Methods for Estimating Change from NSCAW I and NSCAW II

Paul Biemer

Sara Wheeless

Keith Smith

RTI INTERNATIONAL

RTI International is a trade name of Research Triangle Institute

National Survey of Child and Adolescent Well-Being

Course Outline

- Review of NSCAW I and NSCAW II Designs
- Issues in Estimating Between-Cohort Differences
- Calibration Weighting
- Cautions in Using Calibration Weights
- Illustrations and Examples
- Questions and Answers

Overview of the NSCAW I Design

- National representative stratified, two-stage sample
  - 100 PSUs (counties or groups of counties)
  - Secondaries are children with counties selected by a stratified sample using 8 strata (domains)
  - Age restricted to 0 to 14 years
- Four states (representing 8 PSUs) refused to participate (referred to as “agency first contact” states)

NSCAW II Design

NSCAW II was designed to facilitate comparisons with NSCAW I

- an equivalent target population (with exceptions to be noted),
- same PSUs were used as for NSCAW I
- statistically equivalent sampling methodology,
- same interview protocols, respondent selection rules, and nonresponse conversion mechanisms,
- similar questions or characteristics, and
- comparable weighting, post-survey weighting and estimation methods

National Survey of Child and Adolescent Well-Being

The NSCAW II Sample Design

Very similar to NSCAW I except:

- Age range expanded from 14 to 17.5
- 5 within PSU sampling domains instead of 8
- Four new agency first contact states (representing 9 additional PSUs)

→**Target populations for NSCAW I and II are not identical**

National Survey of Child and Adolescent Well-Being

NSCAW I and NSCAW II Target Populations

NSCAW I	NSCAW II
All children **age 0 – 14 years** who are subjects of investigations of child abuse/neglect during the 15-month time period from **October, 1999 through December, 2000**	All children **age 0 – 17.5 years** who are subjects of investigations of child abuse/neglect during the 15-month time period from **February, 2008 through May, 2009**.
Excludes children in **8** "agency first contact" PSUs representing about **5%** of the US child welfare population.	Excludes children in the NSCAW I **8** plus **9 additional** "agency first contact" representing about **10%** of the US child welfare population.

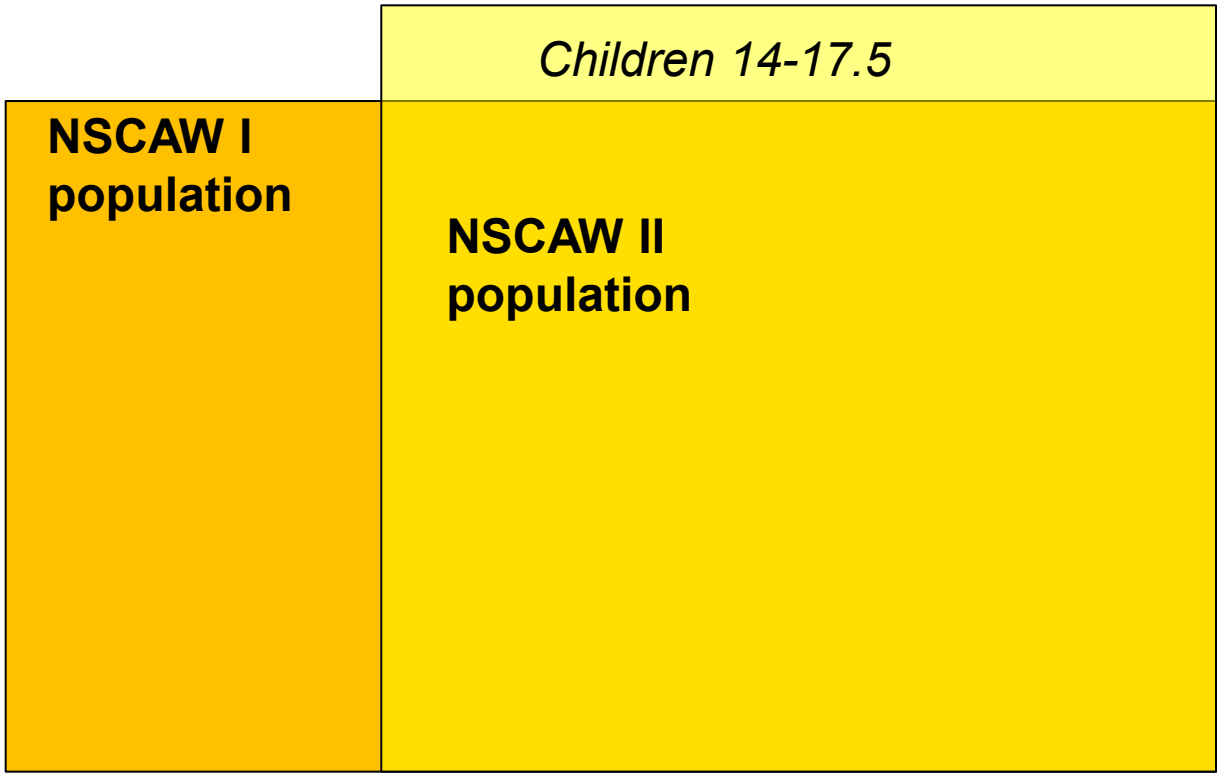
National Survey of Child and Adolescent Well-Being

