

**Sault Ste. Marie Tribe of Chippewa Indians  
Inland Fish and Wildlife Department**

**Technical Report TR16.01  
September 2016**

**Addressing biases associated with spring spawning ground sampling of adult  
walleye populations throughout the 1836 Treaty Ceded Territory**

John H. Powell  
and  
Eric M. Clark

Suggested Citation Format

Powell, J.H., and E.M. Clark. 2016. Addressing biases associated with spring spawning ground sampling of adult walleye populations throughout the 1836 Treaty Ceded Territory. Sault Ste. Marie Tribe of Chippewa Indians Inland Fish and Wildlife Department, Technical Report TR16.01, Sault Ste. Marie, MI.

## ***Introduction***

The draft walleye sampling protocol for the 1836 Treaty Ceded territory seeks to mark and recapture adult fish on the spawning grounds. This sampling protocol is designed to provide an estimate of the walleye population on the spawning grounds. Among parties to the 2007 Inland Consent Decree there is debate as to whether these spawning ground estimates are also estimates of the entire adult population in a lake. Data indicate that walleye spawn annually, except in some northern populations due to shortened growing seasons (Bozek *et al.* 2011). However, in these populations spawning is not always annual due to entire lake failures (Scott and Chapman 1973). Therefore, across the 1836 Treaty Ceded territory for a spawning ground estimate to be valid it must also estimate of the total number of adult walleye in the lake. Differences in the sex-ratio of walleye between spring and fall sampling, however, indicate that the current spring protocol fails to accurately characterize adult walleye populations.

The sampling protocol specifies that the Chapman modification of the Peterson estimator (Chapman 1951) be used to calculate the adult walleye abundance. This estimator requires the following assumptions: 1) constant mortality of marked and unmarked fish, 2) constant vulnerability of marked and unmarked fish, 3) no loss of marks, 4) random mixing of marked and unmarked fish (or proportional sampling relative to marking), 5) all marks are recognized, 6) no recruitment (Ricker 1975). Behavioral differences between male and female walleye on the spawning grounds (Bozek *et al.* 2011), however, indicate that fish are unlikely to have the same probability of capture (assumption 2) or recapture (assumption 4) when sampled in the spring. Furthermore,

When capture probabilities are heterogeneous, population estimates will be negatively biased (Plante *et al.* 1998). This negative bias can be observed in the population estimates from the 1837 and 1842 Treaty Ceded territories whether looking at the number of estimated walleye in a lake (Figure 1) or the density of walleye in the lake (Figure 2). A rich body of literature addresses the effects of relaxing the assumptions of these models (see Otis *et al.* 1978, Pollock *et al.* 1990, Schwarz and Seber 1999 for reviews on this subject). If capture probabilities differ in a predictable way, stratifying individuals into groups that reduce the heterogeneity in capture probabilities can at times correct for negative bias (Pollock *et al.* 1990). However, not all samples recapture the required number of individuals within each stratum to allow for use of this potentially more effective method.

The Inland Fish and Wildlife Department of the Sault Ste. Marie Tribe of Chippewa Indians is proposing two alternate sampling methods and one alternate analysis method to address this small sample problem. First, sampling could be moved to the fall. Second, if spring sampling is preferred individuals should receive unique identifiers so that individual capture histories can be built. Finally, a correction factor is investigated using data from lakes in the 1837 and 1842 Treaty Ceded territories.

Fall sampling of walleye has the potential to avoid some of the pitfalls of spring sampling. First, fall sampling results in a more even sex-distribution of walleye than sampling during other times of the year (Isermann and Parsons 2011). Thus, utilizing methods where sex cannot be determined (i.e., nonlethal sampling) should not result in as biased of estimates as spring sampling. One potential issue would be the inclusion of legal length immature fish. However, these fish are potentially included in spring sampling as well, indicating this is not a novel source of potential bias. Second, fall sampling allows for increased recruitment of younger age classes to the gear (Isermann and Parsons 2011). This increased recruitment should result in a better estimate of the spawning population vulnerable to the spring spawning closure fishery. Finally, gill nets are the most common gear used to sample walleye (Isermann and Parsons 2011), indicating fall sampling may provide more opportunity for range-wide comparisons of walleye stocks.

