# Design-Based Analysis of Survey Data

**Brady T. West, Ph.D.**
Research Associate Professor
Survey Research Center
Institute for Social Research
**bwest@umich.edu**

# Workshop Overview

- What is Complex Sample Survey Data?
- What is Design-Based Analysis?
- Appropriate vs. Inappropriate Analytic Approaches
- So What Can Go Wrong? Case Studies in Analytic Error…
- Many Examples Using Real Survey Data, Including Stata/SAS Code
- References

# What is Complex Sample Survey Data?

- Many survey data sets are collected from analytic units randomly selected according to a **stratified, multistage sample design**
- Introductory statistical methods courses generally assume that data arise from a **simple random sample**, where units are randomly selected from some larger list of units
- This is seldom true in real-world samples, especially those selected from geographically widespread populations
- **Complex Samples** generally feature **stratification**, **cluster sampling**, and **weighting of survey respondents**; any of these features count!

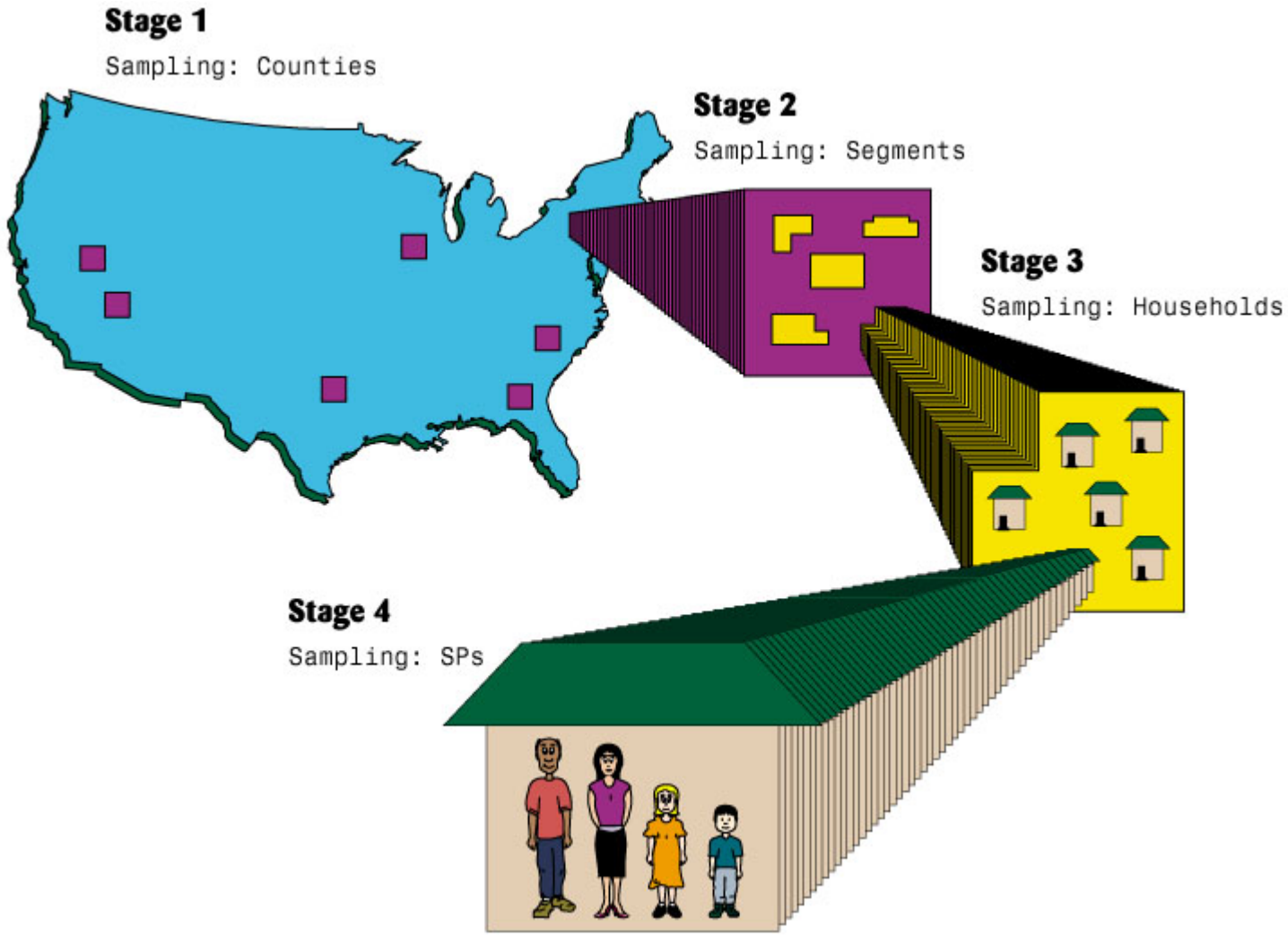
# What is Complex Sample Survey Data?

- **Stratification:**
  - Divide the units in a target population into groups / divisions / **strata** that are homogeneous within and heterogeneous between in terms of the survey measures of interest (requires auxiliary information on the sampling frame)
  - Allocate some fraction of the overall sample to each of the strata, often to minimize the variance of estimates or minimize costs
  - Select samples within each of the strata: could be simple random samples, cluster samples, etc.; **ensures representation from all strata**
  - Combine estimates from the strata into overall estimates; may need weights depending on allocations and probabilities of selection
  - **Reduces the sampling variance of estimates**

# What is Complex Sample Survey Data?

- **Cluster Sampling:**
  - Either within strata or overall, select **clusters** of units of interest (clinics / schools / counties / businesses / etc.); often all that is available on a frame
  - **Saves costs** compared to simple random sampling, especially for geographically widespread populations: sample clusters first, then elements within clusters (rather than units at random)
  - Units within a cluster are often **homogeneous** in terms of measures of interest, but unlike strata, which are fixed by design, clusters are **randomly selected:** we need to account for within-cluster correlation!
  - Clusters are often sampled across **multiple stages** (counties, blocks, etc.)
  - The higher this correlation, the higher the sampling variance introduced by cluster sampling
  - **Increases the sampling variance of estimates**

# What is Complex Sample Survey Data?




Credit: L. Mohadjer, Westat

(Taken from Heeringa et al., 2017, Applied Survey Data Analysis)

# What is Complex Sample Survey Data?

- **Survey Weighting:**
  - In complex samples selected across multiple stages, the probabilities of selection into the sample are often not equal for different units
  - If the probability of selection for a unit is 1/100, that means the unit represents themselves and 99 other units
  - A survey weight begins as the inverse of the probability of selection (would be equal to 100 in this example)
  - Cases with lower probabilities of selection get more weight
  - Weights are often adjusted to correct for differential nonresponse across subgroups, and calibrated to the target population (Valliant and Dever, 2018)
  - **Weights can reduce the bias in estimates (if correlated with key variables)**, but can also **increase sampling variance** if they are highly variable

# Survey Weighting: Save the Date!

- One of the worldwide experts in the development of survey weights, Rick Valliant from the Universities of Maryland and Michigan, will be teaching a hands-on PDHP workshop about survey weighting!
- Save the date: **November 12, 2019, 9am – 1pm**
- This will be a follow-up to today's workshop, providing much more detail (and examples) about where weights come from and how to compute them to reflect different sample designs
- **The Valliant and Dever (2018) book is an essential practical reference on the development of survey weights**

# What is Complex Sample Survey Data?

- **Review:**
  - Stratification ensures population representation and tends to decrease sampling variance
  - Cluster sampling saves costs but tends to increase sampling variance due to within-cluster correlation
  - Weighting for unequal probability of selection and nonresponse can reduce the bias in estimates but can also increase sampling variance if the weights are highly variable
  - **If we wish to make sound inferences about the target populations from which complex samples are selected, we need to carefully account for all of these sample design features in the analysis!**

# What is Design-Based Analysis?

- Most statistical courses introduce **model-based analysis**, where the analyst needs to specify an appropriate model for a variable of interest, and all inference about parameters of interest (means, regression coefficients, etc.) is governed by the specified model
- **Design-based analysis** refers to the expected **sampling distribution** for an estimate of interest that would arise under a given sample design: if we were to draw thousands of samples under the same design, where would it be centered, and how variable would it be?
- With design-based analysis, we estimate the parameters of interest based on the sample, and estimate the variance of those parameters **with respect to the sample design**; all inference is governed by the random sampling process!