Coverage Error

NSCAW I population

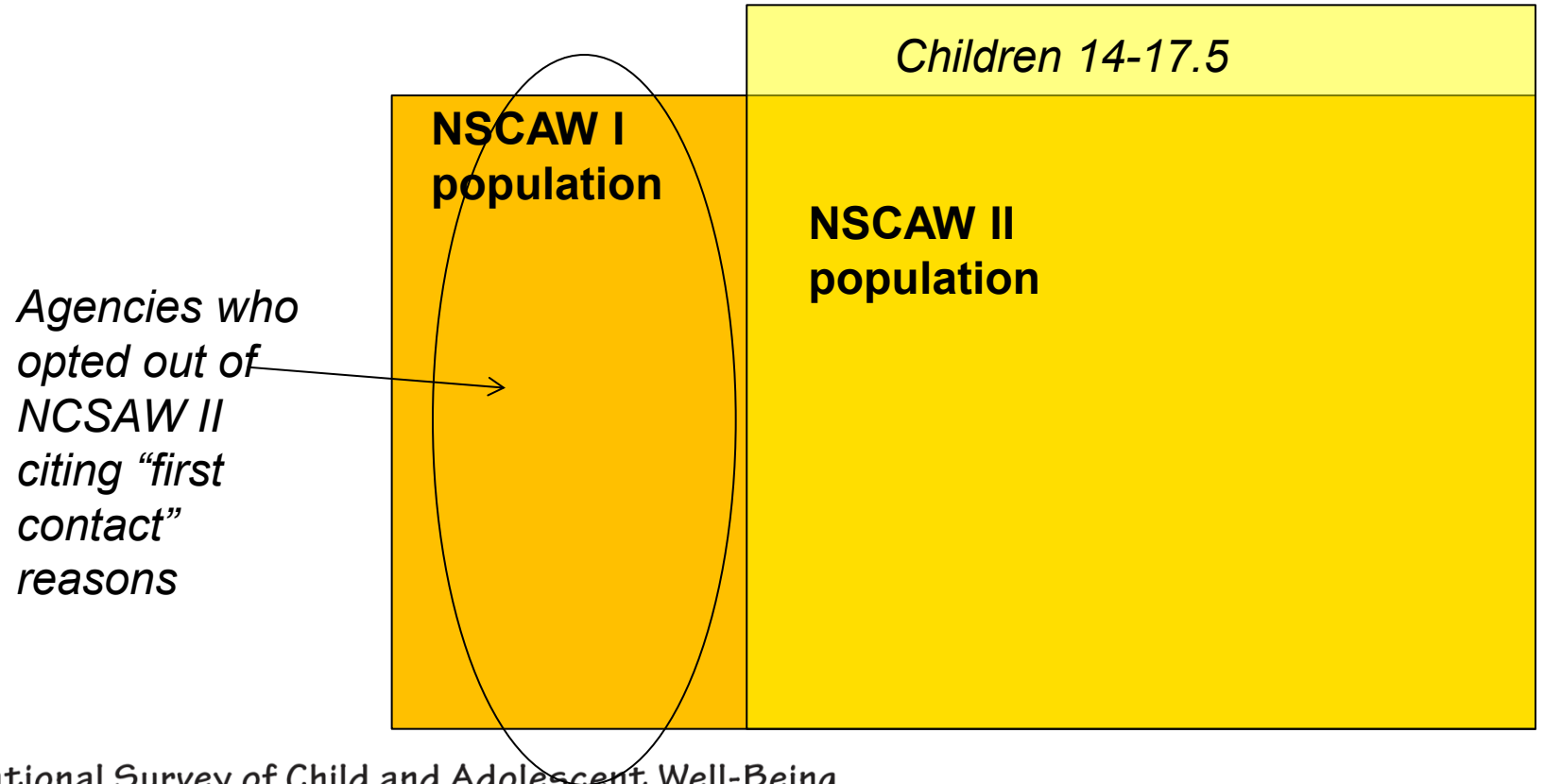
National Survey of Child and Adolescent Well-Being