As an alternative to sampling walleye populations in the fall, the spring protocol could be modified so that biologists could account for heterogeneity in capture probabilities. By individually marking fish, capture histories can be built for all individuals handled during a survey. Multiple analysis packages are available for selecting and fitting the many closed population size estimators available for this data (e.g., Mark, White and Burnham 1999; Rcapture, Rivest and Baillargeon 2014).

Because methods to account for heterogeneity in capture probabilities require knowledge of individual capture histories (see Huggins 1989 for an example), these methods could be cost or time prohibitive. A less desirable alternative that may potentially reduce the negative bias in population estimates is to examine and model bias in previously estimated population estimates. This bias correction factor can then be applied to help adjust for this negative bias in these estimates.

Here we use data collected in the 1837 and 1842 Treaty Ceded Territories to estimate the bias associated with estimating population abundance using a samples where all sexes are pooled into a common estimate. We present a potential correction factor for pooled spring population estimates in the 1836 Treaty Ceded Territory. Finally, a potential method to construct confidence intervals around these adjusted estimates is provided in the appendix.

## ***Methods***

Population estimates used in this analysis were collected on 205 lakes in the 1837 and 1842 Treaty Ceded Territories between 1990 and 2015. For each lake that was sampled in a given year, an estimate of the number of adult fish was made after stratifying the sample by sex, as well as with all fish pooled into a common estimate. Therefore, only those samples that had a minimum of three male and three female recaptures were used in this analysis (M. Leuhring *personal communication*). In addition to the sample size requirements, to be included in this analysis the coefficients of variation for each estimate were required to be below 0.3, and the

estimate was required to have been accepted to set harvest quotas on the lake. These criteria resulted in a total of 455 population estimates being available for this analysis.

Before analysis we examined the strength of the linear relationship between estimated abundance from stratified and pooled samples (Figure 1). Due to the presence of potential leverage points in these relationships that are still present after taking the logarithm of both population estimates, the relationship between densities (abundance divided by surface area) was also examined (Figure 2). As there were no apparent leverage points in the density relationship, this was selected for analysis.

A multilevel model was fit due to its ability to account for repeated measures, and make predictions in new groups not included in the original sample (Gelman and Hill 2007). Following formulation presented in Gelman and Hill (2007) after taking the logarithm of both the stratified density estimate ( $y_i$ ) and the pooled density estimate ( $x_i$ ), and centering the pooled density estimate, the model

$$\log(y_i) \sim N(\alpha_{j[i]} + \beta * (\log(x_i) - \overline{\log(x_i)}), \sigma_{\log(y)}^2), \quad \text{for } i = 1, \dots, 455$$

$$\alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2), \quad \text{for } j = 1, \dots, 205$$

was fit with the *lme4* package (Bates *et al.* 2015) in the statistical computing program R version 3.0.1 (R Core Team 2013). The level of pooling occurring this model was assessed by running a classic regression analysis, as well as a no-pooling model where each lake was modeled separately. Following model fit quantile prediction intervals were constructed from distributions of 10,000 simulated values of the lake level error term  $\alpha$  and the density  $y$  for each of 1,000 predicted densities using functions provided in the *arm* package (Gelman *et al.* 2015).

## ***Results***

It is estimated that the association between the pooled adult walleye population density and the population density from samples stratified by sex (Figure 3) is

$$e^{(1.0397 + 1.0039 * (\log(\text{Pooled Density}) - 0.9783))}$$

The lake level error standard deviation is estimated to be 0.0496, and the individual population density estimate error standard deviation is estimated to be 0.12. This indicates that the standard deviation of population estimates between lakes is equivalent to the standard deviation of six population estimates within a lake. Because there are a small number of revisits in many of the lakes, the regression line falls closer to the classic regression (complete pooling) analysis in many of the lakes (Figure 4).

