Non-Probability Sampling

ICF's Experience

September 25, 2017

R. Lee Harding, Statistician

Challenges with Probability Samples

- **National data have limited usefulness for estimating local needs and evaluating local programs**
  - Lack the sufficient sample size to produce “local” estimates
- **Generally not designed to address topics that are specific to subpopulation or communities**
- **Very few surveys conducted at the community level**
  - Behavioral Risk Factor Surveillance (BRFSS) surveys
- **Probability samples are experiencing lower response rates**
- **Probability samples are costly**
- **Suffer from issues related to timeliness**

Challenges with Non-Probability Sampling (NPS)

- **There is no statistical theory to support non-probability sampling**
- **Panel population is not representative of the population as a whole**
- **Some limitations within small geographic areas**
  - E.g. How many Hispanic Females 18-24 are actually on the panel in Prince George Virginia
- **The sample is often balance across some dimensions using the quota sample but this can distorts the other demographic dimensions**
- **The quality of the NPS is assessed by comparisons to traditional probability survey results**

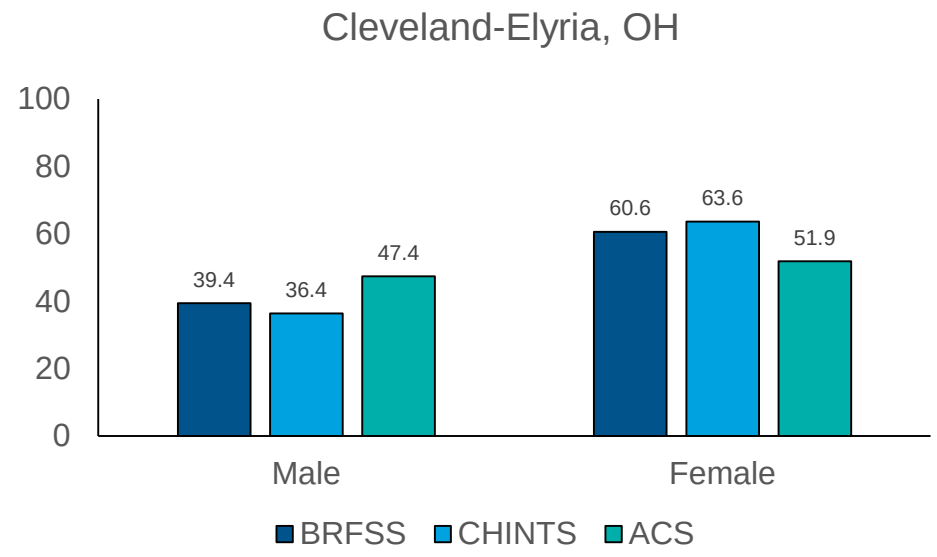
The Big Question Around Non-Probability Samples

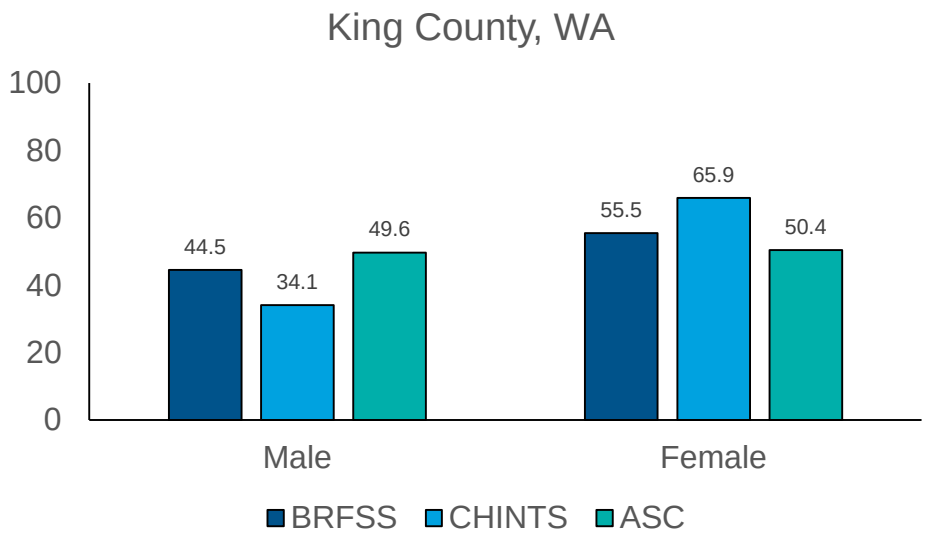
- **In the absence of a statistical theory supporting non-probability sampling, is there a method or reasonable decision rule that allows a non-probability samples to stand alone?**

