# Confidence Intervals for Statistics for Categorical Variables from Complex Samples

Keith Rust and Valerie Hsu
Westat, 1650 Research Blvd., Rockville, MD 20850

## Abstract

The standard large-sample-based approach to deriving a two-sided confidence interval performs poorly for proportions that are close to zero or one. Alternative methods are suggested in the literature, but no one approach appears to be universally superior, and some of the approaches are difficult to implement in practice, especially when the data are collected via a complex sample design. Such problems are even more evident when one estimates certain summary measures for two-way tables, such as the Index of Disparity (Pearcy and Keppel, 2002), where the parameter estimated is defined to be non-negative, but estimates close to zero are common. In this paper we consider the practical implementation of some alternatives to deriving confidence intervals in these cases using statistical software for the analysis of complex survey data (WesVar).

**Keywords:** Asymptotically normal estimators, multi-stage sampling, log transformation, log(-log) transformation, index of disparity, WesVar.

## 1. Introduction

In this paper, we discuss the practical aspects of obtaining confidence intervals for proportions and other summary statistics derived from categorical variables, when the data are obtained from a complex sample survey. More so than is the case with statistics derived from continuous variables, the distributions of the estimators in question are often sufficiently far from normal, even for moderate and large sample sizes. This means that the application of the standard, large-sample confidence intervals derived from inverting appropriate *t*-statistics often leads to resulting intervals with poor coverage properties.

There are approaches for dealing with such problems. For proportions in particular, specialized approaches have been proposed. But more generally, the use of transformations can be effective in obtaining confidence intervals with good coverage properties. However, using these transformations on a large scale in practice can be tedious, and for this reason, practitioners often do not avail themselves of this tool.

The purpose of this paper is to illustrate some practical approaches to deriving sound confidence intervals for statistics derived from categorical variables in a complex survey setting. First, we discuss the literature on confidence intervals for proportions estimated from survey data, and then elaborate on the use of transformations for this purpose. We then introduce two specialized summary measures derived from categorical data, the Index of Disparity and the Mean Deviation, and discuss the use of transformations for obtaining confidence intervals for these. We discussed using data from the National Health and Nutrition Examination Survey (NHANES), for which we describe the relevant features. The WesVar software package was used to carry out the computations.

## 2. Confidence Intervals for Proportions from Survey Data

The technical and practical difficulties in obtaining valid inferences, and in particular confidence intervals with stated coverage, for estimates of small proportions are well known (Brown et al., 2001; Vollset, 1993). There are several alternative solutions available in the general case, including those proposed by Clopper and Pearson (1934), Wilson (1927), and Jeffreys (see Brown et al., 2001). In the case where the data are collected from a complex sample survey, the difficulties are exacerbated.

Several authors have addressed the issue of obtaining confidence intervals for small proportions from survey data. Korn and Graubard (1998) propose a variation on the method of Clopper and Pearson, and show via simulation that the method has good empirical properties. They compared their proposed procedure with the standard approach based on the large-sample normality assumption of the estimator, a method that applies the logit transformation to the proportion, and an approach similar to their own, developed by Breeze (1990). The standard large-sample approach was clearly the least effective method of the four.

Kott and Carr (1997) and Korn and Graubard (1999) proposed a modified Wilson procedure. As with the modified Clopper-Pearson procedure, the chief modification to the original procedure involves replacing the actual sample size by the effective sample size for the proportion in question. We are not aware of any

systematic evaluation of the modified Wilson procedure, or comparisons of its performance with the modified Clopper-Pearson method or any transformations. More recently, Gray et al. (2004) have proposed making such a modification to the Jeffreys method, and show a case study where it produces slightly shorter intervals than the modified Clopper-Pearson procedure.