## *Literature Cited*

- Bates, D., M. Maechler, B. Bolker, and S. Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67:1-48.
- Bozek M.A., D.A. Baccante, and N.P. Lester. 2011. Walleye and sauger life history. Pages 233-301 *in* B.A. Barton, editor, *Biology, Management, and Culture of Walleye and Sauger*. American Fisheries Society, Bethesda, Maryland.
- Chapman, D.G. 1951. Some properties of the hypergeometric distribution with applications to zoological sample censuses. *University of California Publications in Statistics* 1:131-160.
- Gelman, A. Y.-S. Su, M. Yajima, J. Hill, M.G. Pittau, J. Kerman, T. Zheng, and V. Dorie. 2015. Data analysis using regression and multilevel/hierarchical models. URL <http://CRAN.R-project.org/package=arm>.
- Gelman, A., and J. Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, New York, New York.
- Huggins, R.M. 1989. On the statistical analysis of capture experiments. *Biometrika* 76:133-140.
- Isermann, D.A., and B.G. Parsons. 2011. Harvest regulations and sampling. Pages 403-423 *in* B.A. Barton, editor, *Biology, Management, and Culture of Walleye and Sauger*. American Fisheries Society, Bethesda, Maryland.
- Otis, D.L., K.P. Burnham, G.C. White, and D.R. Anderson. 1978. Statistical inference from capture data on closed animal populations. *Wildlife Monographs* 62:3-135.
- Plante, N., L.-P. Rivest, and G. Tremblay. 1998. Stratified capture-recapture estimation of the size of a closed population. *Biometrics* 54:47-60.
- Pollock, K.H., J.D. Nichols, C. Brownie, and J.E. Hines. 1990. Statistical inference for capture-recapture experiments. *Wildlife Monographs* 107:3-97.
- R Core Development Team. 2013. *R a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Ricker, W.E. 1975. *Computation and Interpretation of Biological Statistics of Fish Populations*. Bulletin of the Fisheries Research Board of Canada 191, Ottawa, Canada.
- Rivest, L.-P., and S. Baillargeon. 2014. Rcapture: loglinear models for capture-recapture experiments. R package version 1.4-2. <http://CRAN.R-project.org/package=Rcapture>
- Schwarz, C.J., and G.A.F. Seber. 1999. Estimating animal abundance: review III. *Statistical Science* 14:427-456.

Scott, W.B., and E.J. Crossman. 1973. Freshwater fishes of Canada. Fisheries Research Board of Canada Bulletin 184, Ottawa.

White, G.C., and K.P. Burnham. 1999. Program MARK: survival estimation from populations of marked animals. Bird Study 46 Supplement:120-138.

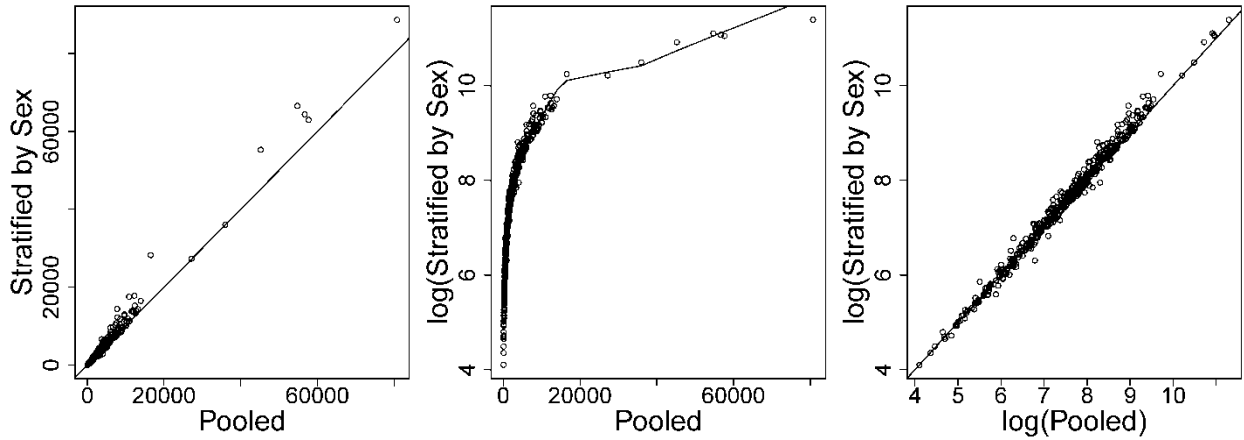


Figure 1. Estimated abundance when all fish are pooled versus estimates calculated when samples are stratified by sex. The left plot depicts the raw data with a one-to-one line added for reference. The middle plot depicts the relationship between the logarithm of the stratified estimate versus the untransformed pooled estimate with a lowess line for reference. The right plot depicts the relationship after taking the logarithm of both variables with a one-to-one line for reference.

