, for the most part, n and SD are uncorrelated. We interpret the 6 cases of significant correlations to represent just random chance; 5 of the 6 cases were from studies with $n \leq 5$. This result supports our use of an estimator that individually accounts for SS for studies individually, which is what equation [2] does.

When $n_i < 5$ for an individual study, we did not calculate SD for that study and omitted that study from calculations of composite variance but included it in calculations of composite mean.

On Intervals: What is Being Predicted?

The purpose of calculating composite means and composite variances across studies is to determine the spread of values of snag dbh, snag density, down wood diameter, and down

wood percent cover used and selected by wildlife species in specific forest vegetation types and forest structural condition classes and geographic areas. It is the spread of values that provides information on what proportions of the observed population of wildlife organisms (the “elemental observations”), use various sizes and amounts of snags and down wood. This is one of the major questions that DecAID answers.

(DecAID also answers questions about the natural range of variability in dead wood, when wildlife use data are not available; see the section on Inventory methods, below)

To organize the data for this purpose, we summarized the spread of values by using statistical intervals. The next question is, what is the appropriate type of statistical interval to use? A different way of phrasing this question is, what type of variation is pertinent to managing for different “comfort” levels of providing for wildlife use of snags and down wood?

To answer these questions we chose to use tolerance intervals.

A Primer on Tolerance Intervals

There are three kinds of statistical intervals: confidence intervals, prediction intervals, and tolerance intervals. *Confidence intervals* (CI) are used to provide estimates of mean values from additional studies. *Prediction intervals* (PI) are used to provide estimates of future values of means (or SDs) based on current variance (or trend) in the sample data. *Tolerance intervals* (TI) are used to provide estimates of the proportion of elemental observations within specific percentages. TIs are best for describing historic or existing data patterns and, in particular, the distribution of values among observations.

TIs are based on the spread of values for the elemental observations, that is, among individuals of a wildlife population, whereas CIs and PIs are based on values of means or SDs from subpopulations of additional or future samples (studies). It is clear that TIs are the most appropriate type of interval to use for the DecAID Advisor, where the question pertains to needing to know snag or down wood values corresponding to a specified proportion of the observed wildlife population, or the proportion of the observed wildlife population that is encompassed by a specified snag or down wood value.

If a data set is normally distributed, the mode is the mean (\bar{y}) and half of all elemental observations fall below the mean and half fall above. Further, about 66% of all observations fall within one standard deviation of the mean, that is within $\bar{y} \pm 1$ SD, so that 33% of all observations are within $\bar{y} - 1$ SD and 33% are within $\bar{y} + 1$ SD, as illustrated in the following figure:

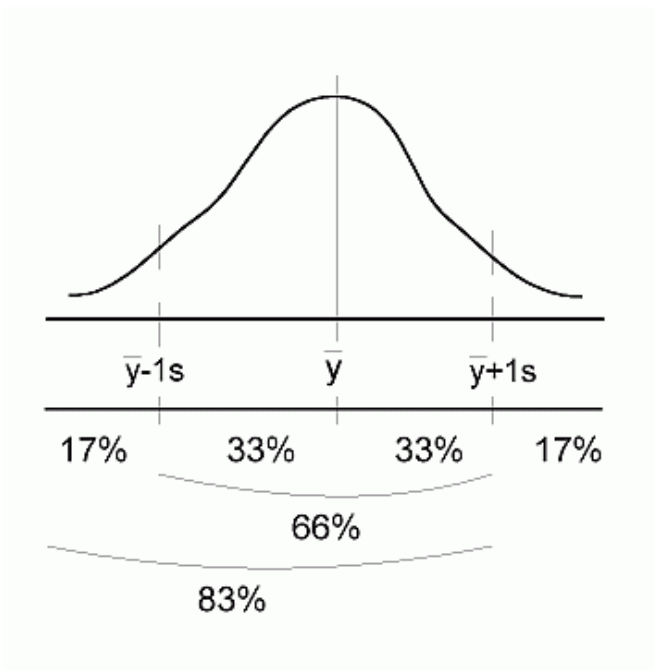


Figure 1. Tolerance intervals from normally distributed data for $\bar{y} \pm 1s$.

The intervals shown in Figure 1 are symmetric tolerance intervals because a normal distribution is symmetric around the mean.

Therefore, the interval $\bar{y} \pm 1s$ corresponds to 66% of elemental observations which fall in the interval $[\bar{y} - 1s, \bar{y} + 1s]$. Alternatively, $(100 - 17\%) = 83\%$ of elemental observations are included in the interval $[0, \bar{y} + 1s)$. This latter interval is a one-sided tolerance interval, which is the type of tolerance interval we used in DecAID. Note that a tolerance interval is a range of values bound by a tolerance limit. For example, in Figure 1, the 83% tolerance limit is the single value occurring at $\bar{y} + 1s$, whereas the tolerance interval is the range of values from $-\infty$ up to $\bar{y} + 1s$.

A broader tolerance interval can be defined for normally distributed data as:

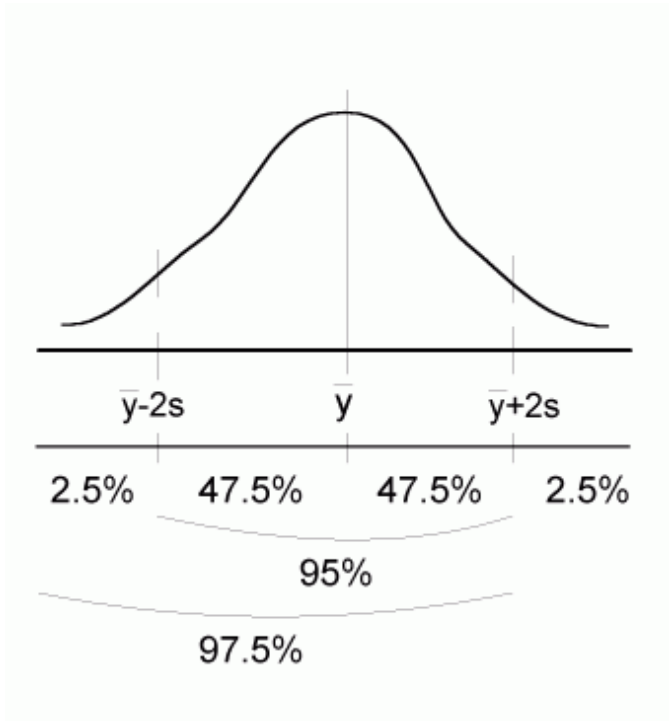


Figure 2. Tolerance intervals from normally distributed data for $\bar{y} \pm 2s$.

(Actually, the exact value for $\bar{y} \pm 2s$ for the percentages shown in Figure 2 is $\bar{y} \pm 1.96s$. The constant 1.96 derives from the Z distribution of normally distributed data.) As above, the interval $\bar{y} \pm 2s$ corresponds to 95% of elemental observations which fall in the interval $[\bar{y} - 2s, \bar{y} + 2s]$. Alternatively, $(100 - 2.5\%) = 97.5\%$ of elemental observations are included in the interval $[0, \bar{y} \pm 2s]$ when $(\bar{y} - 2s) < 0$.

Further, for a particular value, such as a given snag dbh or density, one can calculate the specific percentage of elemental observations falling above or below this value. In this way, one can determine the tolerance limit represented by that value, and the tolerance interval (percent of elemental observations) that tolerance limit represents (where tolerance limits are the specific values that bound tolerance intervals.)

