Trend Analysis of Aggregate Outcomes in Complex Health Survey Data

Wesley M. Durand, Paul C. Stey, PhD, Elizabeth S. Chen, PhD, Indra Neil Sarkar, PhD, MLIS
Alpert Medical School and Center for Biomedical Informatics, Brown University, Providence, RI

Abstract

Public health and clinical practice pattern trends are often analyzed using complex survey data. Use of statistical approaches that do not account for survey design predisposes to error, potentially leading to resource misdirection and inefficiency. This study examined two techniques for analyzing trends in complex survey data: (1) design-corrected logistic regression and (2) jackknife re-weighted linear regression. These approaches were compared to weighted least squares regression, as well as non-design corrected techniques. Data were obtained from NEISS, a complex survey of emergency departments that can be weighted to produce national estimates of injury occurrence. Trends were analyzed in rug-related injuries among male versus female patients ≥65 years of age. All design-corrected techniques performed comparably in assessment of trend within sex-based subgroups. In almost all cases, design-corrected approaches contrasted profoundly with standard statistical techniques. Future analyses may employ these design-corrected approaches to appropriately account for estimate variance in complex survey data.

Introduction

Analysis of complex survey data often requires specialized statistical software or packages. Such procedures utilize survey design parameters to calculate variance associated with estimates from these data (i.e., design-corrected analysis). Use of standard statistical procedures assuming simple random sampling to analyze survey data predisposes to error. Studies of public health trends and clinical practice patterns often utilize complex survey data. Accordingly, analyses utilizing non-design-corrected statistical techniques risk misguided conclusions, potentially leading to resource misdirection and inefficiency.

The National Health and Nutrition Examination Survey (NHANES; CDC), National Inpatient Sample (NIS; HCUP, AHRQ), and National Electronic Injury Surveillance System (NEISS; CPSC, CDC) are complex surveys commonly employed in biomedical and public health research. The longitudinal nature of these datasets is particularly useful for evaluating trends in event occurrence over time. In these analyses, the outcome is often an aggregate within levels of the variable over which trend analysis is desired (e.g., number of liver transplants per year, or aggregate cost of coronary artery bypass graft procedures per year). The requirement to utilize survey design-corrected procedures, however, limits the array of tools available to researchers. This issue is compounded by significant inter-software variability in available procedures and functionality. Accordingly, when examining aggregate outcome trends, many studies in the biomedical literature have either utilized non-design-corrected statistical approaches (e.g., standard linear regression or Cochran Armitage trend tests), refrained from statistical trend analysis, or circumvented direct trend analysis through a variety of methods.

Weighted least-squares regression is a technique for regression analysis of survey data, as published by Gillum, Graves, and Jean in 1996. While widened availability of survey analysis software has reduced use of this technique for continuous, categorical, and count outcomes, select researchers have applied it to trend assessment of aggregate outcomes in recent years. This approach, however, has several disadvantages. It assumes that standard errors are derived from a sufficiently large sample such that they are without sampling error; this may not always be the case. The magnitude of the regression estimate under weighted least squares may differ from that observed without weighting. Finally, the approach may require ad hoc calculation, complicating “out of the box” usage, particularly when analyzing variable interactions.

This study aimed to develop and evaluate approaches for trend assessment of aggregate outcomes in complex survey data that can be easily applied across a broad array of statistical software. Here, the use of two methods is described: (1) design-corrected logistic regression and (2) jackknife re-weighted linear regression. These methods are compared to an existing design-corrected method (weighted least squares regression) as well as non-design-corrected methods.
(standard linear regression and the Cochran Armitage trend test). The impact of using these techniques is illustrated with data from NEISS, analyzing trends in rug-related injuries among patients ≥65 years of age.

Methods

Data Source: Full-year NEISS data (1999-2015) were abstracted from the Consumer Product Safety Commission query builder ([https://www.cpsc.gov/cgibin/NEISSQuery/home.aspx](https://www.cpsc.gov/cgibin/NEISSQuery/home.aspx)). The NEISS is a stratified sample of ~100 hospital emergency departments (EDs) that can be weighted to produce national estimates of product-related injuries. EDs are divided into strata based on hospital size. Year was treated as an additional stratification variable, such that size-based strata were temporarily separated; this replicates the NEISS’s annual sampling approach.

