Constructing better coverage intervals for estimators computed from a complex sample survey

Phillip S.Kott[1]
[1]RTI International; 6110 Executive Blvd; Rockville, MD 20852

Abstract
Coverage intervals for a parameter estimated from a complex probability sample are usually constructed by assuming that the parameter estimate has an asymptotically normal distribution, and the measure of the estimator's variance is roughly chi-squared. The size of the sample and the nature of the parameter being estimated render this conventional "Wald" methodology dubious when constructing coverage intervals, especially for proportions. A revised method of coverage-interval construction has been developed in the literature that "speeds up the asymptotics" by incorporating an estimated skewness measure. We will discuss how skewness-adjusted coverage intervals can be computed in some common situations and why it may be inappropriate to call them "confidence intervals."

Key Words: Wald coverage interval, Skewness, Skewness-adjusted coverage interval

1. Introduction

Statisticians are interested in estimating intervals likely to contain a parameter. Wald intervals are most commonly used for this purpose. Hypothesis test for the location of the parameter can be conducted using these intervals.

Suppose $\hat{t}$ is a nearly (i.e., asymptotically) unbiased estimator for a parameter t estimated with data drawn via a probability survey. The one-sided Wald coverage intervals for t are

$$t \le \hat{t} + \Phi^{-1}(\alpha)\sqrt{v} \quad \text{and} \quad t \ge \hat{t} - \Phi^{-1}(\alpha)\sqrt{v}, \tag{1}$$

where v is a good estimator for V the variance of $\hat{t}$ under either probability-sampling theory or a reasonable model, and
$\Phi(.)$ is the cumulative distribution function of a standard normal distribution.

It is well known that when the sample size is large enough, both inequalities hold for roughly α-percent of samples drawn using the same sampling design as the probability survey.

A symmetric two-sided α-percent Wald interval easily derivable from equation (1) is

$$\hat{t} - \Phi^{-1}([1+\alpha]/2)\sqrt{v} \le t \le \hat{t} + \Phi^{-1}([1+\alpha]/2)\sqrt{v}.$$

We will focus on one-sided intervals because creating a symmetric two-sided interval from two one-sided intervals is easily done.

Much of the research on interval estimation has concentrated on proportions either estimated from an independent and identically distributed (*iid*) sample (e.g., Clopper and Pearson 1934, Hall 1998, Newcombe 1998, Brown et al. 2001, Cai 2004) or a complex probability sample (e.g., Korn and Graubard 1998, Kott and Liu 2009, Franco et al. 2014). Liu and Kott (2009, p. 575) displays how well many of the alternative methods of computing coverage intervals perform for a proportion estimated with a simple random sample (a link to that article is in the references).

Here, we will look at more general estimators computed from complex probability samples; in particular, we focus on estimators for ratios and for differences between domain mains. Critical to this endeavor will be estimating the third central moment of $\hat{t}$. We will use probability-sampling (design-based) theory in this effort. Analogous arguments assuming models are straightforward.

Often the sample size in an application will not be nearly large enough for a one-sided Wald interval to contain ("cover") t with the frequency suggested by the asymptotic theory. We will use the term "coverage interval" here rather than "confidence interval" because one rarely has confidence that whatever the true value of t it falls within the designated interval at least α percent of the time across repeated realizations of the random variable $\hat{t}$ (as it would were $\hat{t}$ normally distributed and $\Phi(.)$ replaced with the appropriate Student's *t*-distributio*n*).

Kott and Liu (2010) proposed using skewness-adjusted one-sided coverage intervals in place of the Wald intervals to "speed up the asymptotics:"

$$t \le \hat{t} + \delta + \sqrt{z^2 v + \delta^2} \quad \text{and} \quad t \ge \hat{t} + \delta - \sqrt{z^2 v + \delta^2}, \tag{2}$$

where $\delta = \frac{1}{6}(1-z^2)\frac{m_3}{v} + \frac{z^2}{2}b,$ (3)

$z = \Phi^{-1}(\alpha)$,
m_3 is a nearly unbiased estimator for the third central moment of $\hat{t}$,
$M_3 = E[(\hat{t}-t)^3]$, and
b is a nearly unbiased estimate or for $B = E[v(\hat{t}-t)]/V$, the regression of v on $\hat{t}-t$.