Calculating Tolerance Intervals

With normally distributed data.-- Assuming normality in the distribution of values, such as for snag dbh or snag densities used by wildlife, among the elemental observations – or even if the distribution is strictly nonnormal but forms a nonmonotonic “mound” distribution with one mode (e.g., as in a skewed normal distribution) -- tolerance intervals can be roughly estimated as follows:

about 66% of elemental observations are ± 1 SD
about 95% of elemental observations are ± 2 SD
about 99% of elemental observations are ± 3 SD

for two-sided tolerance intervals. For one-sided tolerance intervals,

exactly 50% of elemental observations are $\leq \bar{y}$
about 83% of elemental observations are $\leq \bar{y} + 1S$
about 97.5% of elemental observations are $\leq \bar{y} + 2S$

and so on, for any desired percentage or interval. All these calculations presume large sample sizes; corrections for small samples sizes are discussed below.

With non-normally distributed data.-- However, if the distribution of elemental observations is distinctly monotonic or half-normal, that is, not mound-shaped, or is very highly skewed, then Chebychev's Rule or some other such correction or approach should be used to calculate tolerance intervals for any such distribution. This will substantially widen the intervals, that is, will result in much more conservative (broader) estimates of tolerance limits, to ensure inclusion of the correct values. Alternatively, a distribution-free, or non-parametric, method of determining tolerance limits can be used. (See section on inventory field plot data.)

Testing for normality.-- Our statistical analyses of the distribution of the wildlife data suggested no need to apply Chebychev's Rule or use a distribution-free method. We demonstrated this by plotting values of the elemental observations (for those published studies that provided such elemental data values; not all published studies provided this) and determining which distribution patterns they follow. To determine this, we used probability plots and observed the degree of linearity of the scatter. Probability plots are scatter diagrams that plot the actual value of some parameter against its value calculated from a normal distribution, and are used to determine the degree of normality in the distribution of the data. Results for the wildlife data generally suggested strong linearity in the probability plots and thus adherence to normal distributions.

Three examples demonstrate the typical degree of normality in the wildlife data. The more linear the scatter in these probability plots, the greater the degree of normality of the data. The following plots suggest a high degree of normality with some minor variation to be expected from sample data of this type, particularly when combining data across studies.

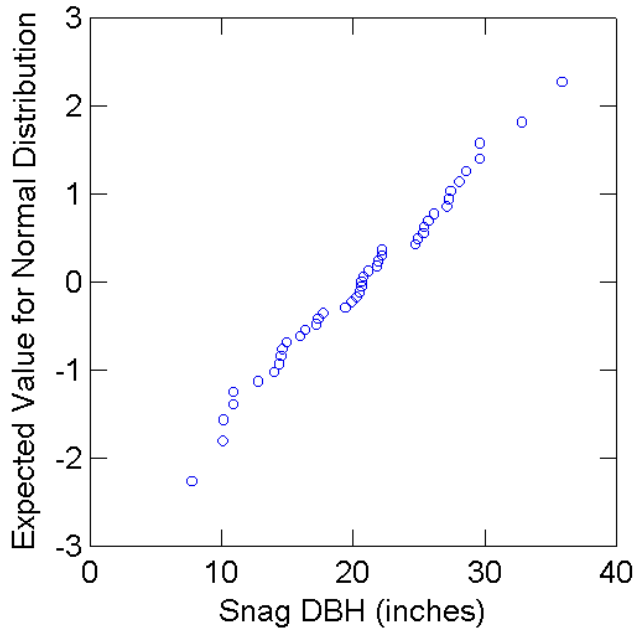


Figure 3. Probability plot of snag DBH nesting data for Eastside Mixed Conifer Forest, Small/Medium Tree structural condition class. The scatter is highly linear, suggesting that the original data are distributed normally.

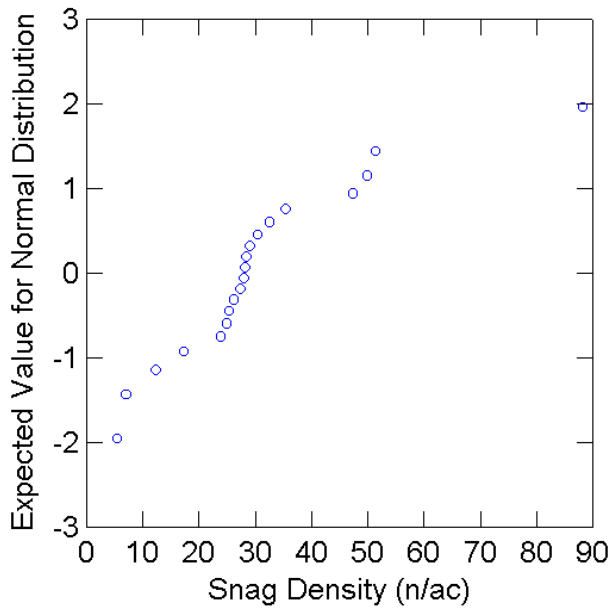


Figure 4. Probability plot of snag density nesting data for Eastside Mixed Conifer and Ponderosa Pine/Douglas-fir Forests, Open Canopy structural condition class in post-fire environments. The data are generally normal, although a slight bend in the plot appears.

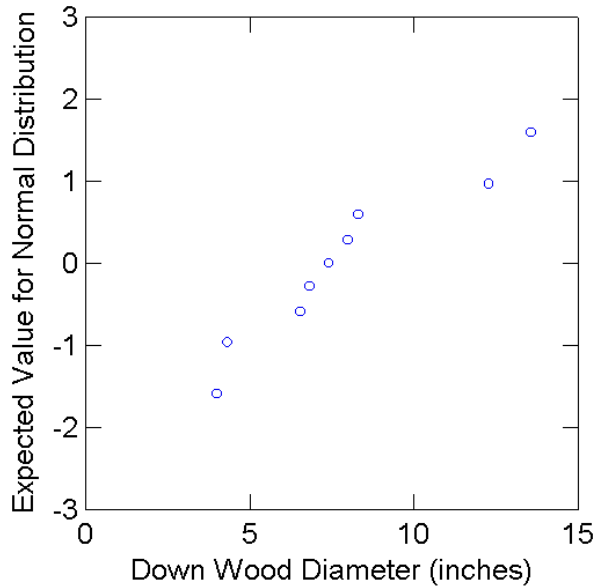


Figure 5. Probability plot of down wood diameter data for Eastside Mixed Conifer Forest, Small/Medium Tree structural condition class. Even with a small sample size such as in this example, the data still appear mostly normal in distribution.

Correcting Tolerance Interval Calculations for Small Sample Sizes

Correcting for degrees of freedom (df).-- For small sample sizes, the calculation of tolerance limit values should follow the above calculations with an additional factor based on degrees of freedom. The factor is a g' statistic derived from textbook table values of tolerance limits.

In this approach, a one-sided lower $100(1-\alpha)\%$ tolerance limit to be exceeded by at least $100p\%$ of a normal population is given by:

$$\tilde{T} = \bar{y} - g'_{(1-\alpha;p,n)} s \quad [\text{Eq. 4}]$$

and a one-sided upper $100(1-\alpha)\%$ tolerance bound to exceed at least $100p\%$ of the population is given by:

$$\tilde{T} = \bar{y} + g'_{(1-\alpha;p,n)} s \quad [\text{Eq. 5}]$$

where \bar{y} = the composite mean (equation 1), s = the composite variance (equation 2), and g' is the table value based on the acceptable level of error $1-\alpha$, the desired confidence level p , and sample size n (Hahn and Meeker 1991:60).

Of course, the 50% tolerance limit for normally distributed data is the mean itself, or

$$T \equiv \bar{y}$$

Calculating composite df.-- The table values for g' can be found in Table A.12 of Hahn and Meeker (1991:312-315). The table is based strictly on sample size n , not df , but it can be used as if $n \approx df$ for practical purposes (T. Max, pers. comm.). Degrees of freedom can be calculated as the composite among all studies, as follows:

	df

Study 1	n_1-1
Study 2	n_2-1
Study 3	n_3-1
.	
.	
Study K	n_K-1

	$\sum_{i=1}^K (n_i - 1)$
	= composite df

Then assume that the n column in Table 12 of Hahn and Meeker is \approx composite df .