# What is Design-Based Analysis?

- Design-based analysis focuses inference on a well-defined and **finite target population** from which the complex sample was selected (**our focus**!)
- With model-based analysis, all inference is based on the specified probability model, and there is no notion of a finite population
- If a model is well-specified and finite population inference is not critical, model-based analyses can be much more efficient and powerful…
- …but if the model is incorrectly specified, the complex sampling features are informative about estimates of interest, and finite population inference is critical, model-based analysis can be critically incorrect!
- See Hansen et al. (1983) for an excellent overview

# Appropriate Design-Based Approaches

1. Use the final survey weights for estimation, especially if they shift the estimates of interest substantially relative to the corresponding shift in the standard errors (weights may not be necessary!)
2. Account for the effects on the sampling variance of weighting, stratification, and cluster sampling via specialized techniques for estimating the variance of the sampling distribution:
   - Taylor Series Linearization (default in most software)
   - Replication Techniques (jackknife, BRR, bootstrap)
3. Generate confidence intervals for parameters of interest with appropriate degrees of freedom based on the sample design (generally, # of clusters - # of strata: clusters matter for variances!)

# Inappropriate Approaches

- **DO NOT:** Ignore the weights entirely, especially if they are correlated with variables of interest and finite population inference is desired
- **DO NOT:** Ignore the stratification and the cluster sampling when estimating sampling variance; information about the design features is usually provided in public-use survey data sets, via design codes or replicate weights
- **DO NOT:** Consider model-based approaches that completely ignore the complex sample design features (e.g., an appropriate model for a given variable in a survey data set should probably use random cluster effects to account for within-cluster correlations)
- **DO NOT:** Use software procedures that do not process the complex sample design features correctly (e.g., frequency weights)

# So What Can Go Wrong?

- A failure to conduct appropriate design-based analyses when finite population inference is desired can lead to erroneous conclusions about populations of interest and incorrect inferences
- Considering the larger **Total Survey Error** framework, an important source of error in survey estimates and inference is therefore **analytic error** of this type
- **Most Common:** Secondary analysts without training in this type of analytic approach apply “standard” methods to complex sample survey data sets, and write papers prone to erroneous inferences

# Analytic Error…

- If secondary analysts of survey data ignore essential features of **complex sample designs** in analyses (whether they are design-based or model-based), inferences based on the survey data may well be biased or erroneous:
  - **Weights** (for unbiased estimation)
  - **Stratum codes** (for increased efficiency of estimates)
  - **Cluster codes** (to capture losses in efficiency due to cluster sampling)
- A failure to account for these features in secondary analysis will negate the substantial federal resources dedicated to minimizing other sources of TSE for a given survey

# Past Research on Analytic Error

- All the way back to Deming (1944): "…Errors in Curve Fitting…"
- Smith (2011): Formally an important component of TSE
- West et al. (2017, *Total Survey Error In Practice*):
  - Review of 100 peer-reviewed journal articles presenting secondary analyses of survey data
  - Public health focus
  - **Apparent** failures to account for weights and complex sampling features in estimation is quite common (do authors provide code for proof?)
  - Subpopulation analyses in particular are rarely performed correctly in design-based analyses

# Question: How Big of a Problem Is This?

1. What is the prevalence of “apparent” analytic error in other peer-reviewed journal articles, conference presentations, book chapters, and technical reports?
2. Are there trends across decades in the prevalence of these errors?
3. Do journal-specific features dictate error rates (e.g., impact factors, having statisticians on editorial boards)?
4. What are the implications of making these errors for estimation and inference?
5. Do the same problems exist for establishment surveys?

# Study Design: NCSES Surveys (West et al. 2016)

- First, perform a **meta-analysis** of approaches used:
  - Stratified sample of 150 research products analyzing survey data from the **Science and Engineers Statistical Data System (SESTAT),** sponsored by the **National Center for Science and Engineering statistics (NCSES)**
  - Google Scholar searches, organized by decade
  - Products ordered by decade; systematic sampling
  - Sample 50 articles presenting secondary analyses of each of three main SESTAT surveys (all having complex samples):
    - National Survey of College Graduates (NSCG)
    - Survey of Doctorate Recipients (SDR) (*2 ineligible, no replacements*)
    - National Survey of Recent College Graduates (NSRCG) (*3 ineligible*)

# Study Design, cont'd

- Code the 150 articles on the following items:
  - Year available online, type of research product
  - **Did the analysis account for weights?**
  - **Did the analysis account for design features in variance estimation?**
  - Was the approach design-based or model-based?
  - Did the authors use appropriate software?
  - **Did the authors perform appropriate subpopulation analyses (for design-based analyses)?**
  - Did the authors describe their results with respect to the target population?
  - **For journal articles:** Impact factor? Statisticians on editorial board? Guidelines for survey data analysis on web site?

# Study Design, cont'd

- Next, examine the **implications of making analytic errors** for analyses of SESTAT data:
  - Download the 2010 public-use SDR and NSCG data
  - Obtain the replicate weights for variance estimation from NCSES
  - Review the sampled articles and work with NCSES staff to identify:
    - Key descriptive estimates
    - Regression models of substantive interest

