An Application of Confidence Interval Methods for Small Proportions in the Health Care Survey of DoD Beneficiaries

Amang Sukasih and Donsig Jang, Mathematica Policy Research. Inc.
Amang Sukasih, 600 Maryland Ave., SW, Suite 550, Washington, D.C. 20024

Keywords: complex survey; coverage probability; simulation; stratified sample.

Introduction

Parameter estimation is often presented in a form of confidence interval. When data is gathered from a complex survey, confidence interval is usually computed under a normality assumption. However, when the parameter of interest is a proportion, and the estimate of proportion is extremely small or large (closed to zero or to one), this approach shows lack of coverage. Alternatively, different approaches have been suggested, such as binomial approach, exact confidence interval, Poisson approach, Logit transformation approach, and Wilson methods. Our paper will evaluate the performance of these methods under a complex survey setting. Application of these methods will be demonstrated with data from the quarterly Health Care Survey of DoD Beneficiaries (HCSDB). Comparison will be done, and a simulation will be performed to investigate the performance of each method in term of coverage probability.

Confidence Interval Methods

A standard confidence interval for a parameter θ is usually constructed using the formula

$$\left\{\hat{\theta}-t_d(1-\alpha/2)\left[\operatorname{var}(\hat{\theta})\right]^{1/2},\ \hat{\theta}+t_d(1-\alpha/2)\left[\operatorname{var}(\hat{\theta})\right]^{1/2}\right\}$$

where $\hat{\theta}$ is the estimate of θ, $\operatorname{var}(\hat{\theta})$ is the estimate of variance of $\hat{\theta}$, and $t_d(1-\alpha/2)$ is the $(1-\alpha/2)$ quantile of a t distribution with degrees of freedom d. If the parameter of interest is a proportion p, similar construction can be done with $\hat{\theta}=\hat{p}$ and $\operatorname{var}(\hat{\theta})$ = $\operatorname{var}(\hat{p})$. When the sample is a random sample of size n, the degrees of freedom is computed as $d=n-1$. When sample size is large, $t_d(1-\alpha/2)$ may be replaced by $z(1-\alpha/2)$, which is the $(1-\alpha/2)$ quantile of a standard normal distribution. For a sample from complex survey data, similar construction is generally used with $\hat{\theta}$ and $\operatorname{var}(\hat{\theta})$ computed through proper estimation method for complex survey data, and d is the number of sampled Primary Sampling Units (PSUs) minus the number of strata[1].

Construction of the above confidence interval is based on normality assumption that relies on large sample size. When the sample size is small or the estimate of proportion is extremely small (closed to zero) or large (closed to one), however, this approach shows lack of coverage (Korn and Graubard 1998, Kott, Andersson and Nerman 2001). In addition, this formula may produce confidence interval that lies outside the permitted 0–1 range.

In this paper we look at some alternative methods of confidence interval construction for a complex survey data. We evaluate and perform comparison of the following methods:

(1) Poisson (Breeze) approach,
(2) Logit transformation,
(3) Binomial approach,
(4) Ad-hoc Quadratic/Wilson method,
(5) Andersson-Nerman method,
(6) Model-based Wilson method,
(7) *t*-adjusted Andersson-Nerman method.

These seven methods were presented in Korn and Graubard (1998), and Kott, Andersson and Nerman (2001). The next sections will present detail formulas for each of confidence intervals above.

Let $\hat{p}$ and $\operatorname{var}(\hat{p})$, respectively, denote the estimate of proportion and its estimate of variance computed through proper estimation method for complex survey data. The effective sample size n^* is defined as

$$n^*=\frac{n}{deff}=\frac{n}{\dfrac{\operatorname{var}(\hat{p})}{\hat{p}(1-\hat{p})/n}}=\frac{\hat{p}(1-\hat{p})}{\operatorname{var}(\hat{p})},$$

where $deff$ denotes design effect due to complex sample design. With a complex survey data, it is expected[2] that $deff>1$, so that $n^*<n$. Additionally,

[1] If the sample design did not implement clustering, then d is the number of sample minus the number of strata.