Choosing error rate.-- We chose to use an error rate of $P = 1-\alpha = 0.90$, which was a compromise between statistical confidence and statistical power (Thomas 1997, Steidl et al. 1997). This means that we accept a 1 in 10 chance that values actually exceed a calculated tolerance limit. We justify this level of confidence on the basis of increased statistical power for the type of error incurred. That is, when there is an error, it errs on the side of including larger snag or down wood diameters or abundance levels rather than smaller ones (because the tolerance limits are one-tailed and bounded on the upper end).

Calculating tolerance limits.-- Then, for the 30% tolerance limit we used $p = 0.700$ to derive the g' table value from which to subtract from the mean (in Equation 3), and for the 80% tolerance limit we used $p = 0.800$ to derive the g' table value from which to add to the mean (in Equation 4), both at the confidence level of $P = 0.90$. For any other tolerance limit, the table values can be read directly or interpolated.

Any $X\%$ tolerance limit provides the size or amount of snags or down wood, for a given wildlife habitat and structural condition class, which includes only $X\%$ of the elemental observations (individual wildlife organisms), in other words, $X\%$ of the observed population used up to those particular size or amount values. For instance, an 80% tolerance limit on snag dbh for a particular wildlife habitat might be, say, 40 inches. This

means that 80% of the observed wildlife, for which there were data reported in the literature, in that particular habitat, used and selected for snags up to 40 inches dbh, and 20% used larger snags. The actual tolerance limit values varied by wildlife habitat.

Tolerance limits and levels of “assurance”.-- If the DecAID user wants very high assurance that snags and down wood sizes and abundances would match what is reported in the literature (that is, would provide for most of the individual wildlife organisms of a given species), then the user would select the 80% (or higher) tolerance limits for those attributes. If the user wishes to provide only a low level of such assurance, they would select a low tolerance limit such as the mean values (50% tolerance limit) or the 30% tolerance limit.

The specific tolerance limits that we explicitly list in the DecAID tables and figures for the wildlife data are 30%, 50% (= mean), and 80%. Any other tolerance limit also can be calculated from the wildlife data for any specified snag or down wood amount, based on the species data we compiled. But we felt that these three tolerance limits nicely span low, medium, and high degrees of “assurance” of including various proportions of observed elemental observations (individuals’ use of snag or down wood sizes or amounts). The user can select the appropriate level of assurance (tolerance limit) depending on their specific management objectives and risk attitude.

Tolerance limits with large n.-- Note that when n is very large, on the order of > 500, then the g’ statistic approximates Z values for normally distributed data. In these cases, Equations [4] and [5] simplify to:

$$\underline{\tilde{T}} = \bar{y} - 0.524s \quad [\text{Eq. 4a}]$$

and

$$\overline{\tilde{T}} = \bar{y} + 0.842s \quad [\text{Eq. 5a}]$$

where the two constants are Z values for large sample size for p=0.700 and p=0.800, respectively. Also, at such large sample sizes, values of these g’ constants are insensitive to confidence levels p.

An Example of Tolerance Intervals From the Wildlife Data

Figure 6 is an example of tolerance intervals of data on wildlife use of snag diameters, from the Eastside Mixed Conifer Forest habitat, Open Canopy structural condition, in the East Cascades/Blue Mountains geographic region of Washington and Oregon.

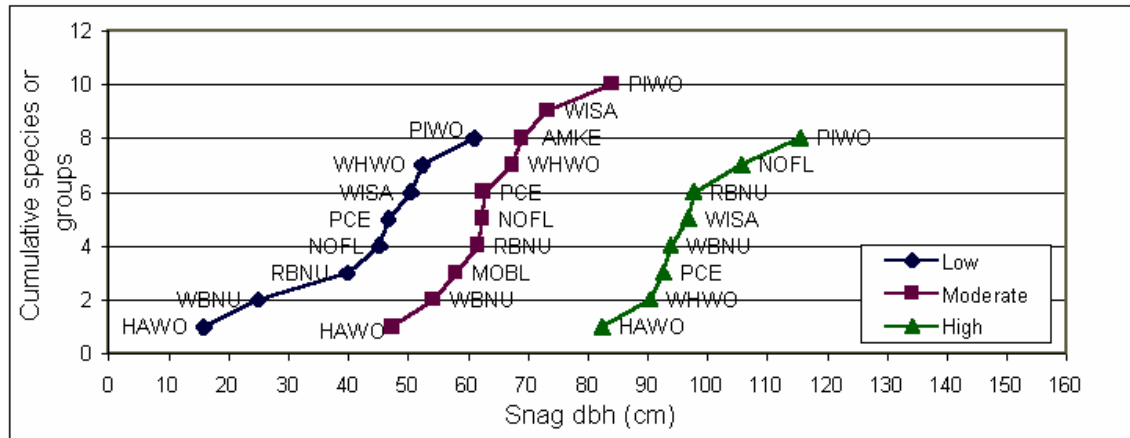


Figure 6. Cumulative wildlife species curves for Eastside Mixed Conifer Forest, Open Canopy vegetation condition, in the East Cascades / Blue Mountains region, depicting three tolerance levels. Low = 30% tolerance level, Moderate = 50%, High = 80%.

The example shows 3 “cumulative species curves” which are wildlife species rank-ordered by, in this case, snag diameter (cm dbh). The 3 curves represent three tolerance levels of 30% (“Low”), 50% (“Moderate”), and 80% (“High”). The points are labeled with codes for names of individual species or species groups, such as PIWO for Pileated Woodpecker, and PCE for primary cavity excavators, depending on how they were reported in the original publications from which the data were taken. The cumulative species curves for each tolerance level do not represent a function in the traditional sense, but serve to depict arrays of species potentially provided by, in this case, various snag diameters at 3 levels of tolerance.

Arrayed in this way, the data can be read in two ways: the snag or down wood condition (e.g, snag dbh) necessary to match use patterns reported in the literature for one or more wildlife species at a given tolerance level, told by reading up from the x-axis; and the number of, and specific, species that could be provided for with a given snag or down wood condition, told by reading over from the y-axis. For example, if a forest stand in this vegetation condition had snags up to 60 cm dbh, this could provide for all of the 8 wildlife species or groups plotted on the Low curve (that is, for 30% of the observed individuals of each species), for about 6 of the 10 species plotted on the Moderate curve (50% of the observed individuals), and for none of the 8 species plotted on the High curve (80% of the observed individuals). To meet the snag dbh use by all 10 species at the Moderate level, snags would need to be about 80 cm dbh. In this way, managers can use the tolerance data as an indication of the percent of the empirical observations of each species using a given snag or down wood condition.

STATISTICAL ANALYSIS OF REGIONAL INVENTORY DATA

We summarized the regional inventory plot data for DecAID in two primary ways: relative frequency distributions and tolerance limits.

Relative Frequency Distributions

Relative frequency distributions are a simple way of graphically portraying the variability in dead wood sampled on inventory plots within a given vegetation condition. The relative frequency distributions characterize variation in the plot data and do not have any statistical properties or confidence levels associated with them. The distributions are useful as stand-alone characterizations of dead wood populations (especially valuable when wildlife data are lacking), and can be compared to variability in the wildlife use data as well. An example of a frequency distribution is shown in Figure 6.