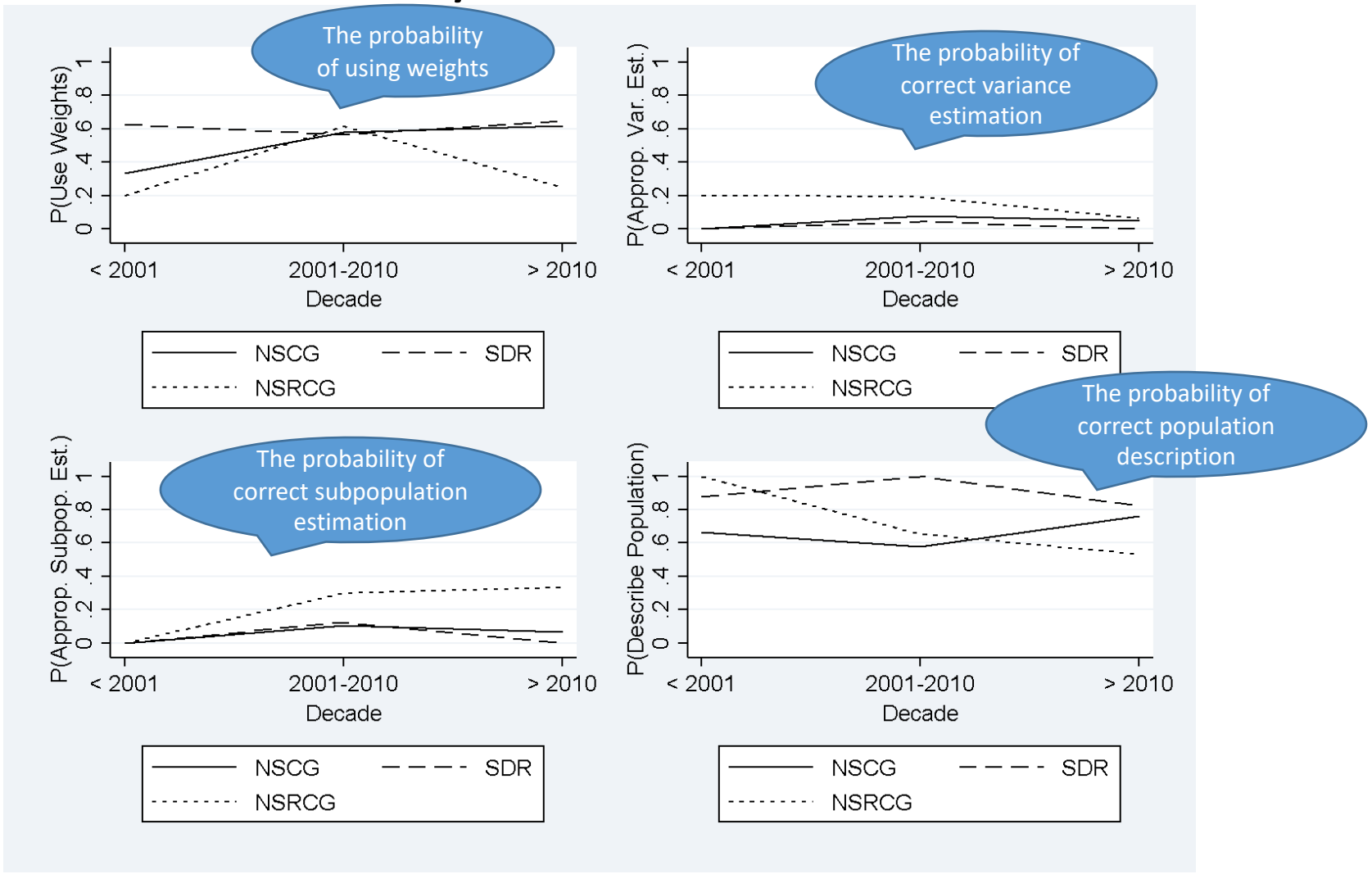
# Study Design, cont'd

- Perform three types of analyses for both descriptive and analytic (regression) parameters:
  - Fully accounting for complex sampling features
  - Accounting for weights only, using Taylor Series Linearization for variance estimation
  - Ignoring complex sampling features entirely
- Consider ratios of variance estimates to assess *misspecification effects,* due to ignoring key sampling features in the analyses

# Results: Meta-Analysis

| | SDR (n = 48) | NSCG (n = 50) | NSRCG (n = 47) | Overall (n = 145) |
|---|---|---|---|---|
| **Indicator** | **% (SE)** | **% (SE)** | **% (SE)** | **% (SE)** |
| Accounted for sampling weights in analyses | 60.4% (7.3%) | 58.0% (7.2%) | 44.7% (6.9%) | **54.5% (4.2%)** |
| Accounted for complex sampling in variance estimation | 2.1% (2.1%) | 6.0% (3.4%) | 14.9% (5.3%) | **7.6% (2.2%)** |
| Used design-based approach (vs. model-based) | 50.0% (7.2%) | 76.0% (6.1%) | 37.0% (7.0%) | 55.6% (4.2%) |
| Used appropriate* subpopulation estimation [2] | 4.2% (4.1%) | 8.1% (4.6%) | 30.8% (13.1%) | **10.7% (3.6%)** |
| Described results with respect to the population (vs. the sample) | 91.7% (4.0%) | 66.0% (6.8%) | 65.2% (6.9%) | 74.3% (3.7%) |

# Results: Meta-Analysis



# Results: Meta-Analysis

- No variance in “apparent” error rates depending on the type of research product
- *Journal articles were the least likely to use appropriate variance estimation (!)*
- Having statisticians on the editorial board or as reviewers increased the probability of using appropriate methods
- Model-based approaches were particularly likely to ignore design features; **vague references to “robust” standard errors**
- Almost no references to the use of appropriate software
- **Journals almost never provide guidance with regard to secondary analysis of survey data on their web sites**

# Results: Implications of Errors

- Among 8 categorical variables in the 2010 SDR, **inferences related to estimated distributions in the population would change for 5 of the variables** if design features were ignored:
  - Current salary, race/ethnicity, attending professional meetings in the past year, major field of study, and labor force status (*changes generally small*)
- Among 10 continuous and categorical variables in the 2010 NSCG, **there would be substantial changes in inference for 9 of the variables!**

# Results: Implications of Errors

- **An extreme example from the 2010 NSCG:**
  - Primary Job in Science and Engineering (key indicator!):
    - Fully accounting for complex sampling:
      - 30.38% (SE = 0.30%)
    - Accounting for final weights only:
      - 30.38% (SE = 0.39%; *no gains from stratification!*)
    - Completely ignoring complex sampling:
      - 54.94% (SE = 0.20%) (!!!!!)
- **Weights were highly correlated with several other measures:** race/ethnicity, highest degree, salary, major degree, etc.

# Results: Implications of Errors

- In a logistic regression model fitted to a binary indicator of having an annual salary > $150K, inference related to the interaction between major degree and race/ethnicity would change:
  - Design-adjusted Wald test, fully accounting: p = 0.019
  - Design-adjusted Wald test, weights only: p = 0.278
  - Wald test, ignoring all features: p = 0.139
- No major effects on an ordinal regression model of hours worked per week, predicted by race/ethnicity, principal job type, and their interaction

# Results: Implications of Errors

- In a regression model fitted to log-transformed current salary in the 2010 NSCG, the main effect of having an S & E degree would change completely:
  - Fully accounting for complex sampling: 0.16 (SE = 0.03)
  - Ignoring complex sampling features: 0.02 (SE = 0.02)
- In a logistic regression model for having an S & E job in the 2010 NSCG, inference about the interaction between race/ethnicity and gender would change radically:
  - Fully accounting: $p = 0.1944$; Ignoring: $p < 0.0001$

# Summary of Results

- **Meta Analysis:** Secondary analysts of SESTAT data only account for weights about 50% of the time, and rarely perform variance estimation or subpopulation analysis correctly (plus, no trends over time)
  - Consistent with initial review of public health research
  - Some journal features help (e.g., statistical reviewers)
- **Implications of Errors:** When failing to account for complex sampling features, implications for inference can be quite severe!
  - Weights can be highly correlated with key variables
  - Gains in efficiency from stratified sampling are missed

# Study Design: 2013 BRDIS (West and Sakshaug, 2018)

- No work to date has considered the analytic error problem in the **establishment survey** context
- We reviewed the methods used in heavily-cited publications presenting analyses of establishment survey data
- We then considered the implications of failing to account for sample design features in a real establishment survey

# Study Design: 2013 BRDIS

- We again used Google Scholar to identify highly-cited articles presenting secondary analyses of establishment survey data
- We then performed alternative analyses of data (e.g., with and without weights) from the 2013 Business Research and Development and Innovation Survey (BRDIS)
- We also considered the effects of ignoring the stratification in the BRDIS sample design

# BRDIS Disclaimer

- This research uses data from the U.S. Census Bureau's Longitudinal Employer Household Dynamics Program, which was partially supported by National Science Foundation Grants SES-9978093, SES-0339191 and ITR-0427889; National Institute on Aging Grant AG018854; and grants from the Alfred P. Sloan Foundation. Any opinions and conclusions expressed herein are those of the author(s) and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed.