The use of transformations as an approach to deriving better confidence intervals for proportions is attractive in application. They are straightforward to apply, and do not require any modifications that depend upon the nature and parameters of the sample design (although of course the transformations may be more effective for some designs than others). Software can be developed relatively straightforwardly for obtaining the confidence intervals (and tests of statistical significance) via this approach. As will be seen later in the paper, the WesVar software (Westat, 2002) is well suited for applying transformations to estimators in survey sampling applications. In addition to the logit transformation, the arcsine transformation is well known for use with estimates of proportions (Wolter, 2007).

Proportions are not the only statistics that can present difficulties for deriving valid confidence intervals in the survey sampling context, even for moderate to large samples. In general, estimators of parameters which have bounded possible values present difficulties. This is because, when the true value lies close to the bound, the sampling distribution of the estimator is far from normal even with large sample sizes, and in particular is generally highly skewed. This makes the application of the assumption of large-sample normality very ineffective as a means of deriving confidence intervals. Examples of such statistics include estimates of population variances that are close to zero, and estimates of correlation coefficients that are close to 1. Typically, for these estimators an approach using transformations will be considered (such as the Fisher-Z transformation for correlation coefficients).

## 3. The Index of Disparity (*ID*) and Mean Deviation (*MD*)

In a recent application, we calculated confidence intervals for a statistic known as the Index of Disparity (Pearcy and Keppel, 2002), for a variety of health measures and a range of demographic subgroups. Consider the case where there are $C$ disjoint subgroups in the population (e.g., age groups or race/ethnic groups), and for each we have an estimate of a characteristic, $R$, which is often, but not necessarily, the rate of some health condition in the population. Let $R_i$ denote the value of the characteristic for subgroup $i$, while the value for the whole population is denoted by $R$. Then the Index of Disparity, *ID*, is defined as

$$ID = \{C^{-1} \sum_{i=1}^{C} | R_i - R | \} / R .$$

This can also be multiplied by 100 and expressed as a percentage. Note that this Index has a lower bound of zero, but has no upper bound, as if there is a very small category for which the value of $R$ is very different from the overall value, the Index can become arbitrarily large.

It is not unlikely in many applications that the true value of *ID* is close to zero. If category membership is unrelated to the characteristic involved, this will be the case. Thus, two-sided confidence intervals constructed using the assumption of large sample normality of the estimator of *ID* will often have a lower endpoint that is negative, and at the same time have true coverage that is substantially different from the nominal rate.

It thus seems reasonable to consider the use of a transformation to address this problem. However, it is not obvious what transformation is likely to be successful in this case. The logit transformation is not a likely contender since the index is not bounded above by 1, and so the logit may be undefined. One can consider the approach of looking for a variance stabilizing transformation, in the hope that it will also lead to an estimator with a more symmetric distribution. However, obtaining such a transformation is not obvious. If the variance of an estimator $X$ is roughly proportional to $X(1 - X)$ (as in the case of a proportion), then an arcsin $(X^{1/2})$ transformation is called for. Again, that will not be appropriate in this case as it is undefined for values in excess of 1. If the variance is roughly proportional to $X^2$ then a log transformation is suggested. If the variance is a function of $(Xlog(X))^2$ then a log(-log) transformation is appropriate. However, since *ID* is a fairly complex function of its component $R$ values, it is difficult to see how its variance could be approximated as a function of its value. In this paper we consider the application of both the log and log(-log) transformations to *ID*.

We also consider the related statistic, Mean Deviation (*MD*), described by Keppel, Pearcy, and Wagener (2002), and defined as

$$MD = C^{-1} \sum_{i=1}^{C} | R_i - R | .$$

When the $R_i$ values are bounded above by 1 (for example, if they are proportions), then *MD* is also bounded above by 1. This leads to a potential difficulty if the log transformation is used. This is because, if *MD* is close to 1, its logarithm will be slightly below zero. A large-sample two-sided confidence interval for the logarithm may well have an upper bound that is greater than zero. When this is back-transformed, it gives rise to an upper bound for the confidence interval for *MD* that is greater than 1. We consider both the log and log(-log) transformations in this case, with the use of the latter being motivated by this problem with the former.