Patient Selection: Records were included based on presence of: a product code for “rugs” (codes 0612, 0676, 0613); patient age ≥65; and diagnosis code for fracture (code = 57).

Variables: The primary dependent variable in this analysis was rug-related fracture occurrence among elderly patients. Independent variables were sex and year.

Statistical Analysis:

Design-Corrected: Logistic Regression: The design-corrected logistic regression approach utilized a scheme for dataset preparation to treat an aggregated outcome as binary. The overall analysis approach is detailed in Figure 1.

![Figure 1. Dataset Creation Strategy for Design-Corrected Logistic Regression.](image)

First, a dataset containing dummy records for each primary sampling unit (PSU) by year was created and merged with isolated patient records, ensuring that all PSUs were represented in the analysis. Patient cases were assigned “cohort=1” and dummy PSU records were assigned “cohort=0” for domain analysis. Note that this step would be required in any analysis seeking to correctly calculate estimate variance, as not all PSUs may have contributed patient records to the analysis.

Subsequently, a denominator dataset was created containing observations for each combination of variables over which trend analysis would occur; in our case, all year*sex variable combinations. These observations were assigned an arbitrary but consistent stratum (stratum = P) and PSU (psu = 999) values. By placing these observations in “lonely PSUs”, estimates from these observations do not add to the overall variance of the sample\(^1\). Observations were weighted as the difference between an arbitrary but consistent large number (in our case, K = 1,000,000) and the actual observed number of injuries for that year*sex combination. Importantly, patient cases were assigned “event=1” and denominator dataset observations were assigned “event=0”. Denominator dataset observations were also assigned “cohort=1”.

Design-corrected logistic regression was accomplished with the `survey` R package, utilizing “event” as the binary dependent variable and sex (reference = male), year, and an interaction term as independent variables, as illustrated in Equation \(^1\). A domain analysis was performed for observations with “cohort=1”. A custom linear hypothesis test was requested to determine the effect of year among females.
Equation 1: \( \text{logit}(\text{sex, year}) = \beta_{\text{intercept}} + \beta_1 \ast \text{sex} + \beta_2 \ast \text{year} + \beta_3 \ast \text{sex} \ast \text{year} \)

where annual change among males is represented as \( \beta_{\text{year, males}} = \beta_2 \), the difference in annual change among females versus males as \( \beta_{\text{year, \Delta}} = \beta_3 \), and the annual change among females as the linear hypothesis test of \( \beta_{\text{year, females}} = \beta_2 + \beta_3 \)

Accordingly, the modeled absolute number of events can be expressed as a function of the probability of event occurrence, calculated according to Equation 2:

Equation 2: \( N(\text{sex, year}) = p(\text{sex, year}) \ast K = \frac{\text{logit}(\text{sex, year})}{1 + \text{logit}(\text{sex, year})} \ast K \)

where \( \text{logit}(\text{sex, year}) \) is calculated according to Equation 1, and \( K \) is the large number used in weighting denominator dataset observations.

**Design-Corrected: Jackknife Re-weighted Linear Regression:** In contrast to the design-corrected logistic regression approach, the jackknife re-weighted linear regression method involved primarily computational steps. The unique hospital dataset and rug-related injury datasets were combined for jackknife re-weighted analysis (the denominator weight dataset was excluded). The jackknife approach to complex survey analysis is available as an alternative to Taylor Series linearization in many statistical packages, though, to our knowledge, there is no literature illustrating use in analyzing aggregate outcomes. In this approach, the analysis is repeated for \( R \) replicates, where \( R \) is equal to the number of PSUs in the sample. For each replicate \( r \), a sequential PSU is selected for deletion from the sample, and observations from the remaining PSUs within the donor stratum (i.e., the stratum of the deleted PSU) are re-weighted according to Equation 3a. A standard linear regression for number of events by year, sex, and their interaction was conducted for each replicate, defined in Equation 4. Estimate magnitude was calculated as the arithmetic mean of replicate estimates, and variance was computed according to Equation 3b. All calculations for the jackknife re-weighted linear regression approach were conducted in Julia 0.6.0, using the RCall.jl package for regression and linear hypothesis testing.