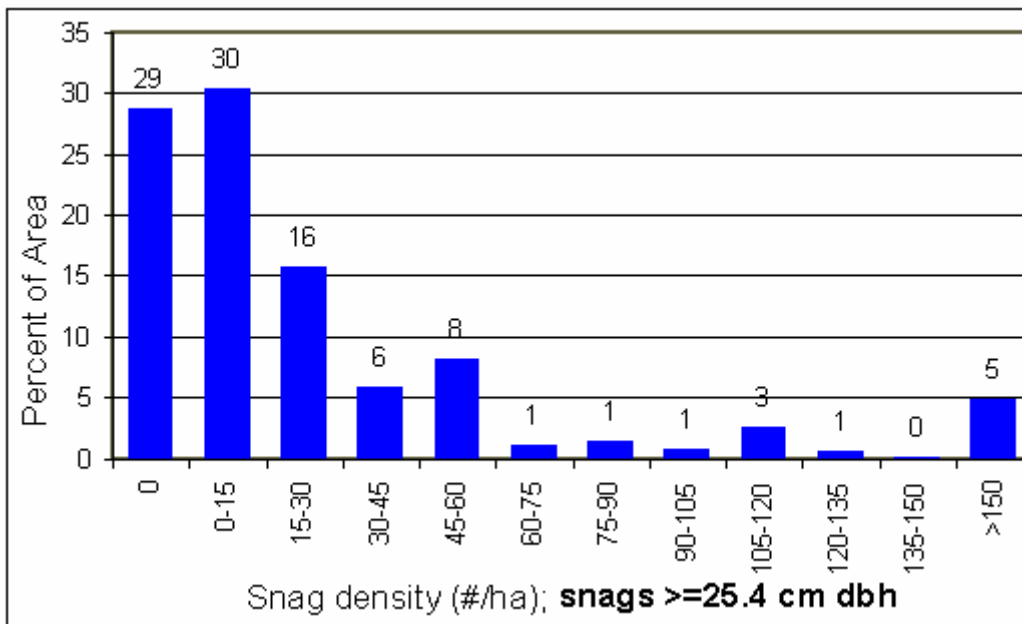


Figure 7. An example of a frequency histogram from the snag inventory data. This histogram shows the distribution of the unharvested area of the Eastside Mixed Conifer Forest in eastern Cascades and Blue Mountains, among snag density classes (no. / ha) for snags ≥ 25.4 cm dbh, based on 132 unharvested inventory plots.

Tolerance Limits

The rationale for using tolerance limits to describe the wildlife use data, described earlier in this paper, apply to the inventory plot data as well. Probabilities and statistical intervals are useful ways to represent variation in the plot data, and thus in forested landscapes, and this variation has implications for forest management. The spread of values in the

plot data provide information on proportions of the observed population of dead wood that occur at different abundance levels across forested landscapes. We chose to describe the plot data using tolerance limits to be consistent with how wildlife use data are being summarized for DecAID, to facilitate comparisons between the field plot and wildlife use information. Nevertheless, describing variation in the plot sample of dead wood presented a different set of statistical challenges than did the combining of wildlife study data, which we describe below.

For each vegetation condition, we present the tolerance limits graphically using a format similar to a “box-and-whisker” diagram. In our version of the box-and-whisker (fig. X), the center of the box represents the 50% tolerance limit, the left and right boundaries of the box indicate the 30% and 80% tolerance limits, respectively, and the left and right ends of the extended lines represent the minimum and maximum values, respectively, present in the data.

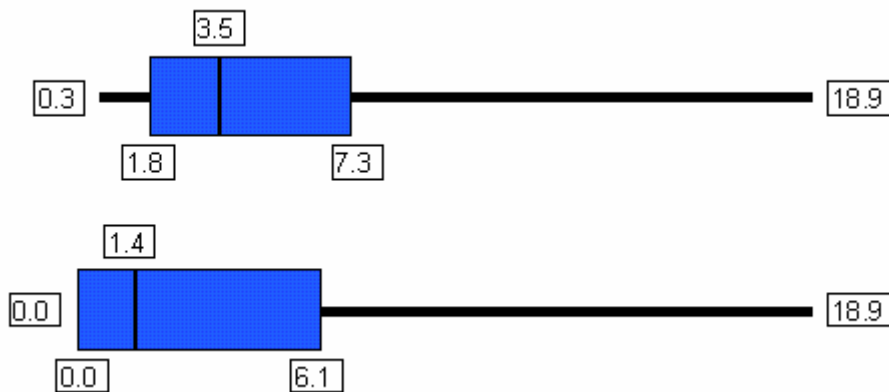


Figure 8. An example of "box-and-whisker" diagrams depicting inventory data of percent ground cover of down wood ≥ 12.5 cm diameter, on unharvested inventory plots (n=67) in Eastside Mixed Conifer Forest, Open Canopy vegetation condition, in the eastern Cascades and Blue Mountains. The upper box depicts inventory plots with measurable down wood, and the low box depicts all unharvested inventory plots including those with no down wood. In the upper box, the values 1.8, 3.5, and 7.3 (percent cover of down wood) are the 30%, 50%, and 80% tolerance levels, and 0.3 and 18.9 are the minimum and maximum observed values.

Combining Information for Areas Sampled at Different Intensities

The field plot observations of dead wood summarized for DecAID were from three different regional forest inventories: Bureau of Land Management (BLM), Current Vegetation Survey (CVS), and Forest Inventory and Analysis (FIA) (Table 2). The inventories used very similar sampling methods at the field plot level. However, sampling intensity – the density of plots per unit area, a function of the average distance between plots -- differed among ownerships, land allocations, and geographic locations. Because dead wood characteristics also vary with these factors, the different inventories (or portions thereof) sampled different populations of dead wood and needed to be weighted differently in our analyses.

Although plots from each of the three inventories were established on systematic grids, the sampling intensity (grid spacing) differed among the three inventories, as well as between land allocations within the CVS inventory, and between geographic locations within the FIA inventory (Table 2). In constructing the *relative frequency distributions*, we used the plot weights in Table 2 in determining the percent of sampled area in different dead wood abundance classes. The weights assigned to plots that contained multiple forest conditions were adjusted further (see below). For calculating *tolerance limits*, we subsampled the component datasets to achieve consistent sampling intensity throughout the study area. This is discussed in detail later in this paper.

The plot weights also were adjusted to account for multiple forest conditions within the plot area. The BLM and CVS plots were laid out on the ground in a fixed configuration and thus often straddled multiple forest conditions. Forest conditions were defined as forest vs. nonforest, and by series (potential vegetation types) within forest. These partial plots, or condition-class-plots, comprised the observations in our database and are referred to here simply as ‘plots.’ In constructing the *relative frequency distributions*, plots confined to a single forest condition were assigned plot weights as shown in Table 2. Partial plots in distinct forest conditions received an apportioned amount of the total plot weight for the entire plot, consistent with the percentage of the plot area occupied by the distinct forest condition. We did not downweight partial plots in calculating *tolerance limits*, based on the rationale that the partial plot is an independent observation of that forest condition but with a smaller plot size.

Non-Normal Distributions of Dead Wood on Field Plots

Before calculating *tolerance limits*, we determined the underlying distribution of the dead wood data on plots. Unlike the assumption of normality made for the data on wildlife use of dead wood, dead wood abundance was distinctly non-normally distributed among the plots in our sample. This was true to varying degrees for all vegetation conditions and for both snags and down wood. Distributions varied from skewed normal, to half-normal or log-normal, to monotonic or log-normal. Especially problematic was the fact that almost all subsets of the data examined contained a large number of ‘zero’ observations. In other words, there were many plots on which dead wood was sampled in the field, but where no dead wood was tallied. The proportion of ‘zero’ observations within a sample is a function of plot size (or transect length) and the abundance and spatial distribution of dead wood. But regardless of the reasons for the large number of zeros, we had to accommodate the resulting non-normal distributions in our statistical summaries. Because of the non-normal distributions of dead wood, we applied distribution-free tolerance limits, which do not assume an underlying (typically normal) distribution to the data.