## 4. The 1999-2002 National Health and Nutrition Examination Survey (NHANES)

The analyses discussed in the remainer of this paper were conducted using data from 1999-2002 NHANES. This ongoing survey conducted by the National Center for Health Statistics (NCHS) uses a stratified, multistage probability sample of the noninstitutionalized civilian U.S. population. The design oversamples certain population groups including Blacks, Mexican Americans (but not other Hispanics), persons aged 60 and above, and low-income individuals. This is done so that more precise estimates can be obtained for these groups than would be otherwise possible.

The survey includes a health questionnaire given to all sampled participants. Interviews are conducted with study participants in their homes to ascertain sociodemographic characteristics and self-reported aspects of health status. The interview includes a question as to whether the respondent has ever been told by a physician that he or she has diabetes (excepting when pregnant). Participants were also given a physical examination, and had a sample of blood drawn. One of the blood values measured was the level of glycosylated hemoglobin (HbA1C), a long-term measure of the body's ability to metabolize glucose, and is the standard for assessing control of diabetes among those already diagnosed with the disease. The analyses discussed in this paper were of the 843 adults (aged 20 years and older) who reported that they had diabetes, and for whom a valid HbA1C measure was obtained, during the 1999-2002 survey period.

As well as estimating the proportion of the diabetic population in various groups, one can use these data to estimate the proportion of the diabetic population within these groups that are classified as having their diabetes not under control. For this purpose, we defined a diabetic as being not in control if their HbA1C level exceeds 7.0. This is a standard established by the American Diabetes Association (American Diabetes Association, 2007). We can then calculate the Index of Disparity (*ID*) and Mean (*MD*) for various deviation sociodemographic subgroups, for the proportion of the diabetic population that have diabetes under control.

## 5. Software for Deriving Confidence Intervals for Analyses of Complex Survey Data Involving Transformations

There are a number of software packages available that are well suited to the analysis of data from complex survey samples (SUDAAN (RTI International, 2005), STATA (StataCorp 2005), SAS (SAS Institute, 2007), SPSS (SPSS, 2006)). However, a general limitation of these packages is that they can only be used for estimators that conform to certain formats. In particular, they are generally not able to obtain confidence intervals for statistics that have had transformations such as the logit, arcsine, log, and log(-log) transformations applied to them. A notable exception is that SUDAAN calculates confidence intervals for proportions via the logit transformation.

The WesVar software (Westat, 2002), however, is well suited to this purpose, although the software itself is not able to back-transform the confidence interval endpoints, which must be performed by the user. WesVar is not able to compute the arcsine transformation. However, one can obtain an approximation to this transformation as

$$\arcsin(p^{0.5}) = p^{0.5}(1 + p/6)$$

by taking a third-order Taylor series approximation for the arcsine function, expanded around 0. This approximation is quite close for values of $p$ below 0.1. Thus, for example, for $p$ = 0.05, $\arcsin(p^{0.5})$ = 0.225513, while $p^{0.5}(1 + p/6)$ = 0.225470.

WesVar is a stand-alone software package that is invoked via a set of Windows© 'point and click' options from menus. Formulae for specialized statistics can either be typed directly into a window, or built using point and click procedures. The COMPUTE and FUNCTION commands are used to do this. Note that it is the formula for the statistic of interest (the logit of a proportion, or an *ID* for example) which is programmed in this way, not the formulae for the variance estimate or the confidence interval limits. Since WesVar uses the replicated variance estimation approaches of the jackknife, or Balanced Repeated Replication (BRR) to calculate sampling variances; there are no special formula needed for specific estimators. The same general formula applies to all

estimators for which the variance estimation method can be validly applied. Thus, for example, for a jackknife sampling variance estimator, for a stratified multistage sample design with two units selected per stratum (provided that the strata can be regarded as being sampled with replacement), one form variance estimator for an estimator $\tau^*$ takes the form

$$Var(\tau^*) = [\sum_{t=1}^{T}\{(\tau^*_{(t)} - \tau^*)^2 + (\tau^*_{(-t)} - \tau^*)^2\}]/2$$

where $\tau^*_{(t)}$ denotes the replicate estimate formed by dropping one of the two PSUs from stratum *t* and doubling the contribution of the second PSU from that stratum, and $\tau^*_{(-t)}$ denotes the replicate formed by dropping the complementary PSU from the stratum *t*. For more detail about replicated variance estimation procedures for complex survey samples, see Rust and Rao (1996).