Equation 3a: \( w_{ij}^{(r)} = \frac{w_{ij}}{\alpha_r} \ast \alpha_r = \frac{n_{hr}^{-1}}{n_{hr}} \)

where \( w_{ij} \) is the weight for the \( j \)th member of the \( i \)th PSU, and \( n_{hr} \) is the number of PSUs in the donor stratum \( h_r \) for replicate \( r = 1, 2, 3 \ldots R \),

Equation 3b: \( V(\beta) = \sum_{r=1}^{R} \alpha_r (\beta_r - \beta)^2 \)

where \( \beta \) are the estimated regression coefficients from the full sample for \( \beta \), and \( \beta_r \) is the regression coefficient from the \( r \)th replicate

Equation 4: \( N(\text{sex, year}) = \beta_{\text{intercept}} + \beta_1 \ast \text{sex} + \beta_2 \ast \text{year} + \beta_3 \ast \text{sex} \ast \text{year} \)

where annual change among males is represented as \( \beta_{\text{year, males}} = \beta_2 \), the difference in annual change among females versus males as \( \beta_{\text{year, \Delta}} = \beta_3 \), and the annual change among females as the linear hypothesis test of \( \beta_{\text{year, females}} = \beta_2 + \beta_3 \)

**Design-Corrected: Weighted Least Squares Regression:** Estimates and standard errors were computed using the survey package for all year*sex combinations. As described by Gillum, Graves, and Jean, regression coefficients and standard errors were calculated according to Equations 5a and 5b. Weighted least squares regression was implemented in Julia 0.6.0.
Equation 5a: $\beta = \frac{\sum w_i x_i y_i - (\sum w_i \bar{x})(\sum w_i \bar{y})}{\sum w_i x_i^2 - (\sum w_i x)^2}$

Equation 5b: $S_\beta = \sqrt{\frac{1}{\sum w_i x_i^2 - (\sum w_i x)^2}}$

where, $w_i = 1/S^2_i$, $\bar{x} = \sum w_i x_i / \sum w_i$, $\bar{y} = \sum w_i y_i / \sum w_i$

For inter-group analysis of trend, the difference between the two groups (i.e., number of injuries among females minus number of injuries among males) was modeled as a function of year. Regression estimate and standard error for this value were calculated as in Equation 5a and 5b, with pooled standard deviation determined according to Equation 6:

Equation 6: $S_{pooled} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$

where, $s_i$ are the standard deviations to be pooled, and $n_i$ are the sample sizes

Standard Analysis: Linear Regression: For comparison, standard linear regression was performed (regression equation identical to Equation 4) for number of events as the dependent variable, with sex, year, and their interaction term as independent variables. Linear regression was implemented in Julia 0.6.0, with use of the RCall.jl package for linear hypothesis testing.

Standard Analysis: Cochran Armitage Trend Test: The Cochran Armitage trend test was implemented, comparing linear trend among male versus female injuries without correction for survey design. This approach was implemented in Julia 0.6.0 using the RCall.jl package and the DescTools R package.

Software: Dataset cleaning and preparation were accomplished using SAS 9.4 (SAS Institute, Cary, NC) and Julia 0.6.0. Analysis was accomplished as detailed above. Plots of predicted versus observed injury counts were produced in Julia with the PlotlyJS.jl package.

Results

Descriptive Statistics: In total, 8,005 unweighted records were analyzed, corresponding to a national estimate of 349,605 patients (SE 10,676) (all future statistics are derived from weighted estimates). The cohort was 82.2% female (SE 0.50%) and 17.8% male (SE 0.50%), corresponding to 287,244 (SE 8,923) and 62,361 (SE 2,603) discharges, respectively.