In δ	
$\frac{z^2}{2}b$	accounts for v varying with $\hat{t}-t$
$\frac{1}{6}(1-z^2)\frac{m_3}{v}$	accounts for $\hat{t}$ being skewed.

If $b \approx m_3/v$, which is true in many contexts, then

$$\delta \approx \left(\frac{1}{6} + \frac{z^2}{3}\right)\frac{m_3}{v}. \tag{4}$$

When equation (4) holds, the coverage intervals in equation (2) can be expressed as

$$t \le \hat{t} + \left\{ \left(\frac{1}{6} + \frac{z^2}{3} \right) \hat{\tau} + z \left[1 + \frac{1}{z^2} \left(\frac{1}{6} + \frac{z^2}{3} \right)^2 \hat{\tau}^2 \right]^{1/2} \right\} \sqrt{v} \text{ and} \tag{5}$$

$$t \ge \hat{t} + \left\{ \left(\frac{1}{6} + \frac{z^2}{3} \right) \hat{\tau} - z \left[1 + \frac{1}{z^2} \left(\frac{1}{6} + \frac{z^2}{3} \right)^2 \hat{\tau}^2 \right]^{1/2} \right\} \sqrt{v},$$

where $\hat{\tau} = m_3 / v^{3/2}$ is the estimated skewness for $\hat{t}$, and
$\tau = M_3 / V^{3/2}$ is the skewness measure being estimated.

The estimated skewness tends to decrease in absolute value as the sample size increases. For an estimated proportion, p, under either simple random sampling with replacement (or an *iid* model):

$$m_3 = p(1-p)(1-2p) / [(n-1)(n-2)],$$
$$b = (1-2p) / (n-1), \text{ and}$$
$$\hat{\tau} \approx [(1-2p) / (np(1-p))]^{1/2}.$$

When v is not too close to zero, the following modification on equation (5) drops terms of a smaller asymptotic order (rendering $\hat{\tau}^2 \approx 0$):

$$\begin{aligned} &t \le \hat{t} + \left\{ \left(\frac{1}{6} + \frac{z^2}{3} \right) \hat{\tau} + z \right\} \sqrt{v} \text{ and } t \ge \hat{t} + \left\{ \left(\frac{1}{6} + \frac{z^2}{3} \right) \hat{\tau} - z \right\} \sqrt{v} \text{ or} \\ &t \le \hat{t} + \left(\frac{1}{6} + \frac{z^2}{3} \right) b + z\sqrt{v} \quad \text{and } t \ge \hat{t} + \left(\frac{1}{6} + \frac{z^2}{3} \right) b - z\sqrt{v} \end{aligned} \tag{6}$$

These are the one-sided Wald intervals shifted by
$(1/6 + z^2/3)\ \hat{\tau}\sqrt{v} = (1/6 + z^2/3)b$.
$b = O_P(1/n)$; $\sqrt{v} = O_P(1/\sqrt{n})$.

2. A Stratified Simple Random Sample

Suppose we are interested in constructing one-sided coverage intervals for a finite-population total or mean based on a stratified simple random sample. The former can be expressed as $T_y = \sum^H \sum_{k \in U_h} y_k$, where y_k is the variable of interest for element k.

The corresponding population mean is $\bar{Y} = T_y / N = \sum^H W_h \bar{Y}_h$, where $N = \sum^H N_h$, $W_h = N_h / N$, and $\bar{Y}_h = N_h^{-1} \sum_{k \in U_h} y_k$.