[2] Analysis based on a complex survey data is commonly done through the weighted analysis, where

the degrees-of-freedom adjusted effective sample size is defined as

$$n_{df}^{*} = \frac{\hat{p}(1-\hat{p})}{\mathrm{var}(\hat{p})}\left(\frac{t_{n-1}(1-\alpha/2)}{t_d(1-\alpha/2)}\right)^2 .$$

Poisson (Breeze) Approach

The interval is developed based on the $(1-\alpha)$ confidence interval for a Poisson random variable. For data from a complex survey the Breeze approach confidence interval can be constructed as follows (Korn and Graubard 1998):

$$\left\{ po_L(\hat{p}n^*)/n^*, po_U(\hat{p}n^*)/n^* \right\}$$

where

$$po_L(x) = 0.5\chi^2_{v_1}(\alpha/2),$$
$$po_U(x) = 0.5\chi^2_{v_2}(1-\alpha/2),$$
$$v_1 = 2x, \quad v_2 = 2(x+1),$$

and $\chi^2_v(\beta)$ denotes β quantile of a chi-square distribution with v degrees of freedom. When $n^* > n$, one can just use n in place of n^*.

Logit Transformation

The lower and upper bounds of this confidence interval are obtained by using logit transformation as follows:

$$y = \log\left(\frac{p}{1-p}\right)$$

and then put bounds on y as follows:

$$y \pm t_d(1-\alpha/2)\left[\mathrm{var}(y)\right]^{1/2},$$

where

$$\mathrm{var}(y) = \mathrm{var}(\hat{p})\left[\frac{\partial y}{\partial \hat{p}}\right]^2 = \frac{\mathrm{var}(\hat{p})}{\left[\hat{p}(1-\hat{p})\right]^2}.$$

So that

$$\log\left(\frac{\hat{p}}{1-\hat{p}}\right) \pm t_d(1-\alpha/2)\frac{\left[\mathrm{var}(\hat{p})\right]^{1/2}}{\hat{p}(1-\hat{p})}.$$

The confidence interval for p is then obtained by inverting back from y to p as follows:

$$p = \frac{\exp(y)}{1+\exp(y)}.$$

Thus, the logit transformation confidence interval can be expressed as:

$$\left\{ \frac{1}{1+\exp(-LLOGIT)}, \frac{1}{1+\exp(-ULOGIT)} \right\},$$

where

$$LLOGIT = \log\frac{\hat{p}}{1-\hat{p}} - t_d(1-\alpha/2)\frac{[\mathrm{var}(\hat{p})]^{1/2}}{\hat{p}(1-\hat{p})},$$
$$ULOGIT = \log\frac{\hat{p}}{1-\hat{p}} + t_d(1-\alpha/2)\frac{[\mathrm{var}(\hat{p})]^{1/2}}{\hat{p}(1-\hat{p})}.$$

When $n^* > n$, one can just use $\hat{p}(1-\hat{p})/n$ in place of $\mathrm{var}(\hat{p})$.

Binomial Approach

The interval is developed based on the binomial distribution approach. Korn and Graubard (Korn and Graubard 1998) developed similar approach for data from a complex survey by replacing n with n_{df}^* and the number of positive count x with $\hat{p}n_{df}^*$ as follows:

$$\left\{ p_L(\hat{p}n_{df}^*, n_{df}^*), p_U(\hat{p}n_{df}^*, n_{df}^*) \right\}$$

where

$$p_L(x,n) = \frac{v_1 F_{v_1,v_3}(\alpha/2)}{v_3 + v_1 F_{v_1,v_3}(\alpha/2)}$$
$$p_U(x,n) = \frac{v_2 F_{v_2,v_4}(1-\alpha/2)}{v_4 + v_2 F_{v_2,v_4}(1-\alpha/2)}$$

$$v_1 = 2x, \; v_2 = 2(x+1), \; v_3 = 2(n-x+1),$$
$$v_4 = 2(n-x),$$

and $F_{d_1,d_2}(\beta)$ denotes β quantile of F-distribution with (d_1, d_2) degrees of freedom.