Coverage Error



National Survey of Child and Adolescent Well-Being

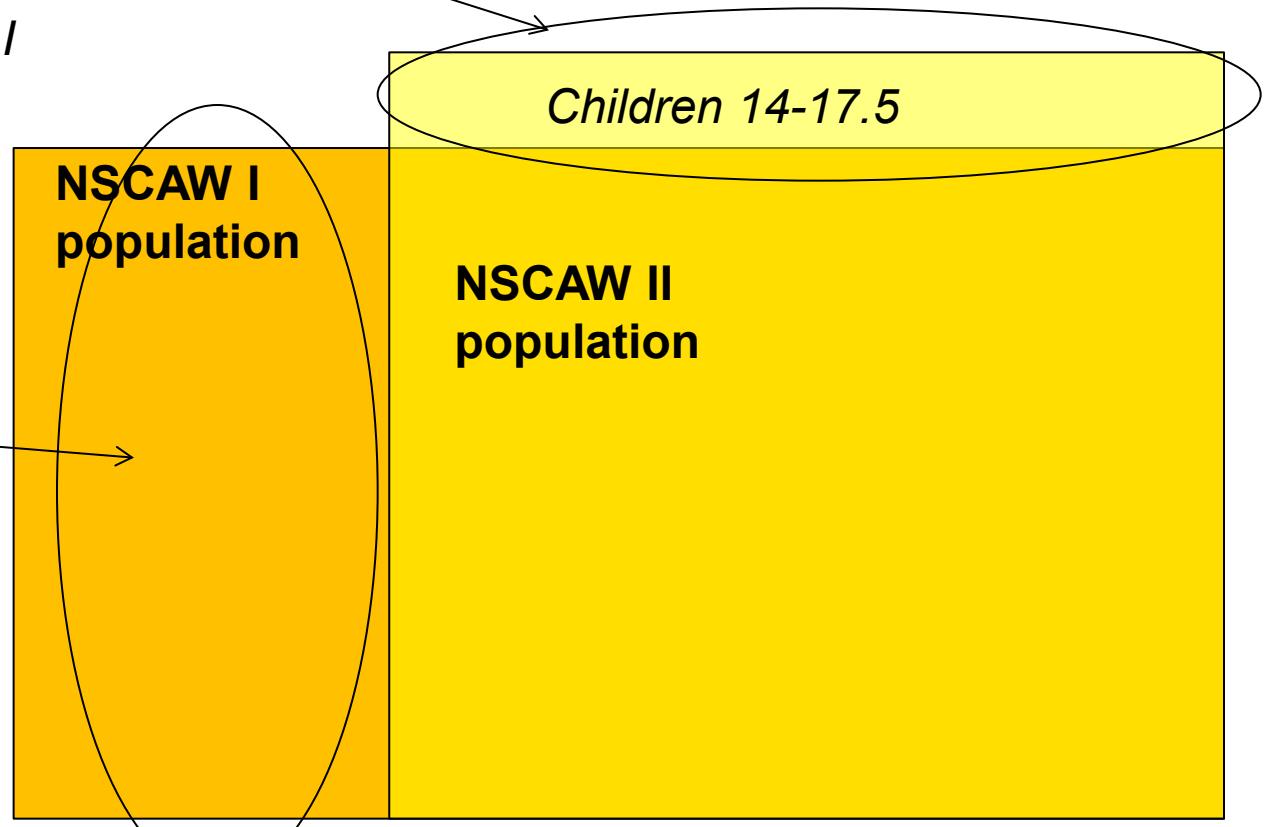
Coverage Error



National Survey of Child and Adolescent Well-Being

Coverage Error

Children age 14 to 17.5 who were not in-scope for NSCAW I



Agencies who opted out of NCSAW II citing "first contact" reasons

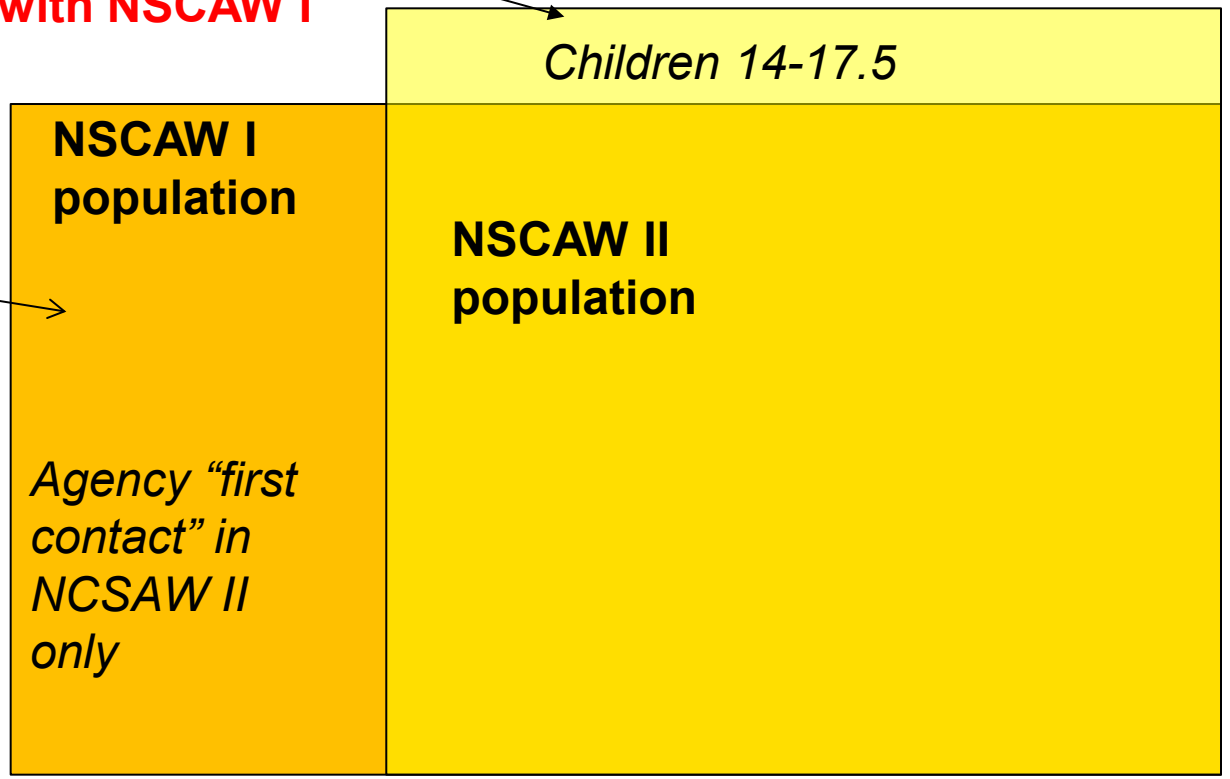
National Survey of Child and Adolescent Well-Being

Coverage Error and Remedies

Remove from NSCAW II for comparisons with NSCAW I

Either:

a. **Remove from NSCAW I for comparisons, or**
b. **Use calibration to adjust NSCAW II weights**