An unbiased estimator for the finite-population total T_y using probability-sampling theory is $\hat{T}_y = N t_y$, where $t_y = \bar{y} = \sum^H W_h \bar{y}_h$ and $\bar{y}_h = n_h^{-1} \sum_{S_h} y_k$.

When every $n_h \geq 3$, one can construct one-sided coverage intervals for $\bar{Y}$ based on probability-sampling theory by setting

$$v = \sum_{h=1}^{H} W_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{\sum_{k \in S_h} (y_k - \bar{y}_h)^2}{n_h(n_h - 1)},$$

$$m_3 = \sum_{h=1}^{H} W_h^3 \left(1 - \frac{n_h}{N_h}\right)\left(1 - \frac{2n_h}{N_h}\right) \frac{\sum_{k \in S_h} (y_k - \bar{y}_h)^3}{n_h(n_h - 1)(n_h - 2)}, \text{ and} \tag{7}$$

$$b = \frac{\sum_{h=1}^{H} W_h^3 \left(1 - \frac{n_h}{N_h}\right)^2 \frac{\sum_{k \in S_h} (y_k - \bar{y}_h)^3}{n_h(n_h - 1)(n_h - 2)}}{v}.$$

One-sided coverage intervals for T_y are simply N times the analogous intervals for $\bar{Y}$. It is easy to see that when all the $2n_h/N_h$ are small enough to be ignored equation (3) effectively collapses into equation (4).

Assuming some mild conditions, the ratio of two totals, T_x/T_z, can be estimated in a consistent manner using without-replacement-stratified-simple-random-sampling data by $\hat{t}_{x/z} = \hat{t}_x / \hat{t}_z$; that is to say, the difference between $\hat{t}_{x/z}$ and T_x/T_z tends to zero in probability as either H or the n_h grow arbitrarily large.

The variance and third central moment of $\hat{t}_{x/z}$ can be estimated as above with each y_k replaced by the linearized term: $e_k = [x_k - (\hat{t}_{x/z})z_k] / \hat{t}_z$, which is asymptotically indistinguishable from $u_k = [x_k - (t_{x/z})z_k] / \hat{t}_z$. Observe that

$$\hat{t}_{x/z} - t_{x/z} = \sum^H W_h \bar{u}_h.$$

A ratio estimator of special interest is the estimator of a domain mean. If $d_k = 1$ for an element in the domain and 0 otherwise, then the estimated mean of y-values in the domain has the form $\hat{t}_{x/z} = \hat{t}_x / \hat{t}_z$, where $z_k = d_k$, and $x_k = d_k y_k$.

3. A Stratified Multistage Sample

Consider now constructing a coverage interval for a parameter t based on stratified multistage sample when a nearly unbiased estimator for that parameter can be put in the form:

$$\hat{t} = \sum_{h=1}^{H} \frac{1}{n_h} \sum_{i=1}^{n_h} \hat{t}_{hi}, \tag{8}$$

where there are n_h primary sampling units (PSU's) in stratum h, and each $\hat{t}_{hi}$ for a PSU i in stratum h is a nearly unbiased estimator for the same value.

The parameter t may be a model parameter or a finite-population parameter. In the latter case, we make the common (but often inaccurate) assumption that that the PSU's were selected randomly but with replacement, while any subsampling was done using probability-sampling principles. In the former, we assume variables of interest are independent across PSUs and that strata are nuisances.

We focus now on the difference between two domain means estimated using data from the same sample, S. Each element in S had a value y_k and a sampling weight w_k attached to it, so that the estimated different in domains means can be expressed as:

$$\frac{\sum_{k \in S} w_k y_k d_k^{(1)}}{\sum_{k \in S} w_k d_k^{(1)}} - \frac{\sum_{k \in S} w_k y_k d_k^{(2)}}{\sum_{k \in S} w_k d_k^{(2)}} = \bar{y}_{(1)} - \bar{y}_{(2)},$$

where $d_k^{(a)}$ = 1 when k is in domain a and 0 otherwise. Here:

$$\hat{t}_{hi} \approx u_{hi} = \sum\nolimits_{k \in S_{hi}} w_k y_k \left(\frac{d_k^{(1)}}{\hat{N}_1} - \frac{d_k^{(2)}}{\hat{N}_2} \right),$$

where S_{hi} is the set of sampled elements in PSU i of stratum h, and $\hat{N}_a = \sum\nolimits_{q \in S} w_q d_q^{(a)}$ is the estimated population size of domain a, that is, N_a. Observe that the

$$\hat{t}_{hi} \approx u_{hi} = \sum\nolimits_{k \in S_{hi}} w_k y_k \left(\frac{d_k^{(1)}}{N_1} - \frac{d_k^{(2)}}{N_2} \right).$$

are independent under probability-sampling theory for PSUs in the same stratum (recall we are assuming with-replacement sampling in the first stage of sample selection).

When all $n_h \geq 3$, the following equalities can be used in equations (2) and (3):

$$v = \sum_{h=1}^{H} \frac{N_h^2}{n_h} \sum_{i=1}^{n_h} \frac{(e_{hi} - \bar{e}_h)^2}{(n_h - 1)}, \;\; m_3 = \sum_{h=1}^{H} \frac{N_h^3}{n_h} \sum_{i=1}^{n_h} \frac{(e_{hi} - \bar{e}_h)^3}{(n_h - 1)(n_h - 2)}, \tag{9}$$

and $b = \frac{m_3}{v}$,

where each e_{hi} has the following linearized expression:

$$e_{hi} = n_h \sum\nolimits_{k \in S_{hi}} w_k \left(\frac{d_k^{(1)}}{\hat{N}_1} [y_k - \bar{y}_{(1)}] - \frac{d_k^{(2)}}{\hat{N}_2} [y_k - \bar{y}_{(2)}] \right).$$

We ignore finite population correction when comparing domain means because an analyst is usually interested in whether there is an underlying process causing the domain means to be different, not the actual difference between the means in in the finite population.

4. An Example

We now look at computing one-sided coverage intervals for two sets of parameters for the MU284 population from Särndal et al. (1992) available at http://lib.stat.cmu.edu/datasets/mu284. The population consists of 284 Swedish administrative municipalities separated into 50 clusters with 8 strata. We collapse the final two strata into a single seventh stratum. We divide the population into two domains, the 26 municipalities with a 1985 population of over 64,000 are in Domain 1 and the remaining 258 in Domain 2. We are interested in constructing coverage intervals for, 1, the arithmetic average across municipalities in 1985 of the municipal taxation per person within each domain and, 2, the fraction of municipalities within each domain with more tax receipts than 9 million kronor per 1,000 persons in 1985. We are also interested in constructing coverage intervals the differences between the domains.

We suppose a cluster sample of three clusters per stratum (n_h = 3) are selected from the MU284 population via simple random sampling with replacement. Letting y_k be either the tax revenue per person in municipality k or a (0/1) indicator of whether that ratio is greater than 9 million kronor per 1,000 persons, we define

$$e_{hi} = N_h \sum\nolimits_{k \in S_{hi}} \left(\frac{d_k^{(a)}}{\hat{N}_1} [y_k - \bar{y}_{(1)}] \right) \qquad \text{for Domain } a \text{ (1 or 2), and}$$

$$e_{hi} = N_h \sum\nolimits_{k \in S_{hi}} \left(\frac{d_k^{(1)}}{\hat{N}_1} [y_k - \bar{y}_{(1)}] - \frac{d_k^{(2)}}{\hat{N}_2} [y_k - \bar{y}_{(2)}] \right)$$

for the difference between the domains,

where N_h is the number of clusters in stratum h, S_{hi} is the sample of municipalities in cluster i of stratum h (in this example, S_{hi} is every municipality in the cluster), $d_k^{(a)}$ = 1 when municipality k is in Domain a, 0 otherwise, $\bar{y}_{(a)}$ is the estimated mean of the y-values in domain a, and $\hat{N}_a$ is the estimated number of municipalities in domain a.

