

FEATURES OF COMPLEX SAMPLE SURVEY DESIGNS

Iris Shimizu

**Office of Research and Methodology
National Center for Health Statistics, CDC**

**National Conference on Health Statistics
August 16, 2010**

The views expressed here are those of the author and not necessarily those of the NCHS or the CDC.



OUTLINE

I. INTRODUCTION

II. SAMPLE DESIGN FEATURES

III. SURVEY WEIGHT FEATURES

IV. FEATURE EFFECTS ON ANALYSIS

V. SUMMARY

INTRODUCTION

- SAMPLE SURVEYS

- PROBABILITY (SCIENTIFIC) SAMPLE
 - NON ZERO CHANCE FOR EVERY UNIT

 - CHANCE CAN BE COMPUTED

SAMPLE DESIGN FEATURES

- **SIMPLE RANDOM SAMPLE**

- **STRATIFIED SAMPLES**

USED TO IMPROVE ESTIMATE PRECISION

- **CLUSTER SAMPLES**

USED TO SAVE DATA COLLECTION COSTS

SAMPLE DESIGN FEATURES (continued)

- **MULTI-STAGE SAMPLING**

 - SAMPLING PROBABILITY AT EACH STAGE**

- **UNEQUAL SAMPLING PROBABILITIES**

 - **POSSIBLE IN STRATIFIED SAMPLES**

 - **PROBABILITY PROPORTIONAL TO SIZE**

SURVEY WEIGHT FEATURES

- MEASURE OF UNITS REPRESENTED
- TYPICAL WEIGHT COMPONENTS
 - INVERSE OF SELECTION PROBABILITY
 - ADJUSTMENT OF NON-RESPONSE
 - MAY HAVE RATIO ADJUSTMENT
 - BENCHMARKING – USES “ACTUAL” TOTALS
 - CALIBRATION – USES CORRELATED DATA FROM NON-SAMPLE SOURCES FOR UNIVERSE

AFFECTS OF COMPLEX DESIGN FEATURES ON ANALYSIS

- **SAMPLES MAY NOT BE SELF WEIGHTING**

FAILURE TO USE PROVIDED SURVEY WEIGHTS
COULD DISTORT ESTIMATES OF:

- TOTALS
- RATES AND PROPORTIONS

VARIABILITY OF SURVEY ESTIMATES

- **SAMPLING VARIABILITY**
- **VALID VARIANCE ESTIMATES MUST
ACCOUNT FOR SAMPLE DESIGNS**

COMPLEX SAMPLE FEATURES AFFECTING VARIANCES

- **CLUSTERING OF ANALYTIC UNITS WITHIN
PRIMARY SAMPLING UNITS (PSUs)**
 - **STRATIFICATION OF PSUs**

DANGERS OF IGNORING SAMPLE DESIGN IN VARIANCE ESTIMATION

- **VARIANCE ESTIMATES PROBABLY TOO**

- **“DEGREES OF FREEDOM”**
SMALL TOO LARGE

DANGER OF USING SAMPLE SUBSETS TO ESTIMATE VARIANCES

- **ESTIMATED VARIANCE**
 - **MAY NOT REFLECT SAMPLE DESIGN**
 - **COULD UNDERSTATE SAMPLING
VARIANCE**

DATA FROM MULTIPLE SURVEYS

- **IF INDEPENDENT SAMPLES, THEN
USE SEPARATE STRATA**
- **REQUIRES COMPARABLE DATA ITEMS
ACROSS SURVEYS**

GENERAL REFERENCE FOR COMPLEX SURVEY ANALYSIS SOFTWARE

<http://www.hcp.med.harvard.edu/statistics/survey-soft/>

PROVIDES LIST OF SOFTWARE PACKAGES FOR
ANALYSIS OF COMPLEX SAMPLE DATA WITH
DESCRIPTIONS AND LINKS TO THOSE PACKAGES.

SUMMARY

- SAMPLES CAN HAVE VARIETY OF COMPLEX

- FOR ALL COMPLEX SAMPLE SURVEYS, DATA USERS
FEATURES.
SHOULD USE:
 - SURVEY WEIGHTS

 - COMPLEX SAMPLE SURVEY ANALYSIS SOFTWARE

 - WHOLE SAMPLE, NOT SUBSET

Contact information

Iris Shimizu

IShimizu@cdc.gov