Ad-hoc Quadratic/Wilson Method

The confidence interval is derived by solving the quadratic function

$$|\hat{p}-p|^2 \le \left[z(1-\alpha/2)\right]^2\left[p(1-p)/n\right]$$

for p, resulting in lower and upper limits:

(continued)
the weights account for unequal probability sampling, nonresponse adjustment, and/or post-stratification adjustment.

$$\hat{p}+\frac{\frac{z^2}{2n}(1-2\hat{p})}{\left(1+\frac{z^2}{n}\right)}\pm z\sqrt{\frac{z^2}{(2n)^2}+\frac{\hat{p}(1-\hat{p})}{n}}\Bigg/\left(1+\frac{z^2}{n}\right).$$

Simplifying the above bounds by dropping the terms $O_p(n^{-3/2})$, the expression becomes

$$\hat{p}+\frac{z^2}{2n}(1-2\hat{p})\pm z\sqrt{\frac{z^2}{(2n)^2}+\frac{\hat{p}(1-\hat{p})}{n}}\ .$$

For a complex survey data the method replaces n with n^* and the model-based variance $\hat{p}(1-\hat{p})/n$ with a variance from complex data $v(\hat{p})$ as follows:

$$\hat{p}+\frac{z^2}{2n^*}(1-2\hat{p})\pm z\sqrt{\frac{z^2}{(2n^*)^2}+v(\hat{p})}\ .$$

Andersson-Nerman Method

This method replaces the n^* in the ad-hoc Wilson confidence interval with n' as follows

$$\hat{p}+\frac{z^2}{2n'}(1-2\hat{p})\pm z\sqrt{\frac{z^2}{(2n')^2}+v(\hat{p})}\ ,$$

where

$$n'=\frac{(1-2\hat{p})[v(\hat{p})]}{e\{(\hat{p}-P)[v(\hat{p})]\}}$$

and

$$e\{(\hat{p}-P)[v(\hat{p})]\}=\sum_{h=1}^{H}\frac{n_h^2}{(n_h-1)(n_h-2)}\sum_{j\in s_h}\left[u_{hj}-\frac{\sum_{i\in S_h}u_{hi}}{n_h}\right]^3$$

$$u_{hj}=\sum_{k\in s_{hj}}\pi_k(y_k-\hat{p})\ ,$$

$$\pi_k=\frac{w_k}{\sum_s w_k}\ ,$$

and w_k is the survey weight for sample unit *k*, and j is the index for PSUs.

Model-based Wilson Method

Kott, et al. (2001) developed a Wilson-based confidence interval by using model-based variance estimator $\hat{p}(1-\hat{p})/n_M$ in place of $v(\hat{p})$ as follows:

$$\hat{p}+\frac{z^2}{2n_M}(1-2\hat{p})\pm z\sqrt{\frac{z^2}{(2n_M)^2}+\frac{\hat{p}(1-\hat{p})}{n_M}}\ ,$$

where

$$n_M=\frac{1}{\sum_s \pi_k^2}\ .$$

t-adjusted Method

This method replaces the z quantile in the Andersson-Nerman confidence interval with the corresponding *t*-distribution quantile with d_t degrees of freedom, where

$$d_t=\frac{2\hat{p}(1-\hat{p})(\sum_s \pi_k^2)^2}{(1-2\hat{p})^2[\sum_s \pi_k^4-(\sum_s \pi_k^3)^2/\sum_s \pi_k^2]}\ .$$

HCSDB Data

The methods above were implemented to a data from the Health Care Survey of the Department of Defense Beneficiaries (HCSDB), a quarterly survey that collects data on the military health system's beneficiaries' opinions about their Department of Defense health care benefits. The HCSDB sample selection implements a stratified random sampling method, where the beneficiaries are partitioned into sampling strata based on enrollment type (enrolled in TRICARE Prime or not), beneficiary type (active duty, active duty family members, retirees and family members under age 65, and non-active duty beneficiaries and their family members age 65 and over), and geographic areas.

Many statistics of interest are expressed in proportions; for instances, proportion of beneficiaries who would rate their health plan 8 or higher on a 0 to 10 scale, proportion of beneficiaries who would have no problem to get a referral to see a specialist, etc. Currently published statistics are based on quarterly survey data as well as annual combined data. Due to sample size limitation, however, estimates based on quarterly data for the smallest geographical area based on Military Treatment Facility (MTF *catchment area*) are produced based on annual combined data. The estimates based on quarterly data for this fine level of domain face the small- or large-proportion problem.