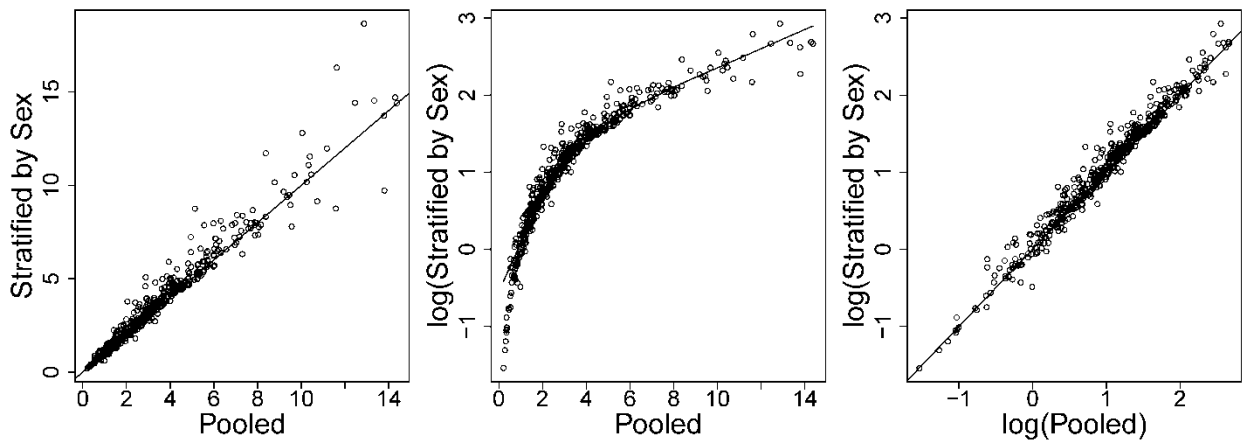


Figure 2. Estimated density when all fish are pooled versus estimates calculated when samples are stratified by sex. The left plot depicts the raw data with a one-to-one line added for reference. The middle plot depicts the relationship between the logarithm of the stratified estimate versus the untransformed pooled estimate with a lowess line for reference. The right plot depicts the relationship after taking the logarithm of both variables with a one-to-one line for reference.

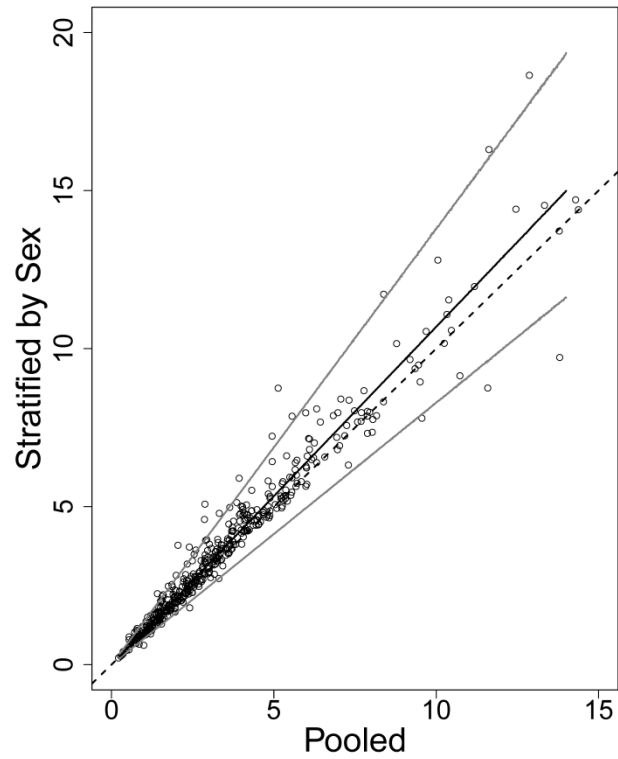


Figure 3. Comparison of pooled density estimates with density estimates stratified by sex. The solid line depicts the mean density in a stratified sample from the multilevel regression model, with gray lines plotting the 95% prediction interval. The dashed line represents what would be expected if there was no difference between the two estimates.