Design-Corrected Logistic Regression: The design-corrected logistic regression approach indicated a robust increasing trend in the annual number of rug-related injuries among both males ($\beta_{year, males} = 5.120 \times 10^{-2}$, SE $9.153 \times 10^{-3}$) and females ($\beta_{year, females} = 2.035 \times 10^{-2}$, SE $6.901 \times 10^{-3}$). The difference between the two was similarly marked ($\beta_{year, \Delta} = -3.085 \times 10^{-2}$, SE $7.150 \times 10^{-3}$). A plot of modeled annual injury count was slightly non-linear. (Table 1) (Figure 2) (Figure 3)

Jackknife Re-weighted Linear Regression: The jackknife re-weighted linear regression approach also exhibited robust positive trends among both males ($\beta_{year, males} = 184.8$, SE $33.06$) and females ($\beta_{year, females} = 337.2$, SE $115.8$). Notably, the t-values for these coefficients were very similar to those observed with design-corrected logistic regression. The difference between the two trends, however, was less significant ($\beta_{year, \Delta} = 152.5$, SE $97.09$), and was not similar to that produced above. (Table 1) (Figure 2) (Figure 3)

Weighted Least Squares Regression: This approach produced standard errors that were relatively similar to those produced in jackknife re-weighted linear regression, though with coefficients differing to a greater degree ($\beta_{year, males} = 153.1$, SE $30.37$) ($\beta_{year, females} = 219.3$, SE $101.2$). The difference observed between the two trends exhibited a comparable standard error to that produced using jackknife re-weighted linear regression ($\beta_{year, \Delta} = 63.08$, SE $92.80$). (Table 1) (Figure 2)
**Standard Linear Regression:** Use of standard linear regression produced coefficients with identical magnitude but differing standard errors as compared to those from jackknife re-weighted linear regression ($\beta_{\text{year,males}} = 184.8$, SE 68.05) ($\beta_{\text{year,females}} = 337.2$, SE 68.05). The results for the difference between the trends was near identical ($\beta_{\text{year,}\Delta} = 152.5$, SE 96.24). (Table 1) (Figure 2)

**Cochran Armitage Trend Test:** The Cochran Armitage test indicated a highly robust difference in trend between the two sexes ($Z = 32.92$). (Table 1)

**Table 1.** Comparison of Select Coefficients for Trend Analysis.

| Design-Corrected Analysis | Key | $\beta_{\text{year,males}}$ | $\beta_{\text{year,females}}$ | $\beta_{\text{year,}\Delta}$ |
|---------------------------|-----|-----------------------------|-----------------------------|-----------------------------|
| Logistic Regression       | Estimate | 5.12E-02 | 2.04E-02 | -3.09E-02 |
|                           | SE     | 9.15E-03 | 6.90E-03 | 7.15E-03 |
|                           | $t$ Value | 5.594 | 2.948 | -4.315 |
| Jackknife Re-weighted Linear Regression | Estimate | 184.8 | 337.2 | 152.5 |
|                           | SE     | 33.06 | 115.8 | 97.09 |
|                           | $t$ Value | 5.59 | 2.914 | 1.571 |
| Weighted Least Squares Regression | Estimate | 153.1 | 219.3 | 63.07 |
|                           | SE     | 30.37 | 101.2 | 92.8 |
|                           | $t$ Value | 5.041 | 2.168 | 0.68 |
| Linear Regression         | Estimate | 184.8 | 337.3 | 152.5 |
|                           | SE     | 68.05 | 68.05 | 96.24 |
|                           | $t$ Value | 2.716 | 4.957 | 1.585 |
| Standard Analysis         | Cochran Armitage Trend Test | Z Value | - | - | 32.92 |
Figure 2. Comparison of \( t \) Values from Regression Trend Analyses. DC: Design-Corrected.

Figure 3. Observed and Predicted Values from Design-Corrected Logistic and Linear Regression. DCLR, Design-Corrected Logistic Regression.
Discussion

Drivers of U.S. healthcare costs will experience increased scrutiny as health-related expenditures continue to grow. Trend analysis of public health and clinical practice patterns will play a significant role in characterizing these dynamics, and these analyses will often be conducted using complex survey data. Use of statistical techniques that account for estimate variance is critical, particularly when resource-allocation decisions may be made based on the presence or absence of a significant trend.