# Results: Meta-Analysis

- Of the 10 most highly-cited articles (median citation count = 412.5, median impact factor = 2.86), only two out of ten even mentioned using weights at all
- None of the articles mentioned accounting for the sample design features (e.g., stratification) when estimating variances

# Results: Estimation of Means

| | | Approach 1 (no weights, no strata) | Approach 2 (weights, no strata) | Approach 3 (weights, strata) |
|---|---|---|---|---|
| **Variable** | Sample Size | Mean (SE) | Mean (SE) [95% CI] | Mean (SE) [95% CI] |
| **Total Salary Expenditures (Millions)** | 31,000 | 539.85 (34.42) | 20.45 (1.04) [18.42, 22.48] | 20.45 (1.00) [18.48, 22.42] |
| **Total Worldwide Employees (Thousands)** | 31,000 | 1.28 (0.09) | 0.07 (<0.01) [0.06, 0.07] | 0.07 (<0.01) [0.06, 0.07] |
| **Total U.S. expenditures on R&D (Thousands)** | 28,000 | 11238.10 (941.62) | 260.31 (20.24) [220.63, 299.99] | 260.31 (17.16) [222.28, 293.33] |
| **Total Worldwide expenditure on R&D (Thousands)** | 28,000 | 9867.83 (831.30) | 229.63 (17.86) [194.62, 264.65] | 229.63 (17.16) [196.00, 263.26] |

# Results: Regression Modeling

| U.S. R&D Exp | Approach 1 | Approach 2 | Approach 3 |
|---|---|---|---|
| Predictor | Coefficient (SE) | Coefficient (SE) | Coefficient (SE) |
| Intercept | 3052.35 (855.22)*** | 254.42 (19.78)*** | 254.42 (16.53)*** |
| Total Salary Expenditures (Millions, Mean-Centered) | 9.48 (0.19)*** | 9.64 (2.58)*** | 9.64 (2.49)*** |
| Total Worldwide Employees (Thousands, Mean-Centered) | 2951.54 (107.46)*** | 1967.67 (894.44)* | 1967.67 (750.43)** |
| Interaction | -0.01 (<0.01)*** | -0.01 (<0.01)*** | -0.01 (<0.01)*** |
| Sample Size | 28,000 | 28,000 | 28,000 |
| R-squared | 0.1883 | 0.1738 | 0.1738 |

# BRDIS 2013: Summary of Results

- Failing to account for weights in the 2013 BRDIS led to substantial changes in inference, for both means and regression coefficients (see the tables of estimates above, and note that sample sizes have been rounded)
- A failure to account for the stratified sample design of the 2013 BRDIS led to overly conservative inferences

# Another Important Recent Study

- This is not just me, tooting my horn from my soapbox! 
- Khera et al. (2017, JAMA)
- Observational study of 120 out of 1,082 published studies in 2015 and 2016 using the National Inpatient Sample (NIS)
- **79 of the 120 studies (68.3%) did not account for weights, cluster sampling or stratification (!!!)**
- **Implication:** More than 2/3 of studies published in peer-reviewed journals using NIS data may report erroneous inferences (**Yikes!)**

# Examples using the 2011 NIS Data

- **So what can go wrong in the NIS?** We now turn to examples of alternative approaches to analyzing data from the 2011 NIS
- **Dependent Variables:** DIED (1 = in-hospital death during admission, 0 = otherwise), TOTCHG (total charges for hospitalization)
- **Covariates:** FEMALE (binary), ELECTIVE (elective or non-elective admission), AGE (age at admission), LOS (length of hospitalization in days), PAY1 (primary payer, categorical), RACE (categorical)
- We will consider descriptive analyses for all variables, and regression models for the two dependent variables

# Examples using the 2011 NIS Data

- **Always check the documentation to make sure that you have a good handle on the complex sample design features!**

  https://www.hcup-us.ahrq.gov/reports/methods/2015-09.pdf

- **Final Survey Weights:** DISCWT
- **Sampling Strata:** NIS_STRATUM
- **Sampling Clusters (Hospitals):** HOSPID

# Descriptive Analysis #1: Ignore Design Features Entirely (naïve approach)

**SAS Code:**

```
proc freq data = desktop.nis2011;
  tables died female elective pay1 race;
run;

proc means data = desktop.nis2011 n mean stderr;
  var totchg age los;
run;
```

**Stata Code:**

```
prop died
prop female
…
prop race

mean totchg
mean age
mean los
```

# Descriptive Analysis #2: Use Weights for Estimation, Ignore Other Design Features

## SAS Code:

```
proc surveyfreq data = desktop.nis2011;
  tables died female elective pay1 race;
  weight discwt;
run;

proc surveymeans data = desktop.nis2011;
  var totchg age los;
  weight discwt;
run;
```

## Stata Code:

```
svyset [pweight = discwt]
svy: prop died
…
svy: prop race

svy: mean totchg
svy: mean age
svy: mean los
```

# Descriptive Analysis #3: Account for All Complex Sample Design Features

**SAS Code:**

```
proc surveyfreq data = desktop.nis2011;
  tables died female elective pay1 race;
  weight discwt;
  cluster hospid;
  stratum nis_stratum;
run;

proc surveymeans data = desktop.nis2011;
  var totchg age los;
  weight discwt;
  cluster hospid;
  stratum nis_stratum;
run;
```

**Stata Code:**

```
svyset hospid [pweight = discwt], strata(nis_stratum)
svy: prop died
…
svy: prop race

svy: mean totchg
svy: mean age
svy: mean los
```

# Comparisons of Descriptive Analysis Approaches

- Estimated Percentages (Standard Errors) for Selected Binary Variables:

| | Approach 1 (Ignore) | Approach 2 (Weights Only) | Approach 3 (Fully Accounting) |
|---|---|---|---|
| DIED | | | |
| Yes | 1.90 (0.005) | 1.91 (0.005) | 1.91 (0.029) |
| No | 98.10 (0.005) | 98.09 (0.005) | 98.09 (0.029) |
| ELECTIVE | | | |
| Yes | 23.62 (0.015) | 23.70 (0.015) | 23.70 (0.567) |
| No | 76.38 (0.015) | 76.30 (0.015) | 76.30 (0.567) |

- Minimal changes in estimates (weights not correlated with indicators!), but notable increases in standard errors when fully accounting for design

# Comparisons of Descriptive Analysis Approaches

- Estimated Means (Standard Errors) for Continuous Variables:

| | Approach 1 (Ignore) | Approach 2 (Weights Only) | Approach 3 (Fully Accounting) |
|---|---|---|---|
| Total Charges | 35466.78 (23.37) | 35415.00 (23.58) | 35415.00 (983.59) |
| Age at Admission | 49.56 (0.01) | 49.62 (0.01) | 49.62 (0.39) |
| LOS (Days) | 4.59 (<0.01) | 4.60 (<0.01) | 4.60 (0.05) |

- Again, minimal changes in estimates (weights not correlated with variables!), and notable increases in standard errors when fully accounting for the complex sample design

# Don’t Forget About Subpopulations!

- A very common mistake when performing design-based analysis (and specifically when using Taylor Series Linearization to estimate sampling variance) is incorrect subpopulation inference
- **Do not simply delete cases that do not belong to the subpopulation before doing the analysis:** treats subsample size as fixed (it usually isn’t), and introduces risk of deleting sample design information
- **Instead**, use built in subpopulation estimation options, with binary indicator variables for subpopulation cases!
  - **Stata:** svy, subpop(indicator_name): mean var_name
  - **SAS:** DOMAIN indicator_name;