For constructing coverage intervals in this example, we replace v, m_3, and $\hat{\tau}$ in equation (9) by what they estimate:

$$V = \sum_{h=1}^{7} N_h^2 \frac{\sum_{i=1}^{N_h} (e_{hi} - \bar{E}_h)^2}{n_h (N_h - 1)},$$

$$M_3 = \sum_{h=1}^{7} N_h^3 \frac{N_h \sum_{i=1}^{N_h} (e_{hi} - \bar{E}_h)^3}{n_h^{\,2} (N_h - 1)(N_h - 2)}, \text{ and}$$

$$\tau = M_3 / V^{3/2},$$

where $\bar{E}_h = \sum_{i=1}^{N_h} e_{hi} / N_h$.

By using these replacements, we produce coverage intervals around the respective estimates close to what the average an infinite number of simulations would produce.

TABLE 1: *Coverage Intervals*

Estimates	Target	One-Sided 95% Intervals (Compared to Estimate) Wald	Skewness-adjusted Lower	Skewness-adjusted Upper	One-Sided 99% Intervals (Compared to Estimate) Wald	Skewness-adjusted Lower	Skewness-adjusted Upper
Municipal Tax Per Person (in 1,000s)							
Domain 1	6.88	±1.19	-1.66	0.72	±1.68	-2.54	0.82
Domain 2	6.90	±0.19	-0.20	0.18	±0.27	-0.29	0.25
Difference	-0.02	±1.15	-1.61	0.68	±1.62	-2.48	0.76
Fraction Above 9 Million Per Thousand							
Domain	0.15	±0.17	-0.10	0.24	±0.24	-0.10	0.37
Domain 2	0.02*	±0.02	-0.01	0.03	±0.03	-0.02*	0.04
Difference	0.13	±0.16	-0.09	0.24	±0.23	-0.10	0.37

The average lower (upper) Wald interval is bound from below (above) by the estimate minus (plus) the unsigned value in the table. The average lower (upper) Skewness-adjusted interval is bound from below (above) by the estimate plus the Lower (Upper) value in the table.

* Computed to another digit, the target is 0.023, while the lower bound is -0.018.

Table 1 compares one-sided 95% and 99% Wald coverage intervals to skewness adjusted Skewness-adjusted coverage intervals computed with equation (6).

The symmetric Wald and asymmetric Skewness-adjusted intervals tend to be closer to each other in Domain 2. This is because the larger sample size in Domain 2 reduces the impact of skewness adjustment. The coverage intervals for the differences tend to be dominated by the smaller Domain 1 samples.

5. Some Simple Approximations

A key to skewness-adjusted coverage intervals, especially when finite population correction can be ignored, is the estimated value $b = m_3/v = \hat{\tau}\sqrt{v}$. From the last section, the value of this term for the difference between proportions estimated for two distinction domains from a simple random sample is approximately (assuming a large sample)

$$b = \frac{m_3}{v} \approx \frac{p_1(1-p_1)(1-2p_1)/n_1^2 - p_2(1-p_2)(1-p_2)/n_2^2}{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2} \tag{10}$$

where p_a is the estimated proportion in domain a based on n_a sampled elements being in domain a. When $p_1 = p_2$, this collapses to

$$b \approx (1-2p_1)\left(\frac{1}{n_1} - \frac{1}{n_2}\right).$$

This appears to suggest that when assessing the difference between proportions in two distinct domains, one should multiply the domain sample sizes by their respective design effects; BUT

the design effect captures the impact of clustering, stratification, and unequal weighting on the variance of an estimator, not on its third central moment.