Two examples illustrate typical distribution patterns for dead wood in the inventory sample: the example of the relative frequency distribution in fig. 6, and the probability plot in fig. 7. A probability plot is the value of each observation plotted against its

normalized value, so a perfectly normally distributed data set will plot out as perfectly linear on a probability plot.

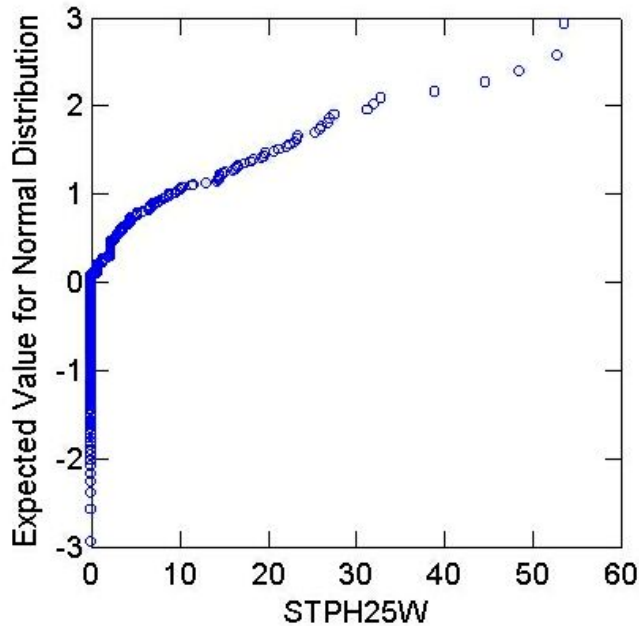


Figure 9. Probability plot of number of snags ≥ 25 cm dbh and ≥ 2 m tall on all plots in Westside Lowlands Conifer-Hardwood Forest, Oregon Coast, Open Canopy conditions. The horizontal axis is the number of snags (STPH25W) on field plots, and the vertical axis is the normalized value.

It is clear that the above probability plot (a typical pattern with the inventory data) illustrates that the inventory data are highly nonnormally distributed. In part this is because of all the plots with zero values, but these are important values to retain in the analyses and cannot be ignored. We conclude that we cannot use the same kind of parametric statistics to calculate tolerance limits as were used with the more normally-distributed wildlife data. Instead, a distribution-free method had to be devised to calculate tolerance limits for the inventory data. This method is described next.

Calculating Distribution-Free Tolerance Limits for the Inventory Data

Definition of Distribution-Free Tolerance Limits.--Distribution-free tolerance limits can be defined as follows (from Mood et al., p. 515-518):

Let $X_1 \dots X_n$ be a random sample from a continuous density function $F(\cdot)$, having a density function $f(\cdot)$.

Let $L_1 = l_1(X_1 \dots X_n) < L_2 = l_2(X_1 \dots X_n)$ be two statistics that satisfy:

- i) the distribution of $F(L_2) - F(L_1)$ does not depend on $F(\cdot)$
- ii) $P[F(L_2) - F(L_1) \geq \beta] = \gamma$

Then L_1 and L_2 will be 100% distribution-free tolerance limits at probability level γ (γ):

$$P\left(\int_{L_1}^{L_2} f(x)dx \geq \beta\right) = \gamma$$

Furthermore, we can order the data with n observations from the smallest value (Y_1) to the largest value (Y_n). For a tolerance limit we want the value of k and j such that the probability that β % of the data are between Y_k and Y_j is $\geq \gamma$. When $j = 0$ we have a one-sided tolerance interval. We can write this as:

$$\begin{aligned} \gamma &= P[F(Y_k) - F(Y_j) \geq \beta] \\ &= 1 - P[F(Y_k) - F(Y_j) < \beta] \\ &= 1 - \sum_{i=k-j}^n \binom{n}{i} \beta^i (1-\beta)^{n-i} \\ &= \sum_{i=0}^{k-j-1} \binom{n}{i} \beta^i (1-\beta)^{n-i} \end{aligned}$$

Calculation of Quantiles Associated with Distribution-Free Tolerance Limits. We calculated quantiles (observation order number, or k) for a range of values of n (sample size), for the following values of γ and β being used for both wildlife and plot data in DecAID. The Appendix contains two programs (SAS and S+), written by Manuela Huso (statistician, Forest Science Department, Oregon State University), that calculate quantiles, or order statistics, for other γ s and β s.

- γ = level of certainty (0.90 for DecAID)
- β_1 = lowest β of interest (0.30 in DecAID)
- β_2 = intermediate β of interest (0.50 in DecAID)
- β_3 = largest β of interest (0.80 in DecAID)

The calculations are based on the binomial probability distribution, where
 p is a numeric probability of success parameter, $0 \leq p \leq 1$
 n is an integer number of independent Bernoulli trials parameter, $n > 0$
 m is an integer number of successes for a random variable, $0 \leq m \leq n$

(see coding for the parameter **PROBBNML(p,n,m)** in the Appendix).

The variables output from the computer program are:

- $SAMP_SZ$ = sample size (n)
- $QUANT_BI$ (k for β_1) = observation whose value you are γ % certain that β_1 % of the population will have values less than its

QUANT_B2 (k for β_2)= observation whose value you are γ % certain that β_2 % of the population will have values less than its

QUANT_B3 (k for β_3)= observation whose value you are γ % certain that β_3 % of the population will have values less than its

Table 3 presents order statistics based on one-sided distribution-free tolerance limits for selected sample sizes, different β s of interest ($\beta_1=0.3$, $\beta_2=0.5$, $\beta_3=0.8$), and $\gamma=0.9$ (90% certainty). Example for $n=10$: The tolerance limit equal to the value (e.g., snag density) of the 5th largest observation will allow you to be at least 90% certain that 30% of the data from the entire population (not your sample) will fall below this value. Note that when $n=500$, the normal approximation gives ks of 170, 272, and 417 for the three β levels, not far from the exact binomial case shown here, but not equal either.

Accounting for Unequal Sampling Intensities in Distribution-Free Tolerance Limits

Characterizing the Current Landscape

We accounted for different sampling intensities in the calculation of distribution-free tolerance limits in two different ways. For characterizing the current landscape (forests of all disturbance histories), we used plots from all three inventories, but only those plots that fell on the 5.5-km grid (see Table 2). This included all BLM plots and all CVS plots within wilderness areas. For CVS plots outside wilderness, this included plots numbered between 1,000,000 and 1,999,999. For FIA plots in western Washington, only those plots numbered ≥ 500 were included; all other FIA plots were used. The reduced sample sizes after subsampling were used in calculating tolerance limits.

Characterizing Unharvested Forests

For characterizing unharvested forests, we used only BLM and CVS plots. We excluded FIA plots since so few had never been harvested. Also, because there were so few CVS plots in wilderness, we weighted all CVS plots equally and did not test for differences between wilderness and non-wilderness. Before combining plot data from the BLM and CVS inventories, we conducted Mann-Whitney tests (nonparametric) and unpaired t-tests (parametric, but relatively robust to moderate violations of normality assumptions) to determine if the snag and down wood densities differed between the two datasets (Table 4). Combining the data would provide greater sample sizes and more reliable estimates of tolerance limits. We conducted the tests on all unharvested inventory plots, including “zero” plots, and including plots on all grids (5.5-km grid as well as all “interstitial” plots). We tested only those vegetation conditions having substantial numbers of BLM plots (Table 2).