For the analyses of the 1999-2002 NHANES data using WesVar, we used the jackknife procedure, with 57 replicates. The replicate weights used for this purpose were created in WesVar, utilizing the stratum and PSU information available in the NHANES data file.

## 6. Examples of Analyses of Confidence Intervals Constructed via the Use of Transformations

Using WesVar, we calculated the following estimates for the population of diagnosed diabetics:

1. The proportion of the population, *p*, with diabetes duration of 15 to 20 years, in each of the four racial-ethnic groups: NonHispanic White, NonHispanic Black, Mexican American, and Other; and in each of the age groups: 20-44, 45-54, 55-64, 65-74, 75+;
2. Logit (*p*), by race/ethnicity, and by age group;
3. $p^{1/2}(1+p/6)$, by race/ethnicity, and by age group;
4. The modified Wilson intervals for *p*, by race/ethnicity, and by age group;
5. The Index of Disparity (*ID*) for poor Diabetes Control (HbA1C>7), for subgroups of age, race/ethnicity, gender, poverty income ratio, and education;
6. The Mean Deviation (*MD*) for poor Diabetes Control (HbA1C>7), for subgroups of age, race/ethnicity, gender, poverty income ratio, and education;
7. Log (*ID*) and Log (*MD*) for the same subgroup classes as listed in 5) and 6); and
8. Log(-log(*ID*)) and Log(-log(*MD*)) for the same subgroup classes as listed in 5) and 6).

We used WesVar to derive two-sided 95 percent confidence intervals for each of these, and then back-transformed the endpoints (where appropriate), to derive confidence intervals for the proportions (*p*), Index of Disparity (*ID*), and Mean Deviation (*MD*). In each case we used the large sample approximation to calculate the interval. That is, the confidence limits were generated as

$$\hat{\theta} \pm t.\sqrt{\text{var}(\hat{\theta})}$$

where the *t* coefficient was taken as the 97.5$^{th}$ percentile of a central *t*-distribution with 29 degrees of freedom. This coefficient is selected based on the number of PSUs in the NHANES design, and hence the number of replicates in the jackknife procedure. For estimates of proportion (*p*) we also used the Wilson procedure, modified for use with complex survey samples. This procedure replaces the sample size parameter used in the standard Wilson confidence interval with the effective sample size, $n_{eff}$, defined as:

$$n_{eff} = \hat{p}(1-\hat{p})/\text{var}(\hat{p})$$

That is, the effective sample size is the actual sample size, divided by the design effect for $\hat{p}$, the estimate of *p*. The confidence limits are given as

$$\frac{\{(2n_{eff}\,\hat{p} + t^2) \pm (t\sqrt{(t^2 + 4\,\hat{p}(1-\hat{p})n_{eff}})\}}{2(n_{eff} + t^2)}$$

where again *t* is the 97.5$^{th}$ percentile point of the central *t* distribution with 29 degrees of freedom. Note that as $n_{eff}$ increases, these limits approach those of the standard Wald approach.

In the case of confidence intervals for proportions obtained via the approximate arcsine transformation described above, the confidence limits for each proportion were obtained by squaring the sines of the upper and lower confidence limits of the transformed parameter.

### 6.1 WesVar Programming

Figure 1 shows a screen shot for WesVar, which illustrates how the user labels the cells of a table, for subsequent use in defining functions of estimates from

within table cells. In this example, using the NHANES data, cells for the different categories of race/ethnicity are being defined.