A wiser procedure might be to estimate $B = M_3/V$ for an estimated proportion $p = \sum_{k\in S} w_k y_k / \sum_{k\in S} w_k$, where p estimates the fraction of the population with $y_k = 0$ or 1, and

$$b_{simple} = \frac{\sum_{k\in S} w_k^3}{\sum_{k\in S} w_k \sum_{k\in S} w_k^2}(1-2p), \tag{11}$$

and then inserting $\hat{\tau}_{simple} = b_{simple}/\sqrt{v}$ into equation (5) or (6). This estimate ignores the impact of stratification and clustering on b. For the proportion in a domain, S in equation (11) and in defining p becomes the sample in the domain. For the difference between two domain means: $\hat{\tau}_{simple} = b_{simple}/\sqrt{v}$ with

$$b_{simple} = \frac{p_1(1-p_1)(1-2p_1)/\tilde{n}_1^2 - p_2(1-p_2)(1-p_2)/\tilde{n}_2^2}{p_1(1-p_1)/n_1^* + p_2(1-p_2)/n_2^*}, \tag{12}$$

where $\tilde{n}_a^2 = \dfrac{\left(\sum_{S_a} w_k\right)^3}{\sum_{S_a} w_k^3}$, and $n_a^* = \dfrac{\left(\sum_{S_a} w_k\right)^2}{\sum_{S_a} w_k^2}$.

For a more general population or domain mean of a *y-variable*, one can replace $\hat{\tau}$ in equation (5) or (6) with

$$\hat{\tau}_{simple} = b_{simple} / \sqrt{v}\ , \text{ where } b_{simple} = \frac{\sum_{k\in S} w_k^3 (y_k - \bar{y})^3}{\sum_{k\in S} w_k \sum_{k\in S} w_k^2 (y_k - \bar{y})^2}, \tag{13}$$

and $\bar{y} = \sum_S w_k y_k / \sum_S w_k$.

TABLE 2: *Approximating B*

Estimates	B	B_{simple} (equation (13))	B_{simple} (equation 11 or 12)
Municipal Tax Per Person (in 1,000s)			
Domain 1	-0.436	-0.486	
Domain 2	-0.011	-0.020	
Difference	-0.436	-0.477	
Fraction Above 9 million Per Thousand			
Domain	0.068	0.073	0.067
Domain 2	0.006	0.008	0.010
Difference	0.069	0.072	0.066

Table 2 assesses equations (11) through (13) with the examples from Section 3, replacing $\sum_{k\in S_a} w_k{}^b z_k$ by $\sum_{k\in U_a} w_k{}^{b-1} z_k$.

The approximations are clearly not perfect, but they are closer to the true $B = M_3/V$ than 0, the value implied when Wald intervals are constructed.

The appeal of these equations results from the practical problem of computing m_3, and consequently $\hat{\tau}$, using either equation (7) or (9): there is no available software routine to do so. Even if there were or a statistician wanted to program the equations herself, there may not be three PSUs in every stratum. Unlike collapsing strata for variance estimation, the direction of the potential bias of $\hat{\tau}$ can be positive or negative when the population means of the strata (the expected value of the y_k in each stratum in equation (7) or the expected values of the u_{hi} corresponding to the e_{hi} for a particular h in equation (9)) being combined are different. Consequently, strata collapsed together should have (near) equal expected population means.

6. Calibration Weighting and the Jackknife

Calibration weighting often removes much of the impact of stratification and clustering from an estimated mean. For example, calibration by region can reduce the impact of stratification by geographical units, while calibration by race and ethnicity can reduce the impact of clustering within neighborhoods. As a result, estimating the skewness of an estimated proportion or mean using equations (10) or (11) may not be unreasonable, although it would often be better to replace the y_k with a calibrated residual. Moreover, when estimating domain means, the impact of calibration weighting like that of stratification and clustering is diminished, except for that due to any increased variability of the weights themselves, making the use of equations (11), (12), or (11) within the intervals in equation (5) or (6) more viable.