We found that statistically significant differences commonly exist in snag and down wood densities between Forest Service and BLM lands (Table 4). Wherever differences were significant, snag or down wood densities were lower on Forest Service lands than

on BLM lands. Therefore, for vegetation conditions where significant differences exist between BLM and CVS (Table 4), combining all plots from both inventories will weight the dead wood summaries towards conditions on Forest Service lands, where plots are more dense. We decided to base our inventory summaries for unharvested forests on all BLM and CVS plots, but to note in our interpretations for the users where these situations occur.

Calculating Inventory Tolerance Limits

For a given vegetation condition and dead wood variable of interest, the sample size, gamma, and betas of interest were used to identify quantiles (Table 3). The actual tolerance limit (value of the dead wood variable) for each quantile (k) is the value associated with the k th observation. A SAS program used to calculate tolerance limits for DecAID can be obtained from Janet Ohmann, Forestry Sciences Lab, 3200 SW Jefferson Way, Corvallis, OR 97331).

CONCLUSIONS

We calculated tolerance intervals to describing wildlife species' use of, and forest inventory plot data on, snag and down wood size and amounts. Tolerance intervals are appropriate for describing proportions of observed (sampled) populations, and lend to viewing the data probabilistically in a risk management framework.

We applied the methods to our synthesis of research data and publications on wildlife use of snag diameter, snag density (n/ha), down wood diameter, and down wood percent cover. We summarized results as cumulative wildlife species curves that depict rank-orders of species for each of three tolerance levels (30%, 50%, and 80%). We also developed and applied methods for calculating snag density and down wood percent cover by diameter class from forest inventory data in harvested and all (harvested and unharvested) conditions for these three same tolerance levels. Results of the analyses are being used in a snag and down wood management advisory system, the DecAID Advisor.

ACKNOWLEDGMENTS

We thank statisticians Tim Max and Manuela Huso for confirming the appropriate use of tolerance intervals and for guiding correct use of equations for calculating tolerance limits. Thanks to Manuela Huso for writing the S+ and SAS computer code presented in the Appendix. The senior author also thanks Scott Overton for teaching theory and use of tolerance intervals in graduate statistical classes at Oregon State University in the early 1980s.

LITERATURE CITED

- Akçakaya, H. R. 2000. Conservation and management for multiple species: integrating field research and modeling into management decisions. *Environmental Management* 26:75-83.
- Bender, L. C., G. J. Roloff, and J. B. Haufler. 1996. Evaluating confidence intervals for habitat suitability models. *Wildl. Soc. Bull.* 24(2):347-352.
- Brook, B. W., J. J. O'Grady, A. P. Chapman, M. A. Burgman, H. R. Akcakaya, and R. Frankham. 2000. Predictive accuracy of population viability analysis in conservation biology. *Nature* 404:385-387.
- Cherry, S. 1996. A comparison of confidence interval methods for habitat use-availability studies. *J. Wildl. Manage.* 60(3):653-658.
- Dix, L. M., and M. A. Strickland. 1986. Sex and age determination for fisher using radiographs of canine teeth: a critique. *Journal of Wildlife Management* 50(2):275-276.
- Dominici, F., G. Parmigiani, R. Wolpert, and K. Reckhow. 1997. Combining information from related regressions. *Journal of Agricultural, Biological, and Environmental Statistics* 2(3):313-332.
- Draper, D., D. P. Gaver, Jr, P. K. Goel, J. B. Greenhouse, L. V. Hedges, C. N. Morris, J. R. Tucker, and C. M. Waternaux. 1992. Combining information. *Statistical issues and opportunities for research. Contemporary statistics No. 1.* National Academy Press, Washington D.C. 217 pp.
- Gurevitch, J., and L. V. Hedges. 1999. Statistical issues in ecological meta-analyses. *Ecology* 80(4):1142-1149.
- Hahn, G. J., and W. Q. Meeker. 1991. *Statistical intervals: a guide for practitioners.* John Wiley & Sons, New York, N.Y. 392 pp.
- Krishnamoorthy, K. 2000. *Statistical distributions with applications and StatCalc software.* Etext.net Publisher, Venice CA. <http://www.etext.net/catalog>
- Marcot, B. G., K. Mellen, J. L. Ohmann, K. L. Waddell, E. A. Willhite, B. B. Hostetler, S. A. Livingston, C. Ogden, and T. Dreisbach. Submitted. DecAID -- work in progress on a decayed wood advisory model for Washington and Oregon forests. Research Note PNW-RN-xxx. USDA Forest Service, Pacific Northwest Region, Portland OR.
- Mood, A. M., F. A. Graybill, and D. C. Boes. 1974. *Introduction to the theory of statistics.* McGraw-Hill, New York. 564 pp.

Neitro, W. A., V. W. Binkley, S. P. Cline, R. W. Mannan, B. G. Marcot, D. Taylor, and F. F. Wagner. 1985. Snags (wildlife trees). Pp. 129-169 in: E. R. Brown, ed. Management of wildlife and fish habitats in forests of western Oregon and Washington. Part 1 - chapter narratives. USDA Forest Service, Portland OR.

Nigh, G. 1998. Prediction intervals for estimates of site index based on ecosystem type. *Environmental Management* 22:197-202.

Pena, D. 1997. Combining information in statistical modeling. *The American Statistician* 51(4):326-332.

Potvin, C., and D. A. Roff. 1993. Distribution-free and robust statistical methods: viable alternatives to parametric statistics? *Ecology* 74(6):1617-1628.

Steidl, R. J., J. P. Haynes, and E. Schaubert. 1997. Statistical power analysis in wildlife research. *J. Wildl. Manage.* 61(2):270-279.

Thomas, J. W., R. G. Anderson, C. Maser, and E. L. Bull. 1979. Snags. Pp. 60-77 in: J. W. Thomas, ed. *Wildlife habitats in managed forests the Blue Mountains of Oregon and Washington*. Agriculture Handbook No. 553. USDA Forest Service, Portland OR.

Thomas, L. 1997. Retrospective power analysis. *Cons. Biol.* 11(1):276-280.

Table 1. Results of correlating n with SD across studies, for various combinations of snag and down wood data, using Pearson correlation. The prevalence of non-significant results ($p > 0.05$ for significant, and $p > 0.10$ for marginal) suggests that use of equation [2] as the estimator of composite variance across studies is appropriate (see text). Veg Type = DecAID vegetation condition code, representing combinations of DecAID wildlife habitat types and DecAID structural condition classes. n/a = sample size inadequate or not given.

Snag DBH

Veg. Type	r^2	p	n
EMC_O.sp-28	0.020	0.50	25
EMC_O.sp-25	0.168	0.27	9
EMC_S.sp-25	0.001	0.89	26
EMC_L.sp-25	0.000	0.94	28
EMC_O.sp-26	0.711	0.07	5
EMC_S.sp-26	0.000	0.98	15
EMC_L.sp-26	0.005	0.80	16
EMC_O.sp-27	n/a	n/a	1
EMC_S/L.sp-27	0.000	1.00	7
PPDF_O.sp-25	0.053	0.85	3
PPDF_S/L.sp-25	0.053	0.50	11
PPDF_O.sp-26	0.782	0.31	3
PPDF_S.sp-26	0.012	0.86	5
PPDF_L.sp-26	0.012	0.86	5
PPDF_S/L.sp-27	0.003	0.93	5
PPDF_O.sp-28	0.019	0.73	9
MMC_S.sp-25	0.373	0.39	4
MMC_L.sp-25	0.330	0.31	5
MMC_S/L.sp-26	0.002	0.96	4
MMC_S/L.sp-27	n/a	n/a	n/a
SWOMC_O.sp-25	n/a	n/a	n/a
SWOMC_S/L.sp-25	0.226	0.42	5
WLCH_O.sp-25	0.061	0.39	14
WLCH_S.sp-25	0.320	0.002	26
WLCH_L.sp-25	0.200	0.03	24
WLCH_O.sp-26	n/a	n/a	n/a
WLCH_S/L.sp-26	0.771	0.32	3
WLCH_S.sp-27	n/a	n/a	n/a
WLCH_L.sp-27	n/a	n/a	n/a