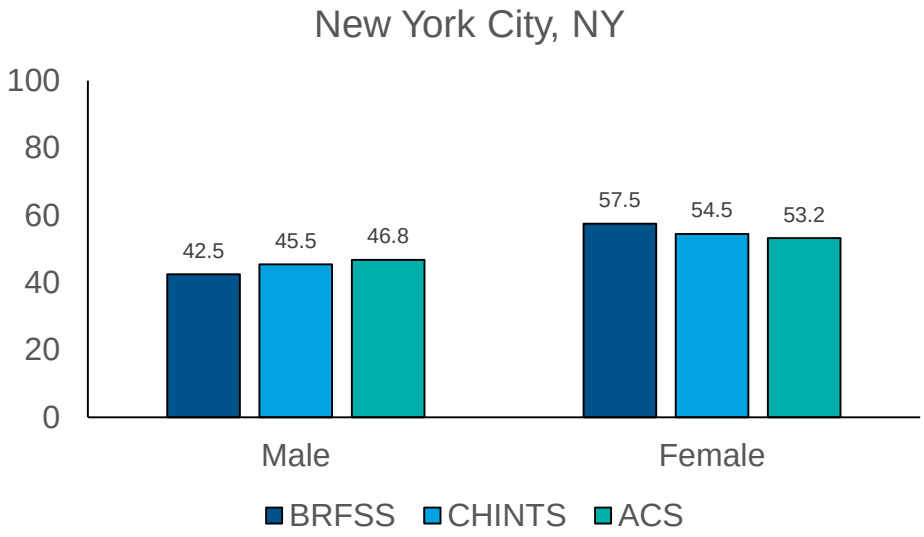
ICF's Experience with Non-Probability Samples

- **ICF initially piloted three NPS Community Health Information National Trends Survey (CHINTS)**
  - We modeled these pilots on the Health Information National Trends Surveys (HINTS)
- **The Los Angeles Health Interview Survey (LA HIS)**
  - Similar to CHINTS we additional health questions from the NHIS Early Reporting Measures as well as BRFSS questions
  - Based on the CHINTS experience we implemented two sampling approaches
    - **Arm 1**: The same methods used in the other three sites: follow ups to induce census balancing
    - **Arm 2**: Stratified random sample followed with a consistent reminder protocol (Enhanced Method)

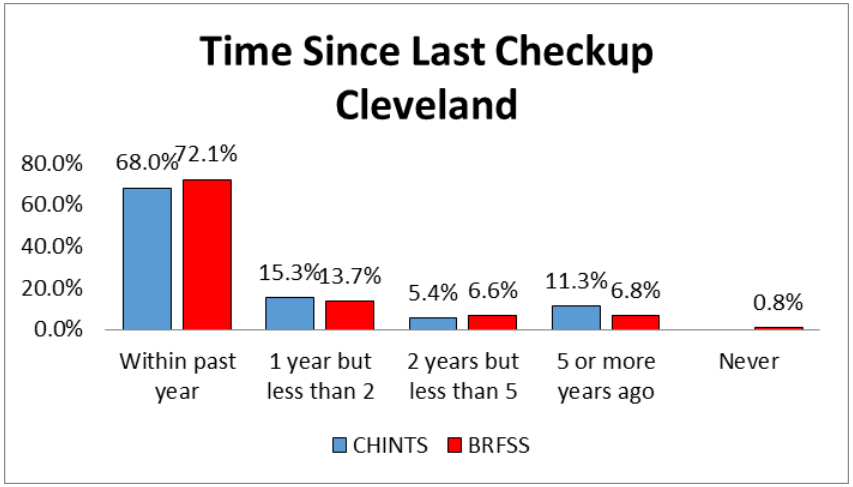
The CHINTS Pilot:
Unweighted frequencies – Gender