Figure 2 shows an example where the user programs the function statements that implement the desired transformation above. The logit and approximate arcsine transformations for proportions are shown in this example. Figure 3 shows how the results of the programmed function are exhibited in the WesVar output.

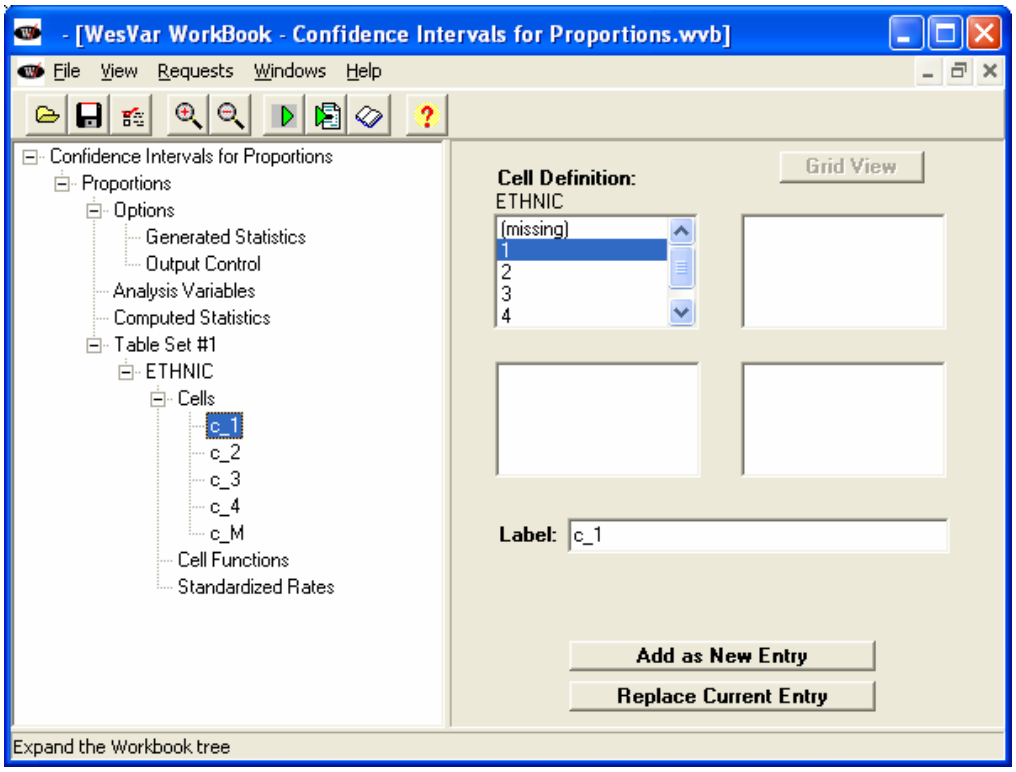


Figure 1. Cell definition screen

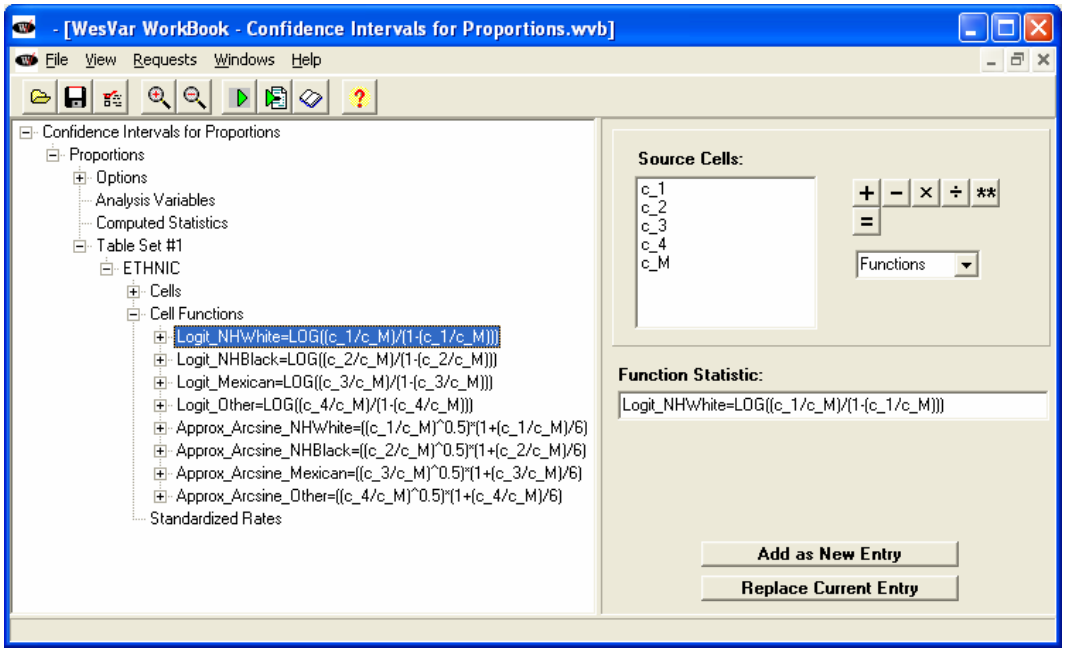


Figure 2. Function statistic screen

| | |
|---|---|
| TABLE : | ETHNIC |
| Cell Definition : | c_1 : ETHNIC = Nh white |
| | c_2 : ETHNIC = Nh black |
| | c_3 : ETHNIC = Mex Am |
| | c_4 : ETHNIC = Other |
| | c_M : ETHNIC = MARGINAL |
| Function Statistics : | Logit_NHWhite=LOG((c_1/c_M)/(1-(c_1/c_M))) |
| | FOR-- SUM_WTS |
| | Logit_NHBlack=LOG((c_2/c_M)/(1-(c_2/c_M))) |
| | FOR-- SUM_WTS |
| | Logit_Mexican=LOG((c_3/c_M)/(1-(c_3/c_M))) |
| | FOR-- SUM_WTS |
| | Logit_Other=LOG((c_4/c_M)/(1-(c_4/c_M))) |
| | FOR-- SUM_WTS |
| | Approx_Arcsine_NHWhite=((c_1/c_M)^0.5)*(1+(c_1/c_M)/6) |
| | FOR-- SUM_WTS |
| | Approx_Arcsine_NHBlack=((c_2/c_M)^0.5)*(1+(c_2/c_M)/6) |
| | FOR-- SUM_WTS |
| | Approx_Arcsine_Mexican=((c_3/c_M)^0.5)*(1+(c_3/c_M)/6) |
| | FOR-- SUM_WTS |
| | Approx_Arcsine_Other=((c_4/c_M)^0.5)*(1+(c_4/c_M)/6) |
| | FOR-- SUM_WTS |

Figure 3. Function statistics in WesVar output

### 6.2 Results

Tables 1 and 2 show the results of the computations listed above. In the cases where WesVar was used to transform the statistic of interest prior to calculating two-sided confidence intervals, the confidence interval endpoints have been back-transformed using EXCEL©.

Table 1 shows the results for proportions. When the subgroup sample sizes are reasonably large, the methods give similar results. When one considers the results for Mexican Americans, however, where the sample size is 69, one can see noticeable differences among the methods.

First, one can see that the standard large-sample, or Wald, confidence interval has a negative lower bound (-2.78 percent), for one age group (20-44 years). The other methods all avoid this drawback. In doing so, they produce substantially higher upper limits than the Wald approach. Even for the other age groups, where the Wald method has a positive lower bound, each of the transformations shifts the interval upwards. The hope is that these transformed intervals give closer to 95 percent coverage. Presumably, they also have tail percentages that are more nearly equal to 2.5 percent than does the Wald-based method.