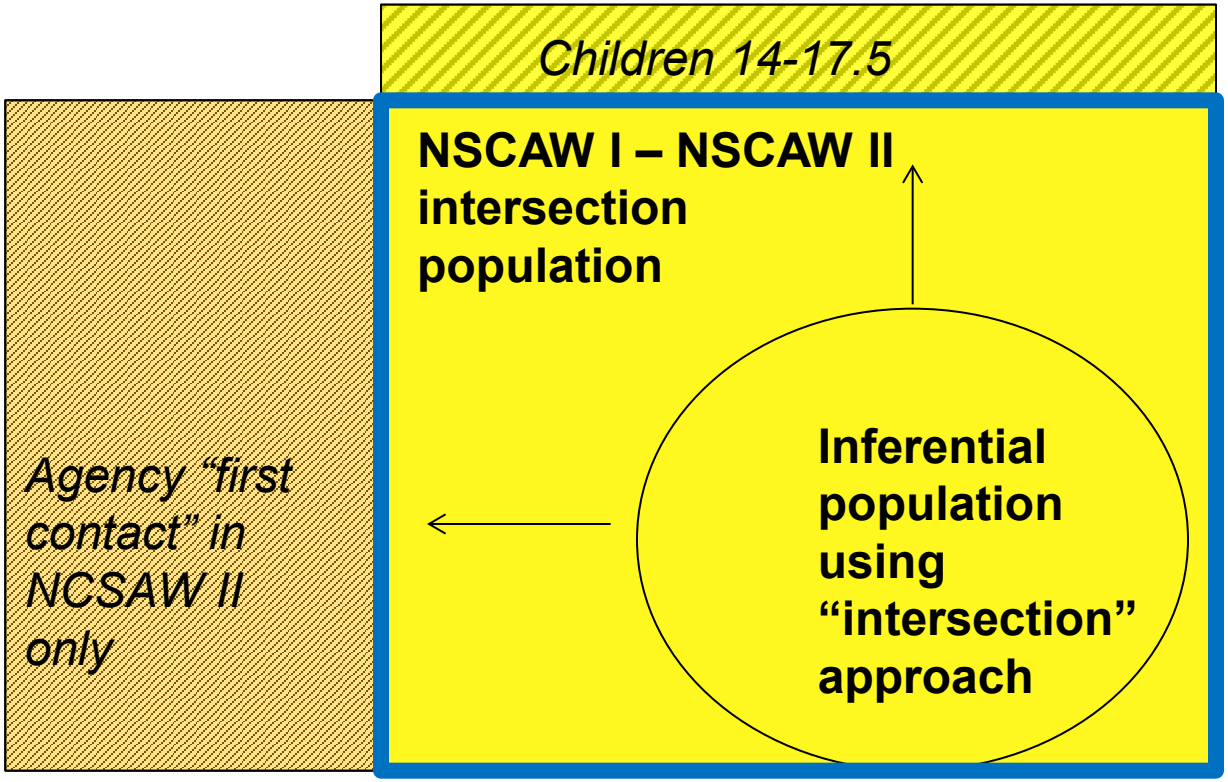
National Survey of Child and Adolescent Well-Being

Option A – Infer to the "Intersection" Population Only

- A simple approach is to analyze the "intersection" population only
- This approach requires that the population of inference for the comparison be restricted to the intersection of NSCAW I and NSCAW II populations; viz.,
  - Children 0-14
  - Areas represented by NSCAW II cooperating agencies

National Survey of Child and Adolescent Well-Being

Inferential Population Using "Intersection" Approach

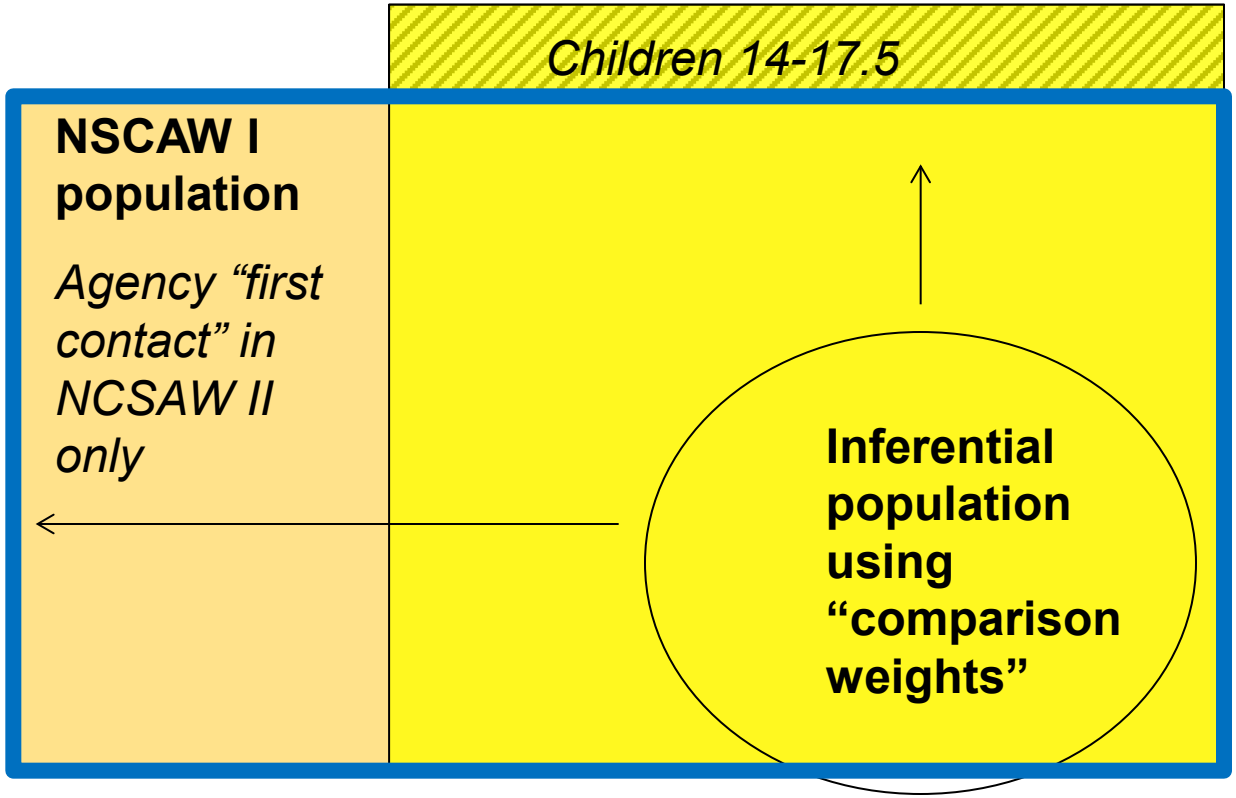


National Survey of Child and Adolescent Well-Being

Option B – Expand Inference to the Entire NSCAW Population Using "Calibration" Weights