# Regression Analysis #1: Ignore Design Features Entirely (naïve approach)

**SAS Code:**

```
proc surveylogistic data = desktop.nis2011;
   class race pay1 / param = ref;
   model died (event = "1") = female elective age los race pay1;
run;

proc surveyreg data = desktop.nis2011;
   class race pay1;
   model lntotchg = female elective age los race pay1 / clparm solution;
run;
```

**Stata Code:**

```
logit died female elective age los i.race i.pay1

reg lntotchg female elective age los i.race i.pay1
```

# Regression Analysis #2: Use Weights for Estimation, Ignore Other Design Features

**SAS Code:**

```
proc surveylogistic data = desktop.nis2011;
  class race pay1 / param = ref;
  model died (event = "1") = female elective age los race pay1;
  weight discwt;
run;

proc surveyreg data = desktop.nis2011;
  class race pay1;
  model lntotchg = female elective age los race pay1 / clparm
solution;
  weight discwt;
run;
```

## Stata Code:

**svyset [pweight = discwt]**

**svy: logit** died female elective age los i.race i.pay1

**svy: reg** lntotchg female elective age los i.race i.pay1

# Regression Analysis #3: Account for All Complex Sample Design Features

**SAS Code:**

```
proc surveylogistic data = desktop.nis2011;
  model died (event = "1") = female elective age los race pay1;
  weight discwt;
  cluster hospid;
  stratum nis_stratum;
run;

proc surveymeans data = desktop.nis2011;
  model lntotchg = female elective age los race pay1 / clparm solution;
  weight discwt;
  cluster hospid;
  stratum nis_stratum;
run;
```

## Stata Code:

**svyset hospid [pweight = discwt], strata(nis_stratum)**

**svy: logit** died female elective age los i.race i.pay1

**svy: reg** lntotchg female elective age los i.race i.pay1

# Comparisons of Linear Regression Analyses

- Estimates (Standard Errors) of Selected Coefficients:

| | Approach 1 (Ignore) | Approach 2 (Weights Only) | Approach 3 (Fully Accounting) |
|---|---|---|---|
| Intercept | 8.711 (0.003) | 8.716 (0.003) | 8.716 (0.070) |
| Female | -0.099 (0.001) | -0.099 (0.001) | -0.099 (0.007) |
| Elective | 0.360 (0.001) | 0.358 (0.001) | 0.358 (0.019) |
| Age | 0.019 (<0.001) | 0.019 (<0.001) | 0.019 (<0.001) |
| LOS | 0.070 (<0.001) | 0.069 (<0.001) | 0.069 (0.003) |
| Race 1 | -0.043 (0.002) | -0.043 (0.002) | -0.043 (0.060) |
| Pay 1 | -0.311 (0.002) | -0.318 (0.002) | -0.318 (0.032) |

- **Same pattern:** Minimal changes in estimates (weights not informative about coefficients!), but notable increases in standard errors when fully accounting for design; model well-specified?

# Comparisons of Logistic Regression Analyses

- Estimates (Standard Errors) of Selected Coefficients:

| | Approach 1 (Ignore) | Approach 2 (Weights Only) | Approach 3 (Fully Accounting) |
|---|---|---|---|
| Intercept | -5.582 (0.024) | -5.588 (0.025) | -5.588 (0.126) |
| Female | -0.308 (0.006) | -0.308 (0.006) | -0.308 (0.007) |
| Elective | -0.691 (0.008) | -0.668 (0.009) | -0.668 (0.050) |
| Age | 0.042 (<0.001) | 0.043 (<0.001) | 0.043 (0.001) |
| LOS | 0.027 (<0.001) | 0.026 (<0.001) | 0.026 (0.002) |
| Race 1 | -0.154 (0.017) | -0.161 (0.017) | -0.161 (0.046) |
| Pay 1 | -0.789 (0.016) | -0.785 (0.016) | -0.785 (0.148) |

- **Same pattern:** Minimal changes in estimates (weights not informative about coefficients!), but notable increases in standard errors when fully accounting for design; model well-specified?

# Summary of 2011 NIS Analyses

- While weights may not have been informative about the estimates of interest in these examples (and the models may have been well-specified), **this will not always be the case**
- Accounting for the strata and clusters used in the actual sample design **matters for inference (especially when forming CIs);** *it is not enough to just use weights!*
- Degrees of freedom for t-tests / confidence intervals: when ignoring design features or only using weights, more than **7 million**
- When fully accounting for the complex sample design: **886** (what if there were only 30 clusters and 15 strata in a design?)

# Working with NIS Data from 2012+

- Beginning in 2012, all NIS microdata are **self-weighting**
  - Every case has the same final weight: no need to use weights in analysis!
- **HOWEVER:** This doesn't mean that you can ignore the sampling strata and sampling clusters when estimating the variances of point estimates!
- We consider an example of what can go wrong on the next slide…inferences could be significantly affected!

# Working with NIS Data from 2012+, cont'd

- Consider simple estimation of means for **total charges**, **age in years at admission**, and **length of stay (days)**
- Use linearization to compute design-adjusted standard errors
- Approach 1: ignore strata (196) and clusters (4,378)
- Approach 2: account for strata and clusters (e.g., svy: mean)

| Variable | Sample Size | Estimated Mean | Approach 1: SE | Approach 2: SE |
|---|---|---|---|---|
| Total Charges | 7,145,492 | 36,704.22 | 25.84 | 446.25 |
| Age in Years | 7,293,627 | 48.59 | 0.01 | 0.20 |
| Length of Stay | 7,294,735 | 4.51 | <0.01 | 0.02 |

# ADDITIONAL EXAMPLES OF DESIGN-BASED ANALYSIS OF SURVEY DATA USING PROCEDURES IN SAS AND STATA

# Getting the Workshop Data Sets into SAS and Stata

- **Download the data sets from the *Applied Survey Data Analysis* web site:**
  - http://www.isr.umich.edu/src/smp/asda
  - Click on “Analysis Examples Data Sets (Stata and SAS Format) – First Edition”
  - Download the .zip file
  - Extract the SAS and Stata files into **a directory that you can access**

- **Example SAS Code for reading in NHANES Data:**

```
libname dir "O:\SMP\Brady\";
data nhanes;
  set dir.nhanes_analysis_ex_c1_c10_2011;
run;
```

- **Example Stata Code for reading in NHANES Data:**

```
. use "O:/SMP/Brady/nhanes_analysis_examples_c1_c10_2011.dta", clear
```

# Important Notes About the Software Packages

**SAS® SURVEY procedures**

- Procedures are readily available as a part of SAS/STAT
- Powerful data management capabilities built in to SAS
- Some minor limitations in terms of available analyses; for example, Poisson regression models cannot be fitted to complex samples.
- Each new version adds more flexibility and features

# Important Notes, cont'd

**Stata `svy` Commands**

- By far the most design and analysis options among the general-purpose statistical software packages
- Very easy-to-use commands for survey estimation (we will look at several examples); after declaring design codes, standard commands are simply modified with `svy:` and any specific options [e.g., `vce(jackknife)`]
- Selected graphing commands enable "pweights" to be used so that graphs reflect estimated population distributions
- Currently many more modeling options than SAS, especially for generalized linear models (e.g., `svy: poisson`)

# Descriptive Statistics: Totals (Stata)
## Total Persons w/ Lifetime MDE by Marital Status (NCS-R)

```
. use "O:/SMP/Brady/ncsr_analysis_examples_c1_c10_2011.dta“, clear
* Generate population weights, given that weights are scaled to sum to 1
. gen popweight = ncsrwtsh * 209128094 / 9282
. svyset seclustr [pweight = popweight], strata(sestrat)
. svy: total mde, over(mar3cat)
. mat list e(b)
. estat effects
```

| Subpop | *n* | Estimated Total Lifetime MDE | Estimated Standard Error | 95% CI | $d^2(\hat{Y})$ |
|---|---|---|---|---|---|
| **Married** | **5322** | **20,304,190** | **1,584,109** | **(17,199,395, 23,408,986)** | **6.07** |
| **Sep./Wid./Div.** | **2017** | **10,360,671** | **702,601** | **(8,983,558, 11,737,783)** | **2.22** |
| **Never Married** | **1943** | **9,427,345** | **773,137** | **(7,912,024, 10,942,667)** | **2.95** |