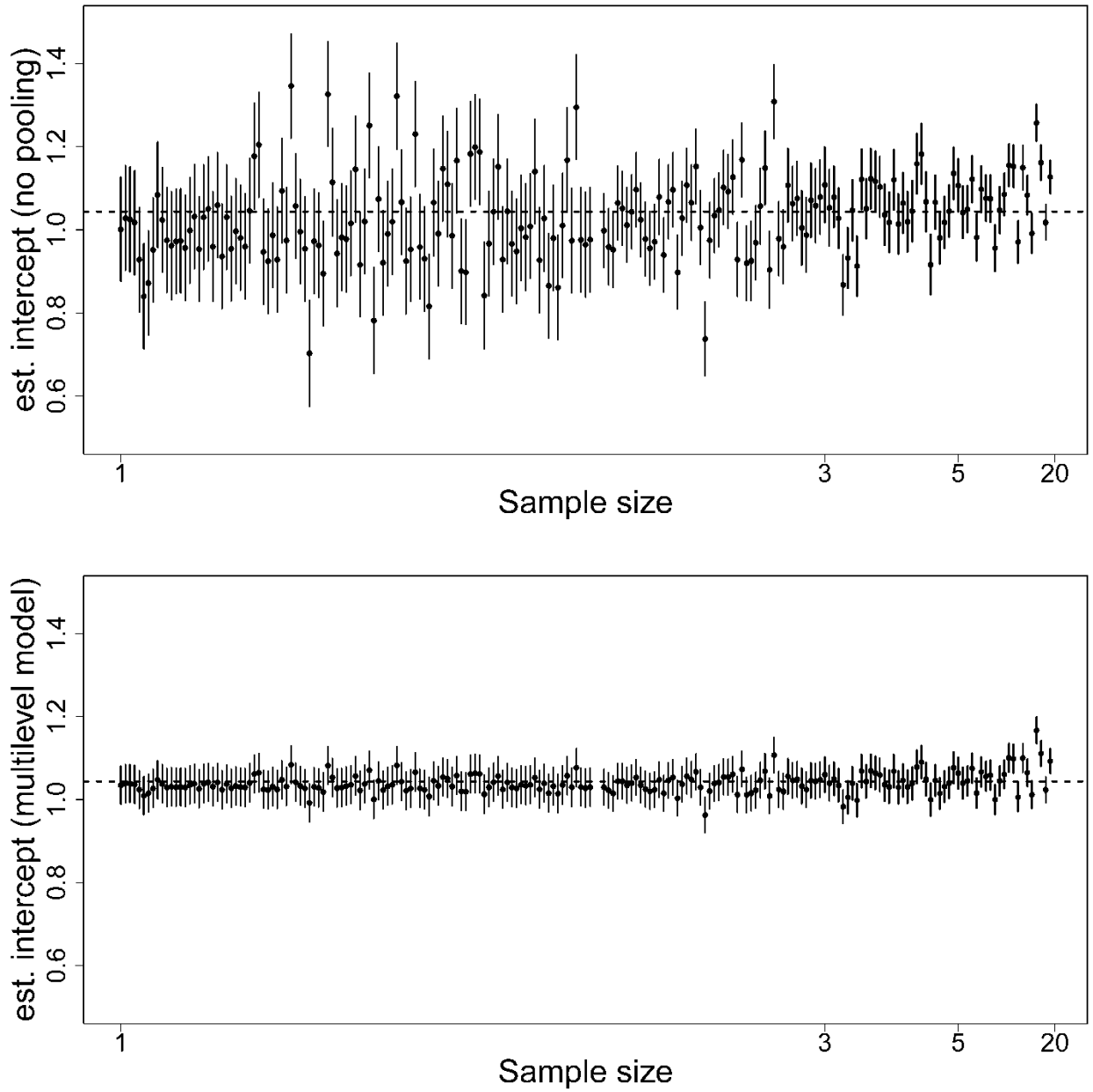


Figure 4. Estimated intercepts ( $\pm$  standard errors) from a model where each lake was estimated separately (top panel) and a multilevel model with a random intercept (bottom panel) are plotted versus the sample size of population estimates available in each lake. The dashed line in each panel depicts the estimated intercept from a classic regression model that does not account for lake.

## Appendix A

Confidence intervals for adjusted adult walleye population estimates can be calculated using a combination of bootstrap resampling and Monte Carlo simulation with the following protocol.

1. Resample the recapture sample to build a bootstrap distribution of the number of recaptured individuals (1000 samples should be taken for 90% confidence intervals, 5000 for 95% confidence intervals, 10000 for 99% confidence intervals; Manly 2007).
2. Calculate a bootstrap distribution of population densities using the Chapman modification of the Peterson formula and dividing by lake surface acres.
3. Simulate the same number of random intercepts as bootstrap resamples from the following distribution.

$$\mathbf{a} \sim N(\boldsymbol{\mu} = 1.039718, \sigma = 0.04955456)$$

4. Simulate the same number of stratified densities as bootstrap resamples from the following distribution.

$$N(\boldsymbol{\mu} = \mathbf{a} + 1.003926 * (\log(\mathbf{B}) - 0.9782681), \sigma = 0.1200354)$$

Where  $\mathbf{B}$  is the one column matrix of bootstrap densities, and  $\mathbf{a}$  is the one column matrix of simulated intercepts

5. Take the appropriate quantile of the resulting distribution and back transform to the original scale.

Provided is a function to calculate the corrected population estimate and a specified confidence interval for the corrected estimate based on the above steps in the statistic computing language R. The variables requiring specification in this function are as follows:

*mark* – The total number of fish marked during the marking phase of the assessment

*capture* – The total number of fish captured during the recapture phase of the assessment

*recapture* – The number of marked fish recaptured during the recapture phase of the assessment

*area* – The surface area of the lake in acres

*alphaLevel* – The desired alpha level (*e.g.*, 0.05 for 95% confidence intervals)

*nSamples* – The number of bootstrap resamples if a number is desired differing from the recommendations of Manly (2007).

```
correctionFactor <- function(mark, capture, recapture, area, alphaLevel = 0.05, nSamples = NA)
{
  if (is.na(nSamples) == FALSE) {
    nSamp <- nSamples
  }
  if (is.na(nSamples) == TRUE) {
    nSamp <- ifelse(alphaLevel == 0.1, 1000, ifelse(alphaLevel == 0.05, 5000,
10000))
  }
}
```

```

    }
    bootstrapRecapture <- colSums(replicate(nSamp, sample(c(rep(1, recapture), rep(0,
capture - recapture)), capture, replace = TRUE), simplify = "matrix"))
    chapmanPeterson <- ((mark + 1) * (capture + 1)) / (bootstrapRecapture + 1)
    aVect <- rnorm(nSamp, mean = 1.039718, sd = 0.04955456)
    yVect <- rnorm(nSamp, mean = aVect + 1.003926 * (log(chapmanPeterson / area) -
0.9782681), sd = 0.1200354)
    confidenceInterval <- exp(quantile(yVect, prob = c(alphaLevel / 2, 1 - alphaLevel / 2))) *
area
    est <- exp(1.039718 + 1.003926 * (log((((mark + 1) * (capture + 1)) / (recapture + 1)) /
area) - 0.9782681)) * area
    list(lowerBound = confidenceInterval[1], estimate = est, upperBound =
confidenceInterval[2])
  }

```

The following is an example of how to use this function based on Sault Ste. Marie Tribe of Chippewa Indian's 2015 adult walleye population estimate on Brevoort Lake in Mackinac County, Michigan.

```
correctionFactor(mark = 1338, capture = 130, recapture = 30, area = 4315)
```