- To expand the inferential population to the entire NSCAW I population, RTI developed the "calibration" weights.
- These weights still restricted the population of inference for the comparison to children 0-14
- However, it includes a sophisticated coverage adjustment that expands inference to include agencies that cooperated in NSCAW I but were first contact agencies in NSCAW II

Inferential Population Using "Calibration" Weights

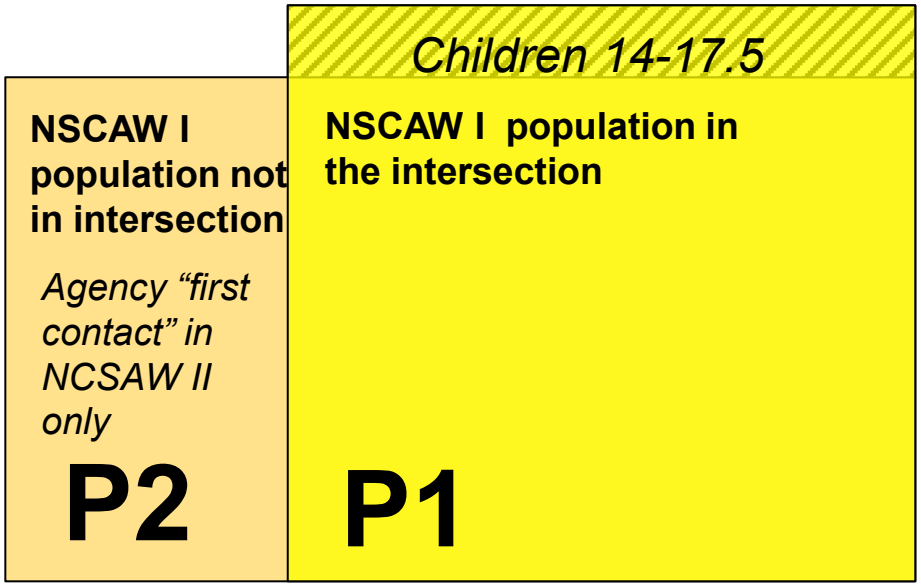


National Survey of Child and Adolescent Well-Being

The Calibration Weighting Process in a Nutshell

The NSCAW I population can be divided into two parts:

- P1 = intersection of the NSCAW I and NSCAW II populations
- P2 = subpopulation in NSCAW I not in NSCAW II (i.e., AFC states omitted from NSCAW II only)

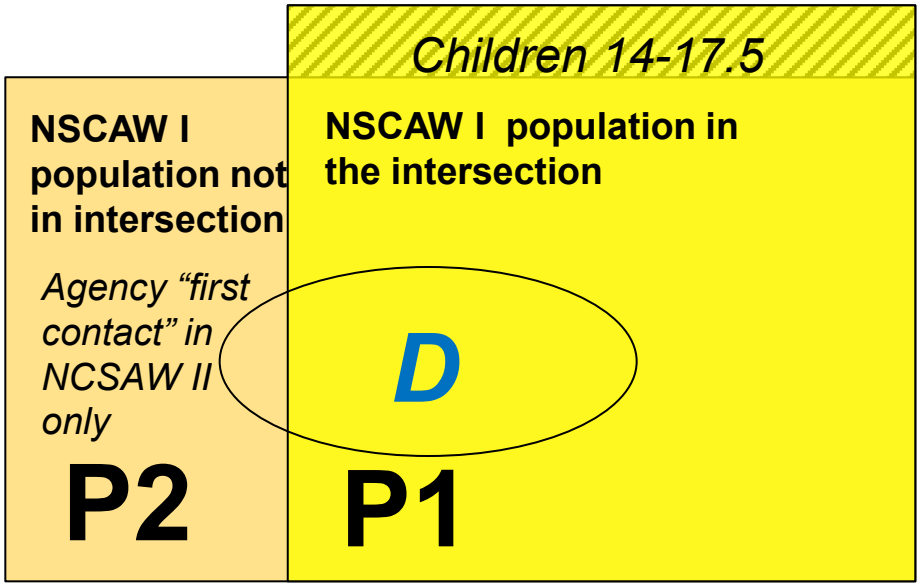


The Calibration Weighting Process in a Nutshell (cont'd)

Step 1. Identify a set of domains, *D*, to be used in the calibration.

These domains should be defined identically for both the NSCAW I and NSCAW II populations.