# Descriptive Statistics: Totals (SAS)
# Total Persons w/ Lifetime MDE by Marital Status (NCS-R)

```
data ncsr;
  set unc.ncsr_analysis_ex_c1_c10_2011;
  ncsrwtsh_pop = ncsrwtsh * 209128094 / 9282;
run;

proc surveymeans data=ncsr nobs sum df stderr clsum ;
  strata sestrat ;
  cluster seclustr ;
  weight ncsrwtsh_pop ;
  var mde ;
  domain mar3cat ;
run ;
```

**(Output not shown here.)**

# Descriptive Statistics: Means and Proportions
# Mean total household assets (HRS 2006).

```
data hrs; set unc.hrs_analysis_ex_c1_c9_2011; run;
proc surveymeans data=hrs nobs df mean stderr clm ;
  strata stratum ;  cluster secu ;  weight kwgthh ;
  domain kfinr ;  /* want kfinr = 1 for financial reporters of households */
  var h8atota ;
run ;
```

**Stata:**

```
. svyset secu [pweight = kwgthh], strata(stratum)
. svy, subpop(if kfinr == 1): mean h8atota
. estat effects
```

| n | df | $\overline{y}_w$ | $se(\overline{y}_w)$ | $CI_{.95}(\overline{y}_w)$ | $d^2(\overline{y}_w)$ |
|---|---|---|---|---|---|
| 11,942 | 56 | $527,313 | $28,012 | ($471,196, $583,429) | 1.52 |

# Estimation of Percentiles: SAS
# Total Household Assets (HRS 2006)

```
proc surveymeans data = hrs q1 median q3 ;
  strata stratum ;
  cluster secu ;
  weight kwgthh ;
  domain kfinr ;
  var h8atota ;
run ;
```

Data Summary

| | |
|---|---|
| Number of Strata | 56 |
| Number of Clusters | 112 |
| Number of Observations | 12558 |
| Number of Observations Used | 11942 |
| Number of Obs with Nonpositive Weights | 616 |
| Sum of Weights | 53853171 |

Quantiles

| Variable | Label | Percentile | Estimate | Std Error | 95% Confidence Limits | |
|---|---|---|---|---|---|---|
| H8ATOTA | Total Assets: HH | 25% Q1 | 39853 | 3258.139382 | 33326.094 | 46379.769 |
| | Total Assets: HH | 50% Median | 183309 | 9977.330641 | 163322.041 | 203296.031 |
| | Total Assets: HH | 75% Q3 | 495931 | 17394 | 461086.518 | 530776.236 |

# Univariate Analysis: Multinomial Variables

- Stata: `svy: tab catvar, obs se`
- SAS: PROC SURVEYFREQ
- NHANES Example in SAS:

```
proc surveyfreq data = nhanes;
  weight wtmec2yr;
  stratum sdmvstra;
  cluster sdmvpsu;
  tables age18p*bp_cat / row cl;
run;
```

# Estimating Proportions for Multinomial Variables (Stata)

```
. svyset sdmvpsu [pweight=wtmec2yr], strata(sdmvstra)
. svy, subpop(age18p): tab bp_cat, obs se deff
* Alternative commands (same results)
. svy, subpop(age18p): prop bp_cat
. estat effects
```

| Blood Pressure Category | *n* | Estimated Proportion | Linearized SE | 95% CI | Design Effect |
|---|---|---|---|---|---|
| Normal | 2441 | 0.471 | 0.011 | (0.449, 0.493) | 2.49 |
| Pre-Hypertension | 1988 | 0.419 | 0.012 | (0.395, 0.442) | 2.89 |
| Stage 1 Hypertension | 470 | 0.086 | 0.006 | (0.074, 0.099) | 2.47 |
| Stage 2 Hypertension | 158 | 0.024 | 0.002 | (0.019, 0.029) | 1.25 |

# A Quick Note: Forming Confidence Intervals for Proportions

- Please be aware that the standard symmetric Wald intervals computed for proportions by default by most software procedures have **very poor coverage properties** (Franco et al., 2019)
- Adjusted intervals using the logit transformation (which is used by svy: tab in Stata) offer improvements, but are still not ideal
- Franco et al. (2019) outline state-of-the-art methods for computing confidence intervals for proportions estimated from survey data that have **substantially improved coverage properties**
- While the article doesn't explicitly provide R software for computing the adjusted intervals, it clearly describes how to perform the computations using output provided by SAS and Stata
- Please consider whether your inferences would change when computing these to supplement your default confidence intervals!
- **Reference:**

Franco, C., Little, R.J.A., Louis, T.A., and Slud, E.V. (2019). Comparative Study of Confidence Intervals for Proportions in Complex Sample Surveys. *Journal of Survey Statistics and Methodology*, 7, 334-364.

# Estimation of Total and Row Proportions (Stata)

```
. svyset seclustr [pweight = ncsrwtlg], strata(sestrat)
. svy: tab sex mde, se ci deff
. svy: tab sex mde, row se ci deff
```

| Description | Parameter | Estimated Proportion | Linearized SE | 95% CI | Design Effect |
|---|---|---|---|---|---|
| Total Proportions | | | | | |
| Male, no MDE | $\pi_{A0}$ | 0.406 | 0.007 | (0.393, 0.421) | 1.87 |
| Male, MDE | $\pi_{A1}$ | 0.072 | 0.003 | (0.066, 0.080) | 1.64 |
| Female, no MDE | $\pi_{B0}$ | 0.402 | 0.005 | (0.391, 0.413) | 1.11 |
| Female, MDE | $\pi_{B1}$ | 0.120 | 0.003 | (0.114, 0.126) | 0.81 |
| Row Proportions | | | | | |
| No MDE \| Male | $\pi_{0\|A}$ | 0.849 | 0.008 | (0.833, 0.864) | 2.08 |
| MDE \| Male | $\pi_{1\|A}$ | 0.151 | 0.008 | (0.136, 0.167) | 2.08 |
| No MDE \| Female | $\pi_{0\|B}$ | 0.770 | 0.006 | (0.759, 0.782) | 0.87 |
| MDE \| Female | $\pi_{1\|B}$ | 0.230 | 0.006 | (0.218, 0.241) | 0.87 |

# Estimation of Total and Row Proportions (SAS)

```
proc surveyfreq data = ncsr;
   weight ncsrwtsh;
   stratum sestrat;
   cluster seclustr;
   tables sexf*mde / row cl deff;
run;
```

# Alcohol Dependence vs. Education Level for Young Adults, Age 18-28. Source: NCS-R.

In Stata:

```
. svyset seclustr [pweight = ncsrwtlg], strata(sestrat)
. svy, subpop(if 18<=age & age<29): tab ed4cat ald, row se ci deff
```

In SAS:

```
proc surveyfreq data=ncsr ;
   strata sestrat ;
   cluster seclustr ;
   weight ncsrwtlg ;
   tables age29*ed4cat*ald / row deff chisq;
run ;
```