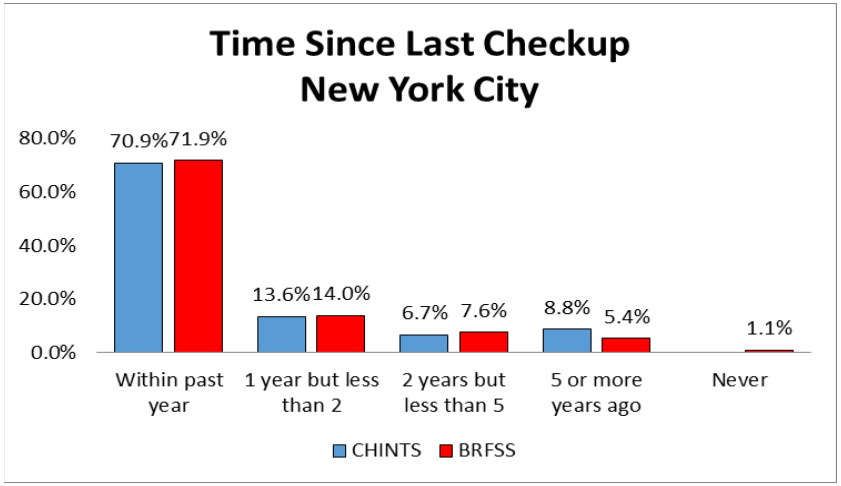


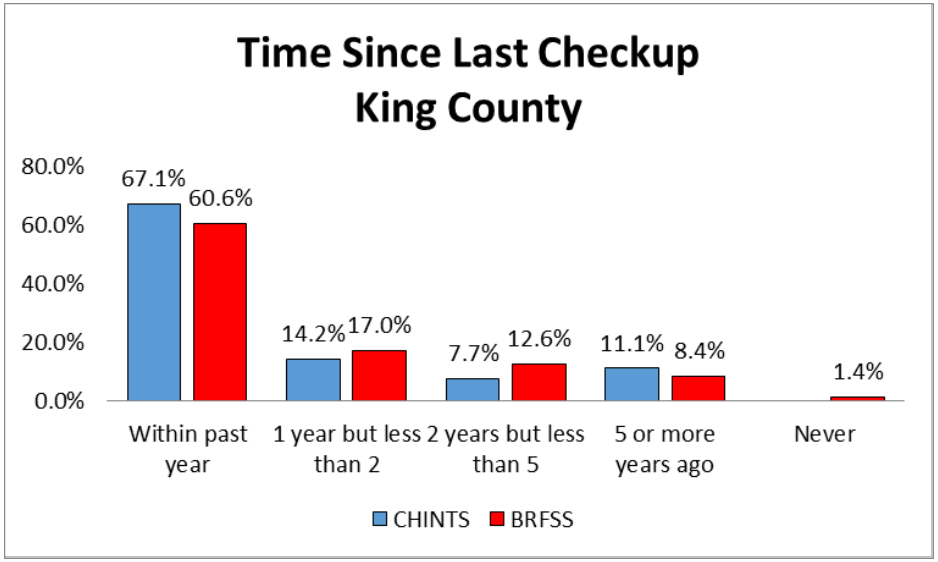




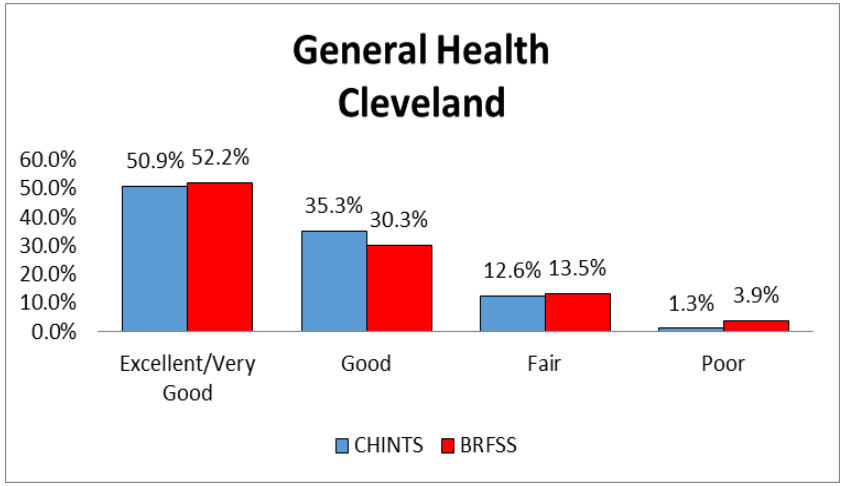