E.g., *D* includes the 5 NSCAW II domains x NSCAW I PSUs



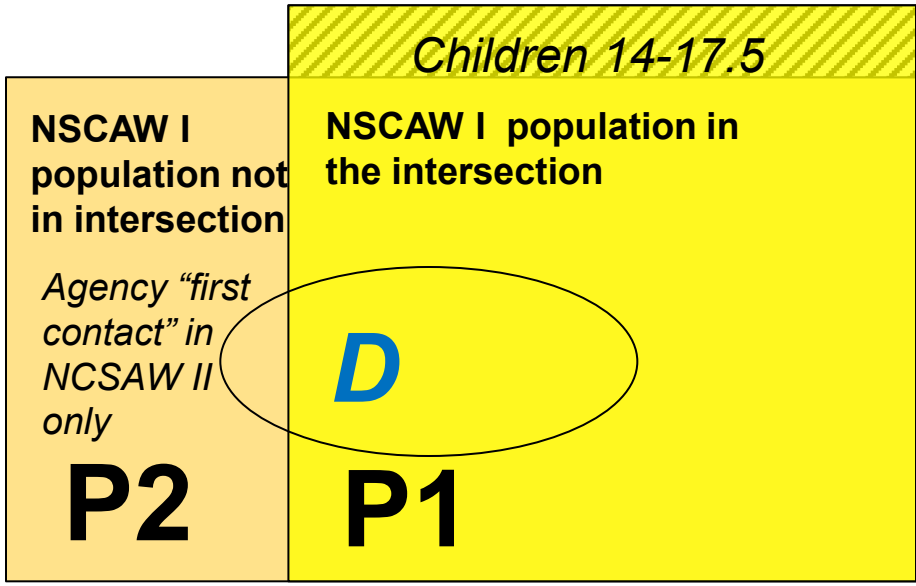
The Calibration Weighting Process in a Nutshell (cont'd)

Step 2. Compute weight adjustment factors, f_D, so that

$$\sum_{\text{all } D \cap P1} f_D w_D^I z_D = \sum_{\text{all } D \cap (P1 \cup P2)} w_D^I z_D$$

w_D^I = sum NSCAW I weight in domain D

z_D = sum of z in domain D



National Survey of Child and Adolescent Well-Being

The Calibration Weighting Process in a Nutshell (cont'd)

Step 3. Now, apply these adjustment factors to the weights corresponding to the same domains in NSCAW II.

$$f_D w_i^{II} \text{ for } i \in D$$

for the NSCAW II sample

$$w_i^{II} = \text{NSCAW II weight}$$

The Calibration Weighting Process in a Nutshell (cont'd)

Step 4. For the final step, these weights are further adjusted so that their totals add to 2006 NCANDS marginal control totals

$$\text{final NSCAW II weight} = a_i f_i w_i^{II}$$

- a_i: NCANDS adjustment weight
- f_i: Calibration adjustment weight
- w_i^{II}: Original NSCAW II weight

National Survey of Child and Adolescent Well-Being

Advantage of Intersection Approach over Calibration Weights

- Consistent with current NSCAW II weights
- NSCAW II target population is the population of inference (excluding 14+ aged children)
- Provide unbiased estimators of the NSCAW II population (excluding 14+ aged children)
- Easy to explain and understand

Disadvantage of Intersection Approach over Calibration Weights

- Limits inference to 90% of the child welfare population (compare to 95% using calibration weights)
- Sample sizes for comparisons are smaller (since part of NSCAW I sample must be discarded)
- Current NSCAW II weights have not been post-stratified to the intersection population – slightly reduces estimate stability

National Survey of Child and Adolescent Well-Being

Advantage of Calibration Weights over Intersection Method

- Expands inference to a larger population, viz., the NSCAW I population
- Uses all the NSCAW I data in estimating change, not just data in the intersection
- Residual coverage bias is small. Bias is 0 if we can assume that the adjustments that solve the calibration equations for NSCAW I would solve similar calibration equations for NSCAW II.

Disadvantage of Calibration Weights over Intersection Method

- Weights are 0 for children aged 14+ in NSCAW II – this may limit their utility for uses other than NSCAW I to NSCAW II comparisons
- Complicated to explain (but not necessarily to use)
- Using NSCAW II standard weights vs NSCAW II calibration weights may produce differences for the same NSCAW II estimates.

When to Use NSCAW II Calibration Weights

- Calibration weights apply only to NSCAW II and are wave specific.
  - So far, we have developed calibration weights for Waves 1 and 2 of NSCAW II
  - Wave 3 weights are currently being developed

When to Use NSCAW II Calibration Weights (cont'd)

- Calibration weights can be used for NSCAW II, Wave 1 analysis instead of the current NSCAW II Wave 1 weights.
  - The advantage is that inference can be expanded to the essentially the entire child welfare population (aged 0-14)
  - Disadvantage is that children aged 14+ must be dropped from the analysis when calibration weights are used

When to Use NSCAW II Calibration Weights (cont'd)

- When using Wave 1 calibration weights, NSCAW II estimates are directly comparable to corresponding estimates from NSCAW I with standard NSCAW I weights.
- Caution: Questions and scales being compared should be identical.
  - Question items: the same question text, response categories, and reference periods
  - Scales: same version of scale calculated the same way

National Survey of Child and Adolescent Well-Being

Illustration - Estimating the Change in Child CBCL for Some Age Group

Let

$\bar{y}_I$ denote the Child CBCL score for NSCAW I, Wave 1 weighted

$\bar{y}_{II}$ denote the Child CBCL score for NSCAW II, Wave 1 calibration weighted

Want to test H_0: $E(\bar{y}_I - \bar{y}_{II}) = 0$

Two sample t-test is a biased test since

$$Cov(\bar{y}_I, \bar{y}_{II}) \neq 0$$

National Survey of Child and Adolescent Well-Being

Proper Way to Proceed

Concatenate the NSCAW I and NSCAW II data sets.