# Design-adjusted Analysis of Alcohol Dependence vs. Education Level for Young Adults Age 18-28. Source: NCS-R.

| **Education Level (Grades)** | **Alcohol Dependence (ALD) Row Percentages (se)** | | |
|---|---|---|---|
| | **0 = No** | **1 = Yes** | **Total** |
| 0-11 | 0.909 (0.029) | 0.091 (0.029) | 1.000 |
| 12 | 0.951 (0.014) | 0.049 (0.014) | 1.000 |
| 13-15 | 0.951 (0.010) | 0.049 (0.010) | 1.000 |
| 16+ | 0.931 (0.014) | 0.069 (0.014) | 1.000 |
| Total | 0.940 (0.009) | 0.060 (0.009) | 1.000 |
| **Test of** | | | |
| Unadjusted $X^2$ | P ( > $X^2_{Pearson}$) | Rao-Scott $F$ | P ($F_{2.75,\ 115.53}$ > $F_{R\text{-}S}$) |
| $X^2_{Pearson}$=27.21 | ***p* < 0.0001** | $F_{R\text{-}S,Pearson}$ = 1.64 | ***p* = 0.18** |
| **Parameters of the Rao-Scott Design-Adjusted Test** | | | |
| $n_{18\text{-}29}$ = 1275 | Design *df* = 42 | ***GDEFF* = 6.62** | *a* = 0.56 |

# Procedures in SAS and Stata for Linear Regression Analysis of Survey Data

- **SAS**: PROC SURVEYREG
- **Stata**: `svy: regress`
- We will look at examples of their use shortly!

# To Weight or Not To Weight?

- There is generally little debate about the need to use weights for estimation of descriptive parameters (means, proportions, totals, etc.) using survey data
- There is more debate regarding the use of weights when fitting regression models to survey data
- In the finite population modeling framework, **model specification is important:**
  - Poor specification of a model, when combined with weighted estimation, will result in unbiased estimates of the regression coefficients in a **poor finite population model**
  - Good specification of a model, when combined with weighted estimation, will result in inflated standard errors (no need to use weights if model has been correctly specified!)
  - See Korn and Graubard (1999, Chapter 4) for good examples

# To Weight or Not To Weight?

- **Compare models fitted with and without weights:**
  - Large changes in parameter estimates suggest model misspecification, and that at the very least weights should be used to obtain unbiased estimates
  - Small changes in parameter estimates and large changes in standard errors would suggest appropriate specification, and minimal problems with using unweighted models
  - Check sensitivity of results to different approaches!
  - **In Practice:** Fit an **unweighted model** that includes all predictors of interest, and add 1) the weight, and 2) all two-way interactions between the weights and other predictors
  - Perform a Wald test of the additional coefficients:
    - Significant: Weights important
    - Not Significant: Ignore weights

# Example: Fitting a Multiple Regression Model to the 2005-2006 NHANES Data.

- **National Health and Nutrition Examination Survey**
  - **Dependent Variable**: Diastolic BP
  - **Predictor Variables**:
    - GENDER (1 = Male, 2 = Female)
    - RACER (1 = White, 2 = Black, 3 = Other)
    - AGE (age in years, centered to estimated mean age of 45.60), AGE SQUARED
    - POVERTY (1 = Not Poor, 2 = Poor)

# Fitting a Multiple Regression Model to the 2005-2006 NHANES Data on Diastolic BP

Stata Code:

```
. svyset sdmvpsu [pweight = wtmec2yr], strata(sdmvstra)
. svy, subpop(age18p): regress bpxdi1_1 i.ridreth1 i.marcat
  i.riagendr agec agecsq
. estat effects
```

SAS Code:

```
proc surveyreg data=nhanes ;
   class ridreth1 marcat ;
   strata sdmvstra ; cluster sdmvpsu ; weight wtmec2yr ;
   domain age18p ;
   model bpxdi1_1 = ridreth1 marcat female agec agecsq / solution deff clparm ;
   output out=outdiag2 p=phat r=resid ;
run ;
```

# Fitting a Multiple Regression Model to 2005-06 NHANES Data on Diastolic BP: Multi-parameter Wald Tests

- **Stata:** use `test`

Example:

```
. test 2.marcat 3.marcat
```

- **SAS:** multi-parameter design-based Wald tests for categorical predictors are included in the output by default; the **contrast** statement (e.g., contrast 'age' agec 1, agecsq 1;) can be used to set up more complex hypothesis tests (see ASDA web site or SAS documentation for examples)

# Fitting a Multiple Regression Model to the 2005-2006 NHANES Data on Diastolic BP: Estimates ($R^2 = 0.134$)

| Predictor | Est. | Linearized SE | *t*-statistic (df) | *p*-value | 95% CI | DEFF |
|---|---|---|---|---|---|---|
| Intercept | 73.859 | 0.455 | 162.37 (15) | < 0.001 | (72.889, 74.829) | 0.65 |
| Ethnicity | | | | | | |
| Other Hispanic | 1.189 | 1.087 | 1.09 (15) | 0.291 | (-1.127, 3.505) | 1.22 |
| White | 1.781 | 0.631 | 2.82 (15) | 0.013 | (0.436, 3.125) | 1.31 |
| Black | 3.465 | 0.779 | 4.45 (15) | < 0.001 | (1.804, 5.126) | 1.16 |
| Other | 1.189 | 0.934 | 1.27 (15) | 0.223 | (-0.803, 3.180) | 1.20 |
| Mexican | -- | -- | -- | -- | -- | -- |
| Marital Status | | | | | | |
| Previously Married | 1.040 | 0.622 | 1.67 (15) | 0.115 | (-0.285, 2.366) | 1.73 |
| Never Married | -0.343 | 0.582 | -0.59 (15) | 0.564 | (-1.583, 0.897) | 1.43 |
| Married | -- | -- | -- | -- | -- | -- |
| Gender | | | | | | |
| Female | -2.721 | 0.338 | -8.06 (15) | < 0.001 | (-3.441, -2.002) | 1.16 |
| Male | -- | -- | -- | -- | -- | -- |
| Age (Centered) | 0.125 | 0.015 | 8.45 (15) | < 0.001 | (0.094, 0.157) | 1.59 |
| Age (Cent.) Squared | -0.012 | 0.001 | -16.34 (15) | < 0.001 | (-0.014, -0.011) | 1.77 |

# Fitting a Multiple Regression Model to 2005-06 NHANES Data on Diastolic BP: Analyze Residuals!

- Stata:

```
. predict ehat1, resid

. predict yhat1, xb

. scatter ehat1 yhat1, name(ehat1xyhat1, replace) title(Residuals v.
  Predicted Y)
```

- In SAS, generate plots using the **outdiag2** data set, which was saved to include the predicted values and residuals based on the fitted model

- Appropriate regression diagnostics for complex sample survey data (Cook's statistics, influence measures, collinearity diagnostics, etc.) is an active area of research in survey statistics, with several recent publications (Li and Valliant, Liao, etc.; see recent issues of *Survey Methodology*); methods have yet to make their way into the software, outside of a contributed R package that is currently under development (see Chapter 7 of ASDA, Second Edition)! Tools above are very simple and should be used to check model structure and outliers.

# Software Procedures for Logistic Regression Analysis of Complex Sample Survey Data (SAS)

- **SAS**: PROC SURVEYLOGISTIC
  - class xxxx; statement to declare categorical predictors
    - (/ param=ref) modifier for “dummy” / indicator coding
    - (ref=‘k”) to change reference category

  Ex: class gender (ref=‘2’) / param=ref;
  - model statement;
    - model mde (event=‘1’)= …;
    - Default is to use highest category as reference for the logit (be careful; default is to model probability of 0!)
  - Automatically generates Wald tests for multi-parameter predictors declared in **class;** statement.