The arcsine intervals are generally closer to the standard Wald interval than those derived via the logit transformation or the Wilson approach. These latter two methods give very similar results, except in the case of the 20-44 years age group. The Wilson intervals are consistently enclosed within the endpoints resulting from the logit procedure, and for the 20-44 year-old group, the Wilson interval is substantially shorter.

For the Index of Disparity and the Mean Deviation, the differences among the three approaches used vary with subgroup (see Table 2), reflecting the fact that the true values for these indexes vary considerably by subgroup. In the case of gender, the disparity indexes are relatively close to zero. This leads to the result that the Wald intervals for both *ID* and *MD* have lower bounds that are below zero. Both of the transformations avoid this problem, as they were designed to do. The log(-log) transformation gives higher confidence limits than the Wald approach, with the log transformation giving higher limits still. With its shorter (but always completely positive) confidence intervals, the log(-log) transformation is more appealing than the log transformation, but of course this analysis says nothing about the true coverage properties of either approach.

Table 1. 95% confidence intervals for the percentage in each racial-ethnic group and age group for the population with diabetes duration of 15 to 20 years *(n = 69)*

| | | | Confidence intervals | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Wald | | Modified Wilson | | Logit | | Arcsine approx. | |
| | Estimate | DEFF | L | U | L | U | L | U | L | U |
| *Race/ethnicity* | | | | | | | | | | |
| NonHispanic White | 61.20 | 1.74 | 45.39 | 77.01 | 45.06 | 75.21 | 44.75 | 75.44 | 44.48 | 70.27 |
| NonHispanic Black | 20.13 | 1.59 | 7.70 | 32.57 | 10.59 | 34.92 | 10.38 | 35.42 | 9.37 | 33.43 |
| Mexican American | 3.43 | 0.35 | 0.77 | 6.10 | 1.60 | 7.22 | 1.55 | 7.44 | 1.27 | 6.60 |
| Other | 15.24 | 2.77 | 0.50 | 29.97 | 5.72 | 34.76 | 5.26 | 36.77 | 3.76 | 32.38 |
| *Age* | | | | | | | | | | |
| 20-44 years | 4.77 | 2.07 | -2.78 | 12.33 | 1.09 | 18.55 | 0.63 | 28.31 | 0.09 | 15.89 |
| 45-54 years | 16.06 | 1.81 | 3.89 | 28.23 | 7.39 | 31.45 | 6.95 | 32.88 | 5.87 | 29.86 |
| 55-64 years | 21.23 | 2.37 | 5.72 | 36.74 | 9.90 | 39.80 | 9.52 | 40.85 | 8.17 | 38.01 |
| 65-74 years | 33.85 | 1.59 | 19.15 | 48.56 | 21.16 | 49.39 | 20.95 | 49.71 | 20.25 | 47.72 |
| >=75 years | 24.08 | 0.84 | 14.44 | 33.72 | 15.85 | 34.82 | 15.72 | 35.04 | 15.16 | 33.84 |

Table 2. 95% confidence intervals for the Index of Disparity and the Mean Deviation for Poor Diabetes Control (HbA1C > 7), for subgroups of age, race/ethnicity, gender, poverty income ratio, and education *(n = 843)*