Use SUDAAN to fit the ANACOV model

$$y_{ijk} = \beta_0 + \beta x_{ijk} + S_i + \varepsilon_{ijk}$$

Cohort

PSU

Person

Covariate

Cohort
I or II

$$E(\bar{y}_I - \bar{y}_{II}) = 0 \quad \Leftrightarrow \quad S_1 = S_2$$

Weight Variables

- **N1N2CWT1** – baseline (Wave 1) comparisons
- **N1N2CWT18MO** –18 month followup comparisons (using NSCAW I Wave 3 and NSCAW II Wave 2)
- **N1N2CWT36MO** – 36 month followup comparisons (using NSCAW I Wave 4 and NSCAW II Wave 3) (coming soon)

Variance Estimation
STRATUM and PSU Variables

- **COMP_STR** and **COMP_PSU** are variables that denote the variance estimation strata and PSU (or cluster)
- **COMP_PSU** is defined so that PSUs that are the same in the two surveys take the same values, so that the covariance is estimated correctly

Software

- Analysts should use the same software currently being used with NSCAW I or NSCAW II data
- Software needs to account for the stratification, clustering, and unequal weighting
- Examples of software: SUDAAN, SPSS Complex Samples, Stata with svyset, SAS survey procedures
- Documentation provided with the calibration weights gives some examples

National Survey of Child and Adolescent Well-Being

Analysis Techniques

- Concatenate/stack the data from the two surveys
- Need an indicator variable to denote the survey or cohort (**COHORT**: 1=NSCAW I, 2=NSCAW II)
- Analysis variables from the two surveys should be named identically
- Tell the software the name of the weight variable
- Tell the software the names of the variables for variance estimation (**COMP_STR** and **COMP_PSU**)

SUDAAN Code

```
proc regress design=wr;
 nest COMP_STR COMP_PSU;
 weight N1N2CWT1;
 model Y = X S
run;
```

COMP_STR, COMP_PSU = combined NSCAW I and NSCAW II Stratum and PSU indicator vectors,

N1N2CWT1 = concatenated NSCAW I and NSCAW II calibrated weight vector for wave 1 (baseline)

Y = $[y_{ijk}]$

X = $[x_{ijk}]$

S = cohort indicator variable.

National Survey of Child and Adolescent Well-Being

Other Examples

- Testing $H_0: E(\bar{y}_{II,d1} - \bar{y}_{II,d2}) = 0$ for two NSCAW II domains, d_1 and d_2, using calibration weights
- Testing $H_0: E(\bar{y}_{I,d1} - \bar{y}_{II,d1}) = 0$ for domain d_1 in NSCAW I and domain and d_1 in NSCAW II
- Others