Snag Density

Diam. (cm)	Veg. Type	r^2	p	n
>25	EMC_O.sp-30	n/a	n/a	1
	EMC_S/L.sp-30	n/a	n/a	1
	EMC_O.sp-35	0.001	0.88	16
>50	EMC_O.sp-30	n/a	n/a	1
	EMC_S/L.sp-30	n/a	n/a	1
	EMC_O.sp-35	0.635	0.20	4
>25	PPDF_O.sp-30	n/a	n/a	n/a

	PPDF_S/L.sp-30	n/a	n/a	n/a
	PPDF_O.sp-35	0.002	0.88	16
>50	PPDF_O.sp-30	n/a	n/a	n/a
	PPDF_S/L.sp-30	n/a	n/a	n/a
	PPDF_O.sp-35	0.640	0.20	4
>25	MMC_S/L.sp-30	n/a	n/a	n/a
>50	MMC_S/L.sp-30	n/a	n/a	n/a
>25	SWOMC_S/L.sp-30	n/a	n/a	n/a
>50	SWOMC_S/L.sp-30	n/a	n/a	n/a
>25	WLCH_O.sp-30	n/a	n/a	n/a
	WLCH_S.sp-30	0.452	0.20	4
	WLCH_L.sp-30	0.635	0.20	4
>50	WLCH_O.sp-30	n/a	n/a	n/a
	WLCH_S.sp-30	0.129	0.76	3
	WLCH_L.sp-30	0.129	0.76	3

Down Wood Diameter

Veg. Type	r^2	p	n
EMC_O.sp-29	n/a	n/a	n/a
EMC_S.sp-29	0.001	0.86	24
EMC_L.sp-29	0.001	0.88	22
PPDF_S/L.sp-29	1.000	0.02	3
MMC_S.sp-29	0.315	0.44	4
MMC_L.sp-29	0.428	0.54	3
SWOMC_L.sp-29	0.015	0.92	3
WLCH_S.sp-29	0.304	0.01	19
WLCH_L.sp-29	0.343	0.01	19

Down Wood % Cover

Veg. Type	r^2	p	n
EMC_S/L.sp-31	0.478	0.51	3
WLCH_S.sp-31	0.689	0.00	27
WLCH_L.sp-31	0.116	0.10	24

Table 2. Characteristics of three regional forest inventories.

<u>Inventory</u>	<u>Lands sampled</u>	<u>Grid spacing</u>	<u>Plot weight</u>
BLM 1.00	BLM, western Oregon		5.5 km
CVS 1.00	National Forest wilderness		5.5 km
0.25	National Forest non-wilderness		2.7 km
FIA 1.00	Nonfederal, all but western Washington		5.5 km
0.50	Nonfederal, western Washington		Two overlapping 5.5 km grids

Table 3. -- Order statistics (quantiles) for one-sided distribution-free tolerance limits; selected sample sizes; betas $\beta_1=0.3, \beta_2=0.5, \beta_3=0.8$; and $\gamma=0.9$ (90% certainty).

<i>SAMP_SZ</i>	<i>QUANT_B1</i>	<i>QUANT_B2</i>	<i>QUANT_B3</i>
10	5	7	10
20	9	13	18
30	12	19	27
40	16	24	35
50	19	30	44
60	23	35	52
70	26	40	60
80	29	46	69
90	33	51	77
100	36	56	85
150	52	83	126
200	68	109	167
250	84	135	208
300	100	161	249
350	116	187	290
400	132	213	330
450	147	239	371
500	163	264	411
1000	319	520	816
1500	473	775	1220
2000	626	1029	1623
2500	779	1282	2026

Table 4. Results of analyzing for differences between US Forest Service and Bureau of Land Management inventory data on snag and down wood density. See text for explanation of approach. SWOMC = Southwest Oregon Mixed Conifer Forest, WLCH_OCO = Westside Lowlands Conifer-Hardwood Forest, Oregon Coast Range geographic region, WLCH_OCA = Westside Lowlands Conifer-Hardwood Forests, Oregon Cascades geographic region; _O = Open Tree, _S = Smaller Tree, _L = Larger Tree structural condition classes. WSTPH25 = snag density (> 24.5 cm dbh), WSTPH50 = snag density (> 50 cm dbh), WCWD12 = down wood percent cover (> 12 cm diameter), WCWD50 = down wood percent cover (> 50 cm diameter).

Vegetation Condition	Snag or Down Wood Parameter	Mann-Whitney non-parametric probability (p)	Pooled variance probability (p) from unpaired t-test
SWOMC_O	WSTPH25	n.s. (p = 0.364)	n.s. (p = 0.436)
SWOMC_O	WSTPH50	n.s. (p = 0.584)	n.s. (p = 0.818)
SWOMC_O	WCWD12	** p < 0.001	** p < 0.001 (many FS plots zero)
SWOMC_O	WCWD50	** p < 0.001	** p < 0.001
SWOMC_S	WSTPH25	* p = 0.034	* p = 0.023
SWOMC_S	WSTPH50	+ p = 0.071	** p < 0.001
SWOMC_S	WCWD12	** p < 0.001	** p < 0.001
SWOMC_S	WCWD50	** p = 0.001	** p < 0.001
SWOMC_L	WSTPH25	* p = 0.016	** p < 0.001
SWOMC_L	WSTPH50	* p = 0.013	** p < 0.001
SWOMC_L	WCWD12	n.s. (p = 0.131)	** p < 0.001
SWOMC_L	WCWD50	n.s. (p = 0.352)	** p = 0.007

WLCH_OCO_O	WSTPH25	n.s. (p = 0.354)	NT (FS n = 1)
WLCH_OCO_O	WSTPH50	n.s. (p = 0.344)	NT (FS n = 1)
WLCH_OCO_O	WCWD12	n.s. (p = 0.468)	NT (FS n = 1)
WLCH_OCO_O	WCWD50	n.s. (p = 0.238)	NT (FS n = 1)
WLCH_OCO_S	WSTPH25	n.s. (p = 0.724)	n.s. (p = 0.362)
WLCH_OCO_S	WSTPH50	n.s. (p = 0.353)	n.s. (p = 0.770)
WLCH_OCO_S	WCWD12	** p = 0.001	** p = 0.001
WLCH_OCO_S	WCWD50	** p = 0.005	** p = 0.001
WLCH_OCO_L	WSTPH25	** p < 0.001	** p < 0.001
WLCH_OCO_L	WSTPH50	** p = 0.005	** p < 0.001
WLCH_OCO_L	WCWD12	** p < 0.001	** p < 0.001
WLCH_OCO_L	WCWD50	** p = 0.002	** p < 0.001

WLCH_OCA_O	WSTPH25	* p = 0.019	n.s. (p = 0.173)
WLCH_OCA_O	WSTPH50	n.s. (p = 0.246)	n.s. (p = 0.476)
WLCH_OCA_O	WCWD12	** p = 0.002	* p = 0.017
WLCH_OCA_O	WCWD50	n.s. (p = 0.161)	n.s. (p = 0.568)
WLCH_OCA_S	WSTPH25	n.s. (p = 0.957)	n.s. (p = 0.846)
WLCH_OCA_S	WSTPH50	n.s. (p = 0.191)	n.s. (p = 0.892)
WLCH_OCA_S	WCWD12	** p < 0.001	** p < 0.001
WLCH_OCA_S	WCWD50	** p = 0.002	* p = 0.026
WLCH_OCA_L	WSTPH25	** p < 0.001	** p < 0.001
WLCH_OCA_L	WSTPH50	** p < 0.001	** p = 0.002
WLCH_OCA_L	WCWD12	** p < 0.001	** p < 0.001
WLCH_OCA_L	WCWD50	+ (p = 0.068)	** p < 0.001