In this paper, we focus on a catchment-area level estimate of proportion for a selected HCSDB variable HP_SMOKH ("proportion of smokers under HEDIS definition") from the Quarter 1 of the 2003 survey data. In this particular catchment area, the sample size for proportion of smokers is 155 across 5 strata. The weighted estimate of the proportion is 9% with standard error[3] 2.3%. This leads to a coefficient of variation of 25.6%. Table 1 presents the confidence intervals based on the eight methods described in the previous sections.

Table 1. Lower and upper bounds of confidence interval for proportion of smokers by the confidence interval method

Confidence interval method	Lower bound	Upper bound
Normal approximation	4.5%	13.6%
Poisson (Breeze)	5.0%	15.2%
Logit transformation	5.4%	14.8%
Binomial approach	5.1%	14.7%
Ad-hoc Quadratic/Wilson	5.4%	14.7%
Andersson-Nerman	5.4%	14.7%
Model-based Wilson	5.4%	14.8%
t-adjusted Andersson-Nerman	5.4%	14.7%

Source: The HCSDB data, Quarter 1, 2003

The normal-based method resulted in a confidence interval that is shorter than that of other methods. The lower and upper bounds based on other methods are consistent to each other. We will investigate the coverage probability of these methods using a simulation.

Simulation

We generate data to simulate population proportions of, respectively, $p = 0.08$ and $p = 0.60$, under several sample designs below. For each p and each design, we generate data through Bernoulli random number generators with sample of size 100. We replicate data generation for a total of 10,000 times. We then evaluate the coverage probability of confidence intervals and their average lengths. Statistical software SUDAAN (RTI 2004) is used to estimate $\hat{p}$ and $v(\hat{p})$.

Design 1: Simple Random Sample

A binary variable is generated using Bernoulli random number generator with parameter p. Two sample are generated with, respectively, $p = 0.08$ and $p = 0.60$. An equal sampling weight of 95 is assigned to every case. To simulate weighting adjustment to account for nonresponses, the sampling weights are further inflated by factors: 0%, 5%, 10%, 15% or 20%, where these factors reflect differential nonresponse rates within five adjustment cells. Membership of being in the adjustment cells is randomized with probabilities 0.4, 0.25, 0.15, 0.12, and 0.08, of being in the cell 1, 2, 3, 4 and 5, respectively.

Design 2: Stratified Random Sample, Proportional Allocation

Two sets of stratified random sample from 3 strata are generated, where the binary variable is generated using the following parameters:

	$p = 0.08$	$p = 0.60$		
Stratum	p_h	p_h	N_h	n_h
1	10%	70%	4,000	42
2	7.5%	60%	3,500	37
3	5%	40%	2,000	21

where N_h and n_h, respectively indicates the stratum population and sample size. An equal sampling weight of 95 is assigned to every body. The weights are also inflated as in Design 1 to account for weighting adjustment for nonresponses.

Design 3: Stratified Random Sample, Disproportionate Allocation

Two sets of stratified sample are generated using the following parameters:

	$p = 0.08$	$p = 0.60$			
Stratum	p_h	p_h	N_h	n_h	w_h
1	10%	70%	4,000	20	200
2	7.5%	60%	3,500	30	117
3	5%	40%	2,000	50	40

The notation w_h denotes sampling weight for stratum *h*. The weights are also inflated as in Design 1 to account for weighting adjustment for nonresponses.

Design 4: Stratified Random Sample, Disproportionate Allocation, Unequal Weights

Two sets of stratified sample are generated using the same parameters (disproportional sample) as used in the Design 3 above. However, to simulate weighting adjustment through the use of individual inverse propensity scores, the individual weight is generated using a lognormal distribution with mean 4.31 and variance 0.49, and then post-stratified by stratum population size.

Table 2.1 and 2.2 present results of the simulation.

Conclusion and Discussion

Tables 2.1 shows that the normal-based confidence interval can be misleading for small proportion. It is obvious that the coverage probability is lower than the nominal level. The coverage probability is worst when variability of the weights is large, as shown in

[3] The point and variance estimates were computed using SUDAAN (RTI 2004).

simulation with Design 4. All other alternative methods provide better coverage probability.