If calibrated jackknife weights have been constructed to compute v for an estimator $\hat{t}$, then these weights can also be used in estimating the third central moment of $\hat{t}$:

$$m_{3(J)} = \sum_{h=1}^{H} \frac{(n_h - 1)^2}{n_h(n_h - 2)} \sum_{i=1}^{n_h} (\hat{t} - \hat{t}_{(hi)})^3,$$

where $\hat{t}$ is computed with calibrated weights, and $\hat{t}_{(hi)}$ is computed with the calibrated weights for the stratum-*h* PSU-*i* jackknife replicate.

Estimating the variance and third central moment of an estimator whose weights incorporate more than one calibration step can be difficult using the linearization methods of Section 2. Computing jackknife measures is much simpler.

References

Abramovitch, L. and Singh, K. (1985) Edgeworth Corrected Pivotal Statistics and the Bootstrap, *Annals of Statistics* 13: 116-132.

Anderson, P. and Nerman, O. (2000) A Balanced Adjusted Confidence Interval Procedure Applied to Finite Population Sampling, Presented at the Second International Conference of Establishment Surveys, Buffalo, NY.

Brown, L., Cai, T. and Dasgupta.,A. (2001) Interval Estimation for a Binomial Proportion. *Statistical Science,* 16, 101–133. https://projecteuclid.org/download/pdf_1/euclid.ss/1009213286

Cai, T. (2004) One-sided Confidence Intervals in Discrete Distributions. *Journal of Statistical Planning and Inference,* 131, 63–88.

Clopper, C. and Pearson, E. (1934) The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial, *Biometrika* 26, 404–413.

Deville, J-C. and Särndal, C-E. (1992) Calibration estimators in Survey Sampling. *Journal of the American Statistical Association,* 87, 376-382.

Hall, P. (1982) Improving the Normal Approximation When Constructing One-sided Confidence Intervals for Binomial or Poisson Parameters, *Biometrik*a, 69, 647–652.

Franco, C., Little, R. Lewis, T., and Slud. E. (2014) Coverage Properties of Confidence Intervals for Proportions in Complex Sample Surveys, *Proceedings of the ASA Survey Research Methods Section.* Latest version on internet: http://www.math.umd.edu/~slud/s770/SurveyConfidenceIntervals/JSSAM-2017-065-FINAL.pdf.

Korn, E. and Graubard, B. (1998) Confidence Intervals for Proportions with Small Expected Number of Positive Counts Estimated from Survey Data, *Survey Methodology* 24, 193–201.

Kott, P. (2006) Using Calibration Weighting to Adjust for nonresponse and Coverage Errors. *Survey Methodology,* 32, 133–142.

Kott, P. (2017) A Note on Wilson Coverage Intervals for Proportions Estimated from Complex Samples, *Survey Methodology,* 43, 235-240.

Kott, P., Anderson, P. and Nerman, O. (2001) Two-sided Coverage Intervals for Small Proportions Based on Survey Data, Presented at Federal Committee on Statistical Methodology Research Conference, Washington, DC. https://pdfs.semanticscholar.org/ff0a/0a1af716681771d49f555c35304b002d9ff3.pdf

Kott, P. and Liu, Y. (2009) One-sided Coverage Intervals for a Proportion Estimated from a Stratified Simple Random Sample, *Internal Statistical Review,* 77, 251–265.

Kott, P. and Liu, Y. (2010) Speeding Up the Asymptotics When Constructing One-sided Coverage Intervals with Survey Data, *Metron* 63, 137–151. Proceedings version: http://www.asasrms.org/Proceedings/y2009/Files/304414.pdf

Liu, Y. and Kott, P. (2009) Evaluating Alternative One-sided Coverage Intervals for a Proportion, *Journal of Official Statistics,* 25*,* 569–588. https://www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/evaluating-alternative-one-sided-coverage-intervals-for-a-proportion.pdf

Newcombe, R. (1998) Two-sided Confidence Intervals for the Single Proportion: Comparison of Seven Methods. *Statistics in Medicine*, 17, 857–872.