| | | Confidence intervals | | | | | |
|---|---|---|---|---|---|---|---|
| | | Wald | | Log | | Log(-log) | |
| Statistic | Estimate | L | U | L | U | L | U |
| *Age* | | | | | | | |
| ID | 10.93 | 3.58 | 18.27 | 5.65 | 21.14 | 4.72 | 18.76 |
| MD | 0.06 | 0.02 | 0.10 | 0.03 | 0.11 | 0.03 | 0.11 |
| *Race/ethnicity* | | | | | | | |
| ID | 12.55 | 3.86 | 21.25 | 6.24 | 25.26 | 4.72 | 21.67 |
| MD | 0.07 | 0.02 | 0.11 | 0.03 | 0.13 | 0.03 | 0.12 |
| *Gender* | | | | | | | |
| ID | 4.24 | -2.63 | 11.12 | 0.83 | 21.76 | 0.43 | 13.94 |
| MD | 0.02 | -0.01 | 0.06 | 0.00 | 0.11 | 0.00 | 0.08 |
| *Poverty income ratio* | | | | | | | |
| ID | 4.68 | -1.25 | 10.60 | 1.30 | 16.84 | 0.87 | 12.52 |
| MD | 0.02 | -0.01 | 0.06 | 0.01 | 0.09 | 0.01 | 0.07 |
| *Education* | | | | | | | |
| ID | 4.62 | -3.75 | 12.99 | 0.74 | 29.02 | 0.30 | 16.47 |
| MD | 0.02 | -0.02 | 0.07 | 0.00 | 0.16 | 0.00 | 0.10 |

## 7. Conclusions and Future Research

This paper discusses the use of transformations to stabilize the calculation of confidence intervals. The application in the case of complex survey data has been discussed, and we have demonstrated how these "nonstandard" statistics can be successfully computed, along with their confidence intervals, with relatively little effort using WesVar.

However, it is far from clear which transformations are likely to be the most successful in any application, and, in the case of *ID* and *MD*, it is not even clear whether there is a transformation that can be effective in creating two-sided confidence intervals with adequate coverage properties.

We hope to investigate some of these issues via simulation studies. This will involve some cases where no

complex sampling is involved, since, if no suitable transformation can be found in the simple random sampling case, it seems unlikely that one could be found that will apply to a range of complex survey situations. Simulations in the case of complex survey designs are also warranted, building on the work of Korn and Graubard, Kott, and others in this area.

## References


American Diabetes Association. (2007). *Diabetes Care*, 30, S4-S41.

Breeze, E. (1990). *General Household Survey: Report on sampling error*. London: Her Majesty's Stationary Office (Office of Population Censuses and Surveys).

Brown, L.D., Cai, T.T., and DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statistical Scienc*e, 16, 101-133.

Clopper, C.J., and Pearson, E.S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26, 404-413.

Gray, A., Haslett, S., and Kuzmicich, G. (2004). Confidence intervals for proportions estimated from complex sample designs. *Journal of Official Statistics*, 20, 705-723.

Keppel, K.G., Pearcy, J.N., and Wagener, D.K. (2002). Trends in racial and ethnic-specific rates for the health status indicators: United States, 1990-1998. *Healthy People 2000 Stat Notes*. Jan 2002 (23), 1-16.

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.

Korn, E.L., and Graubard, B. I. (1999). *Analysis of Health Surveys*. New York: Wiley.

Kott, P.S., and Carr, D.A. (1997). Developing an estimation strategy for a pesticide data program. *Journal of Official Statistics*, 13, 367-383.

Pearcy, J.N., and Keppel, K.G. (2002). A summary measure of health disparities. *Public Health Reports*, 117, 273 – 280.

RTI International. (2005). *SUDAAN release 9.0.1*. Research Triangle Park, NC: RTI International.

Rust, K.F. and Rao, J.N.K. (1996). Variance estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research*, 5, 283-310.

SAS Institute, Inc. (2007). *SAS release 9.1.3*. Cary, NC: SAS Institute, Inc.

SPSS. (2006). *SPSS for Windows, Rel. 15.0*. Chicago, IL: SPSS, Inc.

StataCorp. (2005). *Stata statistical software: Release 9*. College Station, TX: StataCorp LP.

Vollset, S.E. (1993). Confidence intervals for a binomial proportion. *Statistics in Medicine*, 12, 809-824.

Westat. (2002). *WesVar 4.2*. Rockville, MD: Westat.

Wilson, E.B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 22, 209-212.

Wolter, K.M. (2007). *Introduction to variance estimation: Second edition*. New York: Springer-Verlag.