This investigation illustrated two methods for assessing trends in aggregate outcomes from complex survey data: design-corrected logistic and jackknife re-weighted linear regression. While this study utilized the NEISS dataset, these results are readily generalizable to analysis of other complex survey data (e.g., NIS and NHANES). These approaches were compared to weighted least squares regression—a previously-employed technique for complex survey trend analysis—and standard, non-design-corrected analytic approaches. Both design-corrected and non-design-corrected techniques incorporated survey weights and thus produced identical point estimates for annual event occurrence. Non-design-corrected techniques, however, did not account for survey design parameters and therefore produced skewed values for the variability of these estimates. Ultimately, there was generally greater similarity between design-corrected approaches than when compared to techniques that did not account for survey design.

Comparison of Trends Among Subgroups:

Comparison between logistic and linear regression techniques is more apt with t-values than standard errors. Considering this, all design-corrected techniques produced comparable results for trends among both male and female subgroups, and the logistic and jackknife re-weighted linear techniques yielded profoundly similar values. The design-corrected approaches were highly distinct computationally; such mutual corroboration increases confidence in their validity. All design-corrected techniques exhibited markedly greater t-values for subgroup trends than did standard linear regression. Accordingly, public health and biomedical research that utilizes standard linear regression to analyze trends from complex survey data may be predisposed to misguided conclusions; researchers should instead utilize only design-corrected approaches.

Among techniques employing linear regression, jackknife re-weighted and standard linear regression yielded near-identical point estimates for regression parameters. Weighted least squares regression exhibited generally lower values. This likely occurred due to weighting, with greater importance given to observations with lower standard errors. Higher y-values may therefore be disproportionately penalized, yielding attenuated trend magnitude. This finding has implications for researchers interested in preserving equal-weighting in trend analysis, particularly when their data exhibits a wide range of outcome values.

Comparison of Inter-Group Trends Between Subgroups:

The difference in trend among males versus females varied greatly between design-corrected approaches. For the design-corrected logistic regression approach, this is attributable to the modeling of relative versus absolute annual change (i.e., log-odds of event occurrence versus number of events). As expected, the trend modeled with logistic regression is non-linear. Accordingly, while linear regression coefficients indicated that the annual absolute change in event number among females is greater than among males, logistic regression coefficients showed that the annual relative change was greater among males than among females. Researchers conducting trend analysis should therefore consider whether their hypothesis is best assessed in relative or absolute terms and choose their technique accordingly.

While standard errors were very similar between jackknife re-weighted and weighted least squares regression, regression parameter magnitude was markedly different. As above, this is likely attributable to disproportionate penalizing of larger observations under the weighted least squares approach. Indeed, the difference between jackknife re-weighted and weighted least squares regression coefficients for \( \beta_{\text{year,male}} \) and \( \beta_{\text{year,female}} \) is approximately equal to the difference in differences for the two approaches between \( \beta_{\text{year,male}} \) and \( \beta_{\text{year,female}} \).

Notably, jackknife re-weighted linear regression produced comparable results for the difference in trend between males and females as compared to those from standard linear regression. It is possible that this observation is related to features of our dataset, rather than of the method employed; we will test this hypothesis in planned simulation studies. In contrast, all regression-based approaches (including standard linear regression) yielded profoundly different outcomes as compared to a non-design-corrected Cochran Armitage trend test, which showed a highly significant difference between trends among males versus females. This finding has significant implications for future trend analyses of complex survey data. Use of non-design-corrected techniques may not only predispose to error, but the direction of such error (i.e., overestimation versus underestimation) may be dependent on the technique employed.
Holistic Method Comparison:

Performance aside, it is important to consider ease of implementation for design-corrected techniques. We observed design-corrected logistic regression to be the most straightforward of those tested, followed by weighted least squares and jackknife re-weighted linear regression. Design-corrected logistic regression required only a single incremental data cleaning step, leveraging existing survey data analysis techniques for further computation. Implementation was straightforward in R, SAS, and Julia, and required very little de novo programming. In contrast, the other approaches required a fair degree of customization. Jackknife re-weighted linear regression required not only replicate creation, but also custom variance, standard error, and t-value calculations. Additionally, the considerable number of replicates required for jackknife re-weighted linear regression (1717 replicates in total) made this approach more computationally intensive. Weighted least squares regression required custom regression parameter and variance calculations, as well as pooling of standard errors, techniques with which many researchers may not be familiar.