Table 2.1 also shows that when the variability of the weights is small, all alternative methods are conservative where the coverage probability is a little bit larger than the nominal level. On the other hand, when the variability of the weights is large, some methods such as the Binomial approach, Ad-hoc Wilson, and Andersson-Nerman confidence intervals have coverage probability lower than the nominal level.

In terms of computation, some methods are simpler than the others. The methods require computation of weighted proportion and/or its variance (or standard error) based on proper estimation technique for a complex survey data. They also require quantile value of certain distributions. Having these values, some methods then only require simple calculation that can be done using a simple calculator, while other methods require aggregating values of weights, weight squares, etc, which require more computation. Note that some methods will not produce confidence bounds when the value of estimate proportion is exactly 0 or 1.

Table 2.1. Simulated coverage probability and average length of the 95% confidence intervals by sample design and method: for $p = 0.08$

	Design 1		Design 2		Design 3		Design 4	
Confidence interval method	Coverage	Average length	Coverage	Average length	Coverage	Average length	Coverage	Average length
Normal approximation	92.0%	10.6%	92.8%	10.7%	90.1%	12.9%	87.1%	13.9%
Poisson (Breeze)	97.6%	12.3%	97.8%	12.3%	95.6%	15.6%	95.0%	17.1%
Logit transformation	96.4%	11.3%	96.5%	11.3%	94.9%	14.3%	94.5%	15.6%
Binomial approach	97.4%	11.5%	97.5%	11.6%	94.4%	14.5%	93.2%	15.7%
Ad-hoc Quadratic (Wilson)	96.2%	11.2%	96.2%	11.2%	93.5%	14.3%	92.7%	15.6%
Andersson-Nerman	96.2%	11.2%	96.2%	11.3%	94.2%	14.7%	92.8%	16.5%
Model-based Wilson	96.2%	11.2%	96.2%	11.2%	95.2%	13.6%	96.0%	15.4%
t-adjusted Andersson-Nerman	96.2%	11.2%	96.2%	11.3%	95.1%	14.9%	96.7%	18.9%

Table 2.2. Simulated coverage probability and average length of the 95% confidence intervals by sample design and method: for $p = 0.6$

	Design 1		Design 2		Design 3		Design 4	
Confidence interval method	Coverage	Average length	Coverage	Average length	Coverage	Average length	Coverage	Average length
Normal approximation	95.1%	19.5%	94.9%	19.0%	94.5%	22.3%	93.5%	24.8%
Poisson (Breeze)	99.8%	31.6%	99.9%	31.5%	99.9%	36.5%	99.7%	40.7%
Logit transformation	95.7%	19.2%	95.8%	19.2%	95.5%	22.0%	94.5%	24.3%
Binomial approach	96.4%	20.0%	96.0%	19.4%	96.1%	23.0%	95.3%	25.5%
Ad-hoc Quadratic (Wilson)	96.2%	19.6%	95.8%	19.1%	96.0%	22.6%	95.3%	25.3%
Andersson-Nerman	96.1%	19.6%	96.3%	19.2%	97.4%	27.2%	96.2%	32.2%
Model-based Wilson	96.0%	19.5%	96.3%	19.5%	97.1%	23.5%	97.1%	26.3%
t-adjusted Andersson-Nerman	96.0%	19.6%	96.1%	19.2%	97.4%	27.2%	96.2%	32.2%

References

Agresti, Alan, and Coull, Brent A. (1998), "Approximate is better than "exact" for interval estimation of binomial proportions", The American Statistician, 52, 119-126.

Brown, Lawrence D., Cai, T. Tony, and Dasgupta, Anirban (2001), "Interval estimation for a binomial proportion (Pkg: p101-133)"', Statistical Science, 16 (2), 101-133.

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

Kott, P. S., Andersson, P. G., and Nerman, O. (2001), "Two-sided Coverage Intervals for Small Proportion Based on Survey Data," Paper presented at Federal Committee on Statistical Methodology, Washington, D.C.

Newcombe, Robert G. (1998), "`Two-sided confidence intervals for the single proportion: Comparison of seven methods", Statistics in Medicine, 17, 857-872.

Research Triangle Institute (2004), *SUDAAN Language Manual, Release 9.0.*, Research Triangle Park, NC: Research Triangle Institute.