Example Code

Examples of SAS/SUDAAN Code and Output

National Survey of Child and Adolescent Well-Being

```
libname cps 'C:\Documents and Settings\kesmith\My Documents\Data Delivery\final w5 data';
libname nscaw2 'C:\Documents and Settings\kesmith\My Documents\NSCAW II\Freq review';
libname calwgt "L:\Sampling\NSCAW_II\Calibration Weights\delivery";
option nofmterr;

/* Sort NSCAW I Survey data by child ID */
PROC SORT DATA= cps.cps OUT=NSCAW1_SURVEY (KEEP = NSCAWID SERVC YYB_TPT CRA13A
CHDGENDR);
BY NSCAWID;
RUN;

/* Sort NSCAW I Wave 1 Weights and Sample Variables by child ID */
PROC SORT DATA= calwgt.N1_W1Calib OUT=NSCAW1_Calib;
BY NSCAWID;
RUN;


/* Sort NSCAW II Survey data by child ID */
PROC SORT DATA=nscaw2.CPS_N2 OUT=NSCAW2_SURVEY (KEEP = NSCAWID SERVC YYB_TPT
CRA13A CHDGENDR);
BY NSCAWID;
RUN;

/* Sort NSCAW II Wave 1 Calibrated Weights and Sample Variables by child ID */
PROC SORT DATA=calwgt.N2_W1Calib OUT=NSCAW2_Calib;
BY NSCAWID;
RUN;


/* Next the comparison variables are merged with the survey variables, separately for each year */
/* Merge survey data with calibrated weights file */
DATA NSCAWI;
MERGE NSCAW1_SURVEY NSCAW1_Calib;
BY NSCAWID;
RUN;

DATA NSCAWII;
      MERGE NSCAW2_SURVEY NSCAW2_Calib;
      BY NSCAWID;
      RUN;

/* After the variables are merged on, the resulting NSCAW I and NSCAW II data sets should be
concatenated or stacked. */

/* Stack the datasets */
DATA COMPARISON;
SET NSCAWI NSCAWII ;

/* recode SERVC to 1-0 instead of 1-2 */
RSERVC=SERVC;
if SERVC=2 then RSERVC=0; else
if SERVC<0 then RSERVC=.;
/* Set negative values of CRA13A to missing */
```

```
if CRA13A < 0 then CRA13A=.;
run;

/* QC check on the SERVC recode */
proc freq; tables RSERVC*SERVC/list missprint;
run;

/* Once the datasets are stacked the data is ready for analysis using a survey software package such as SUDAAN.
To run the analysis, the stacked dataset must first be sorted by the variables on the NEST statement,
specifically COMP_PSU and COMP_STR. */

/* Begin SUDAAN Analysis */
PROC SORT DATA=COMPARISON;
BY COMP_STR COMP_PSU;
RUN;



/* EXAMPLE ANALYSIS OUTPUT IN SUDAAN */

/* Generate a crosstab and chi-square of the cohort by a categorical variable
Services at time of sampling */
PROC CROSSTAB DESIGN=WR DATA=COMPARISON;
WEIGHT N1N2CWT1;
NEST COMP_STR COMP_PSU;
class
NCOHORT
RSERVC
/nofreq;
tables
NCOHORT*RSERVC
;
test chisq;
print nsum wsum totper setot rowper serow colper secol / stest=default;
run;

/* Generate another crosstab and chi-square of the cohort by categorical variable
CW Risk Assessment (CRA13A) Active alcohol abuse by primary CG */
PROC CROSSTAB DESIGN=WR DATA=COMPARISON;
WEIGHT N1N2CWT1;
NEST COMP_STR COMP_PSU;
class
NCOHORT
CRA13A
/nofreq;
tables
NCOHORT*CRA13A
;
test chisq;
print nsum wsum totper setot rowper serow colper secol / stest=default;
run;


/* Test whether the estimated mean of a variable (for example SERVC) is the same in both survey years.
```

```
The PROC DESCRIPT procedure is used to conduct this test. */
PROC DESCRIPT DESIGN=WR DATA=COMPARISON;
WEIGHT N1N2CWT1;
NEST COMP_STR COMP_PSU;
VAR RSERVC;
CLASS NCOHORT;
DIFFVAR NCOHORT = (1 2);
rtitle "Difference in proportion receiving services (according to frame) in NSCAW I versus NSCAW II";
run;


/* Also, use SUDAAN to obtain the estimated mean of a variable (for example SERVC) for each cohort. */
PROC DESCRIPT DESIGN=WR DATA=COMPARISON;
WEIGHT N1N2CWT1;
NEST COMP_STR COMP_PSU;
VAR RSERVC;
CLASS NCOHORT;
Tables NCOHORT;
rtitle "Proportion receiving services (according to frame) in NSCAW I and NSCAW II";
RUN;



/* Child CBCL Comparison */
PROC DESCRIPT DESIGN=WR DATA=COMPARISON;
WEIGHT N1N2CWT1;
NEST COMP_STR COMP_PSU;
VAR YYB_TPT;
CLASS NCOHORT;
DIFFVAR NCOHORT = (1 2);
rtitle "Difference in the Child CBCL in NSCAW I and NSCAW II";
run;

PROC DESCRIPT DESIGN=WR DATA=COMPARISON;
WEIGHT N1N2CWT1;
NEST COMP_STR COMP_PSU;
VAR YYB_TPT;
CLASS NCOHORT;
Tables NCOHORT;
rtitle "Comparison of the Child CBCL in NSCAW I and NSCAW II";
RUN;

proc regress data=COMPARISON;
NEST COMP_STR COMP_PSU;
WEIGHT N1N2CWT1;
MODEL YYB_TPT = CHDGENDR NCOHORT;
SUBGROUP CHDGENDR NCOHORT;
LEVELS     2       2 ;
rtitle "Regression model: CBCL = Gender Cohort";
run;
```

```
libname nscaw2 'C:\Documents and Settings\kesmith\My Documents\NSCAW II\Freq review';
libname cps 'C:\Documents and Settings\kesmith\My Documents\Data Delivery\final w5 data';
libname library "L:\NSCAW_II\Master\Data";
libname calwgt18 'L:\Sampling\NSCAW_II\Calibration Weights\18months';

option nofmterr;


/* Sort NSCAW I Wave 3 Survey data by Child ID*/
PROC SORT DATA= cps.cps_w3 OUT=NSCAW1_SURVEY_W3 (KEEP = NSCAWID YB3_TPT);
BY NSCAWID;
RUN;

/* Sort NSCAW I Wave 3 Weights by Child ID */
PROC SORT DATA= calwgt18.N1_compwts OUT=NSCAW1_Calib_W3;
BY NSCAWID;
RUN;

/* Sort NSCAW II Wave 2 Survey data by Child ID*/
PROC SORT DATA= nscaw2.cps_n2_w2 OUT=NSCAW2_SURVEY_W2 (keep = NSCAWID YB2_TPT);
BY NSCAWID;
RUN;

/* Sort NSCAW II Wave 2 Calibrated Weights and Sample Variables by Child ID */
PROC SORT DATA=calwgt18.N2_compwts OUT=NSCAW2_Calib_W2;
BY NSCAWID;
RUN;


/* Next the comparison variables are merged with the survey variables, separately for each year */
/* Merge survey data with calibrated weights file */
DATA NSCAWI_W3;
MERGE NSCAW1_SURVEY_W3 NSCAW1_Calib_W3;
BY NSCAWID;
RUN;

DATA NSCAWII_W2;
MERGE NSCAW2_SURVEY_W2 NSCAW2_Calib_W2;
BY NSCAWID;
RUN;


/* Added Step for Making 18-Month Comparisons */

/* Prior to stacking the datasets,
rename NSCAW I W3 variables so they have the same name
as the NSCAW II W2 variables */
DATA NSCAWI_W3 (RENAME = (YB3_TPT = YB2_TPT)); SET NSCAWI_W3; RUN;


/* Stack the datasets */
DATA COMPARISON_W3_W2;
SET NSCAWI_W3  NSCAWII_W2 ;
```

```
run;


/* Once the datasets are stacked the data is ready for analysis using a survey software package such as SUDAAN.
To use SUDAAN, the stacked dataset must first be sorted by the variables on the NEST statement,
specifically COMP_PSU and COMP_STR. */

PROC SORT DATA=COMPARISON_W3_W2;
BY COMP_STR COMP_PSU;
RUN;


/* SUDAAN ANALYSIS */

/* Child CBCL Comparison */
PROC DESCRIPT DESIGN=WR DATA=COMPARISON_W3_W2;
WEIGHT N1N2Cwt18mo;
NEST COMP_STR COMP_PSU;
VAR YB2_TPT;
CLASS NCOHORT;
DIFFVAR NCOHORT = (1 2);
rtitle "Comparison of the Child CBCL in NSCAW I and NSCAW II";
run;

PROC DESCRIPT DESIGN=WR DATA=COMPARISON_W3_W2;
WEIGHT N1N2Cwt18mo;
NEST COMP_STR COMP_PSU;
VAR YB2_TPT;
CLASS NCOHORT;
Tables NCOHORT;
rtitle "Comparison of the Child CBCL in NSCAW I and NSCAW II";
RUN;
```