NT = No test; insufficient sample size (number of inventory plots).
n.s. = t-test could not be conducted; non-significant difference ($P > 0.10$)
+ $0.05 \leq p < 0.10$ (marginally significant difference)
* $0.01 \leq p < 0.05$ (significant difference)
** $p < 0.01$ (highly significant difference)

Appendix -- Two programs (for S+ and SAS software programs) for calculating distribution-free tolerance limits. The programs calculate quantiles (observation order number, or k) for a range of values of n (sample size), for specified values of γ and β . Actual values of a random variable that are associated with the quantiles need to be determined by the user by applying them to the actual dataset. The programs were written by Manuela Huso (statistician, Forest Science Department, Oregon State University).

(1) S+ code:

Arguments: maximum sample size possible, i.e. n .

lowest β of interest

intermediate β of interest

largest β of interest

γ , i.e. level of certainty

vector of sample sizes of interest, say (10,12,15,20,25,50,100)

```
tolint <- function(maxsamp, beta1, beta2, beta3, gamma, sampsize)
{
  k <- matrix(rep(0, 4 * length(sampsize)), ncol = 4)
  k[,1]<-sampsize
  dimnames(k)<-list(NULL,c("sample size", "k:beta1", "k:beta2", "k:beta3"))
  n <- seq(1:maxsamp + 1) - 1
  for(i in 1:length(sampsize)) {
    k[i, 2] <- match(T, pbinom(n, sampsize[i], beta1) > gamma) - 1
    k[i, 3] <- match(T, pbinom(n, sampsize[i], beta2) > gamma) - 1
    k[i, 4] <- match(T, pbinom(n, sampsize[i], beta3) > gamma) - 1
  }
  return(k)
}
```

Example:

USAGE

`pbinom(q, size, prob)`

OPTIONAL ARGUMENTS

q vector of (positive) quantiles (number of successes obtained in size binomial trials with probability `prob` of success). Missing values (NAs) are allowed.

size vector of (positive integer) numbers of coin flips for which the Binomial distribution measures the number of heads.

prob vector of probabilities of a head. If `length(n)` is larger than 1, then `length(n)` random values are returned.

```
> n<-seq(1:(20+1))-1
```

```

> n
[1] 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
> pbinom(n, 15, .3)
[1] 0.0047476 0.0352676 0.1268277 0.2968679 0.5154911 0.7216214 0.8688574
0.9499875 0.9847575
[10] 0.9963475 0.9993278 0.9999083 0.9999913 0.9999995 1.0000000 1.0000000
1.0000000 1.0000000
[19] 1.0000000 1.0000000 1.0000000
> maxs
[1] 100
> b1
[1] 0.3
> b2
[1] 0.5
> b3
[1] 0.8
> gam
[1] 0.9
> samp
[1] 10 12 15 20 50 100
> tolint(maxs, b1, b2, b3, gam, samp)
      sample size k:beta1 k:beta2 k:beta3
[1,]      10      5      7      10
[2,]      12      6      8      11
[3,]      15      7     10     14
[4,]      20      9     13     18
[5,]      50     19     30     44
[6,]     100     36     56     85

```

(2) SAS code: These are SAS macros to calculate quantiles for tolerance intervals from samples of any size.

* TOLINT.SAS - Macros to calculate quantiles for tolerance intervals

* from samples of any size*;

%MACRO *tolint*(maxsamp, beta1, beta2, beta3, gamma);

/* Arguments:

maximum sample size possible, i.e. n.

lowest beta of interest

intermediate beta of interest

largest beta of interest

gamma, i.e. level of certainty

* OUTPUT file is called TOLLIM

* with the following variables:

* samp_sz = sample size

```

* quant_b1 = observation whose value you are gamma% certain that
*   beta1% of the population will have values less than its
* quant_b2 = observation whose value you are gamma% certain that
*   beta2% of the population will have values less than its
* quant_b3 = observation whose value you are gamma% certain that
*   beta3% of the population will have values less than its
*/
data tollim (keep=samp_sz quant_b1 quant_b2 quant_b3);
  array ss[4] (0 0 0 0);
  array b1b[0:&MAXSAMP] (0 (&MAXSAMP*0));
  array b2b[0:&MAXSAMP] (0 (&MAXSAMP*0));
  array b3b[0:&MAXSAMP] (0 (&MAXSAMP*0));
/* PROBBNML(p,n,m)
Arguments
p is a numeric probability of success parameter. 0<=p<=1
n is an integer number of independent Bernoulli trials parameter. n>0
m is an integer number of successes random variable. 0<=m<=n
*/
  do i=3 to &maxsamp;
    ss[1]=i;
    do m=0 to i;
      b1b[m]=probbnml(&BETA1, i, m);
      b2b[m]=probbnml(&BETA2, i, m);
      b3b[m]=probbnml(&BETA3, i, m);
      if m>0 then do;
        if (((b1b[m-1]<&gamma) - (b1b[m]<&gamma))=1) then ss[2]=m;
        if (((b2b[m-1]<&gamma) - (b2b[m]<&gamma))=1) then ss[3]=m;
        if (((b3b[m-1]<&gamma) - (b3b[m]<&gamma))=1) then ss[4]=m;
      end; end;
    samp_sz=ss[1]; quant_b1=ss[2]; quant_b2=ss[3]; quant_b3=ss[4]; output;
  end;
run;
%MEND;

%MACRO TOLLEVEL(sampsize);
/* Arguments: sample size for which tolerance limits are desired
* OUTPUT file is tollev with the same variables as tollim
* described above. It will have only one observation,
* the tolerance limits for the particular sample size
* given as an argument. %TOLINT must have been run prior
* to this call with a large enough max sample size
*/
data tollev;
  set tollim;
  if samp_sz=&sampsize;
run;

```

```
proc print data=tollev;  
run;  
%MEND;
```

* Example:

What are the quantiles for the tolerance limits for sample sizes
from 3 to 10, at three levels of $\beta=0.3, 0.5, 0.8$
respectively, with 90% certainty

```
*;  
%tolint(10, .3, .5, .8, .9);  
proc print data=tollim;  
%tolint(20, .3, .5, .8, .9);  
proc print data=tollim;  
%tolint(500, .3, .5, .8, .9);  
%tolint(2500, .3, .5, .8, .9);  
* DON'T print this one :-) *;
```

* What are the quantiles for the tolerance limits
when a sample of 977 was taken?

```
*;  
%tollevel(977);
```

It is a quirk of statistics that TIs are seldom discussed or presented in standard statistical textbooks (but see Mood et al. 1974), although their basis is well represented in statistical theory and in the published statistical literature and is used in ecological practice (e.g., Dix and Strickland 1986). The term “tolerance limit” also has been used in ecological literature mostly to refer to limits of physiological stress from environmental conditions. We show 0 as the closed lower limit of this interval because values of the parameters of interest here -- snag or down wood sizes or amounts – cannot be negative. When this happened, we provided further biological interpretation in the narrative; obviously zero snags or down wood would provide only for zero associated species.

To verify this approach, we also compared the table values from Hahn and Meeker against values we generated from StatCalc, a freeware statistical computer program that calculates one-sided tolerance limits (Krishnamoorthy 2000). Results were identical.