Nevertheless, design-corrected logistic regression had a notable disadvantage: it modeled the linear annual change in log-odds and therefore assessed non-linear trends in annual event occurrence. As we observed, there may be significant differences between trends modeled in absolute versus relative annual change. While multivariate models are commonplace in epidemiology, researchers desirous of modeling strictly linear absolute annual change may opt for jackknife re-weighted regression. The choice of technique may ultimately be influenced by a priori assumptions regarding relationships between dependent and independent variables, as well as patterns observed in exploratory analyses.

Future Directions

Further characterization of these techniques will involve simulated survey datasets to assess reproducibility of trend assessment with replicate sampling. Reproducibility will be measured as the distribution of trend t-values as assessed by each technique across replicate samples. If one believes that a method accurately accounts for sampling approach (i.e., is “design-corrected”), then one should believe that the method will produce relatively more consistent results with replicate sampling than would a method that is not design-corrected. We hypothesize that design-corrected techniques will offer superior reproducibility over non-design-corrected techniques.

Conclusion

This investigation showed that, with proper dataset cleaning, design-corrected logistic regression is an accurate and easily implemented method to account for variance in trend analysis of aggregate outcomes from complex survey data. In the context of rug-related fractures among the elderly from the NEISS dataset, the results produced with design-corrected logistic regression were highly similar to those from jackknife re-weighted linear regression; the latter may also be employed when researchers are desirous of modeling strictly linear change in absolute quantities. Generally, all design-corrected techniques produced markedly different results as compared to those from standard linear regression and Cochran Armitage trend test. Thus, the results from non-design-corrected analysis may not only lead to misguided conclusions regarding trend robustness, but also manifest in either overestimation or underestimation, depending on the non-design-corrected technique employed. Future analyses of trends in aggregate outcomes from complex survey data should be conducted using only statistical techniques that account for estimate variance. Use of standard statistical procedures predisposes to error and should be avoided.

Acknowledgments

This work was funded in part by the Scholarly Concentration Program in The Warren Alpert Medical School of Brown University and National Institutes of Health grant U54GM115677. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. Lumley T. Analysis of complex survey samples. Journal of Statistical Software. 2004;9(1):1-19.
2. Chen X, Gorrell P. An introduction to the SAS survey analysis PROCs. NESUG. 2008.
26. Gillum BS, Graves EJ, Jean L. Trends in hospital utilization: united states, 1988-92. Vital and health statistics Series 13, Data from the National Health Survey. 1996(124):1-71.

27. Benedict K, Derado G, Mody RK. Histoplasmosis-Associated Hospitalizations in the United States, 2001–2012. Open Forum Infectious Diseases. 2016;3(1):ofv219-ofv.

28. White AM, Hingson RW, Pan I, yi H. Hospitalizations for Alcohol and Drug Overdoses in Young Adults Ages 18–24 in the United States, 1999–2008: Results From the Nationwide Inpatient Sample. Journal of Studies on Alcohol and Drugs. 2011;72(5):774-86.

29. Holman RC, Belay ED, Christensen KY, Folkema AM, Steiner CA, Schonberger LB. Hospitalizations for Kawasaki syndrome among children in the United States, 1997–2007. The Pediatric infectious disease journal. 2010;29(6):483-8.

30. Cox S, Johnson CH, Meikle S, Jamieson DJ, Posner SF. Trends in Rates of Hospitalization with a Diagnosis of Substance Abuse among Reproductive-Age Women, 1998 to 2003. Women's Health Issues. 2007;17(2):75-83.

31. Vallabhaneni S, Benedict K, Derado G, Mody RK. Trends in Hospitalizations Related to Invasive Aspergillosis and Mucormycosis in the United States, 2000–2013. Open Forum Infectious Diseases. 2017;4(1):ofw268-ofw.