# Software Procedures for Logistic Regression Analysis of Complex Sample Survey Data (Stata)

- **Stata**: the svy: logit and svy: logistic commands
  - svy: logistic defaults to odds ratio output; **coef** option yields logistic model parameter estimates
  - svy: logit defaults to log-odds (B) output; **or** option yields odds ratios
  - "i." prefix defines categorical predictors
    - Default to lowest alphanumeric category for reference
    - Change reference category for categorical predictors using ib#.
  - Post-estimation **test** statement for Wald tests of multi-parameter hypotheses
- Archer and Lemeshow (2006) developed a version of the Hosmer-Lemeshow goodness-of-fit test for logistic regression models fitted to complex samples; implemented in the

```
estat gof
```

post-estimation command!

# Example: Logistic regression analysis of the NCS-R data: Specifying the Preliminary Model

- In this example, we assess the significance of potential predictors of having lifetime major depression for adults greater than 17 years of age.

- **Dependent Variable**: (MDE: 1=Yes, 0=No)

- **Predictor Variables**:
  - AGE4CAT (1=18-29; 2=30-44;  3=45-49; 4=60+)
  - SEX (1 = Male; 2 = Female)
  - ALD (1 = alcohol dependent; 0= not ALD)
  - EDCAT (1=<12; 2=12; 3=13-15; 4=16+)
  - MARCAT (1=Married; 2=Previously; 3=Never)

# Example: Logistic regression analysis of the NCS-R data: Fit example model.

Stata code:

```
. svy: logistic mde i.ag4cat ib2.sex ald i.ed4cat i.mar3cat
. svy: logistic mde i.ag4cat ib2.sex ald i.ed4cat i.mar3cat, coef

. svy: logit mde i.ag4cat ib2.sex ald i.ed4cat i.mar3cat
```

Estimated odds ratios and 95% CIs can be generated in svy: logit by adding the or option:

```
. svy: logit mde i.ag4cat ib2.sex ald i.ed4cat i.mar3cat, or
```

Other software systems (e.g., SAS PROC SURVEYLOGISTIC) will output both the estimated logistic regression coefficients (and standard errors) and the corresponding odds ratio estimates.

# Example: Logistic regression analysis of the NCS-R data: Fit example model.

SAS Code:

```
proc surveylogistic data=ncsr ;
  strata sestrat ;
  cluster seclustr ;
  weight ncsrwtlg ;
  class ag4cat (ref=first) ed4cat (ref=first) mar3cat (ref=first) / param=ref  ;
  model mde (event='1')= ag4cat sexm ald ed4cat mar3cat;
run ;
```

SAS PROC SURVEYLOGISTIC will output both the estimated logistic regression coefficients (and standard errors) and the corresponding odds ratio estimates.

# Example: Logistic regression analysis of the NCS-R data: Estimated Coefficients in Example Model.

| Predictor* | Category | $\hat{B}$ | $se(\hat{B})$ | *t* | $P(t_{42} > t)$ |
|---|---|---|---|---|---|
| INTERCEPT | | -1.583 | 0.121 | -13.12 | <0.001 |
| AGE4CAT | 30-44 | 0.255 | 0.122 | 2.71 | 0.100 |
| | 45-59 | 0.206 | 0.113 | 2.26 | 0.029 |
| | 60+ | -0.676 | 0.072 | -4.78 | < 0.001 |
| SEX | Male | -0.577 | 0.138 | -7.48 | < 0.001 |
| ALD | Yes | 1.424 | 0.640 | 9.24 | < 0.001 |
| ED4CAT | 12 | 0.079 | 0.105 | 0.82 | 0.418 |
| | 13-15 | 0.231 | 0.117 | 2.48 | 0.017 |
| | 16+ | 0.163 | 0.130 | 1.47 | 0.148 |
| MAR3CAT | Previously | 0.486 | 0.139 | 5.69 | <0.001 |
| | Never | 0.116 | 0.121 | 1.07 | 0.290 |

* Source: NCS-R, $n = 5{,}692$, adjusted Wald test for all parameters: $F(10,33) = 28.07$, $p < 0.001$. Reference categories for categorical predictors are: AGE4CAT (18-29); SEX (Female); ALD (No); ED4CAT (<12 yrs); MAR3CAT (Married).

# Example: Logistic regression analysis of the NCS-R data: Interpret Model Results as Odds Ratio Estimates.

| Predictor* | Category | $\hat{\psi}$ | 95% CI for $\hat{\psi}$ |
|---|---|---|---|
| AGE4CAT | 30-44 | 1.29 | (1.067, 1.562) |
| | 45-59 | 1.23 | (1.022, 1.479) |
| | 60+ | 0.51 | (0.383, 0.677) |
| SEX | Male | 0.56 | (0.480, 0.656) |
| ALD | Yes | 4.15 | (3.042, 5.668) |
| ED4CAT | 12 | 1.08 | (0.890, 1.316) |
| | 13-15 | 1.26 | (1.044, 1.519) |
| | 16+ | 1.18 | (0.941, 1.471) |
| MAR3CAT | Previously | 1.63 | (1.369, 1.932) |
| | Never | 1.12 | (0.903, 1.396) |

# Example: Logistic regression analysis of NCS-R Wald Tests of Multi-parameter Predictors.

NOTE: These tests are automatically printed by SAS PROC SURVEYLOGISTIC, and Stata users need to run the `test` post-estimation command indicated earlier.

| **Categorical Predictor** | ***F*-test Statistic** | $P(\mathcal{F} > F)$ |
|---|---|---|
| AG4CAT | $F_{(3,40)} = 19.03$ | < 0.001 |
| ED4CAT | $F_{(3,40)} = 2.13$ | 0.112 |
| MAR3CAT | $F_{(2,40)} = 16.60$ | < 0.001 |

# References


- Hansen, M.H., Madow, W.G., and Tepping, B.J. (1983). An evaluation of model-dependent and probability sampling inferences in sample surveys. *Journal of the American Statistical Association*, 78, 776-793.
- Heeringa, S.G., West, B.T., and Berglund, P.A. (2017). *Applied Survey Data Analysis.* Chapman & Hall / CRC Press: Boca Raton, FL.
- Khera, R., Angraal, S., et al. (2017). Adherence to Methodological Standards in Research Using the National Inpatient Sample. *Journal of the American Medical Association*, 318(20), 2011-2018.
- Smith, T.W. (2011). Refining the Total Survey Error Perspective. *International Journal of Public Opinion Research*, 23, 464-484.
- Vallliant, R. and Dever, J. (2018). *Survey Weights: A Step-by-Step Guide to Calculation*. Stata Press: College Station, TX.
- West, B.T., Sakshaug, J.W. and Aurelien, G.A.S. (2016). How Big of a Problem is Analytic Error in Secondary Analyses of Survey Data? *PLOS One*, 11(6), e0158120.
- West, B.T., Sakshaug, J.W. and Kim, Y. (2017). Analytic Error as an Important Component of Total Survey Error: Results from a Meta-Analysis. In P. Biemer, E. de Leeuw, S. Eckman, B. Edwards, F. Kreuter, L.E. Lyberg, N.C. Tucker, and B. West (Eds.). *Total Survey Error in Practice*. Hoboken, NJ: John Wiley and Sons, pp. 489-508.
- West, B.T. and Sakshaug, J.W. (2018). The Need to Account for Complex Sampling Features when Analyzing Establishment Survey Data: An Illustration using the 2013 Business Research and Development Innovation Survey (BRDIS). *Survey Methods: Insights from the Field*. https://doi.org/10.13094/SMIF-2018-00001.

# Thank You!

- Please direct any comments or questions to **bwest@umich.edu**