Length of Time Since Last Routine Checkup

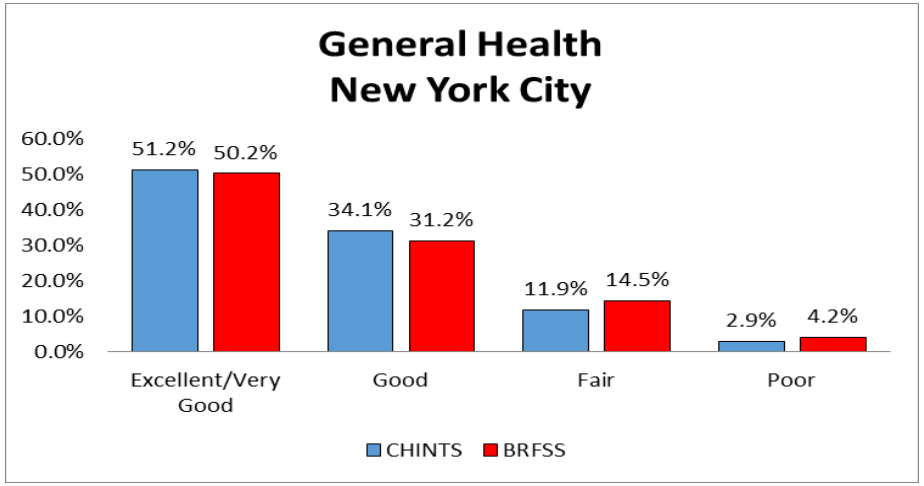


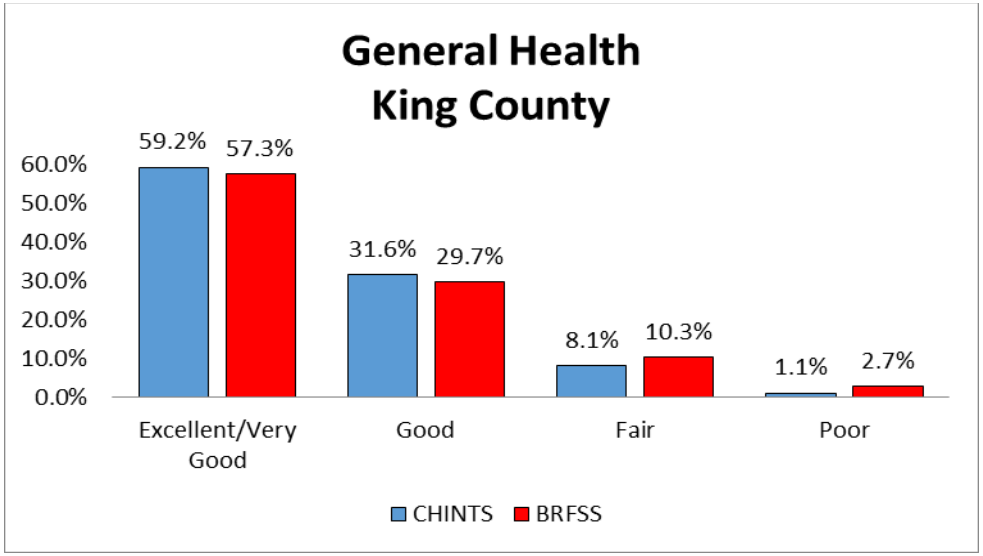




General Health





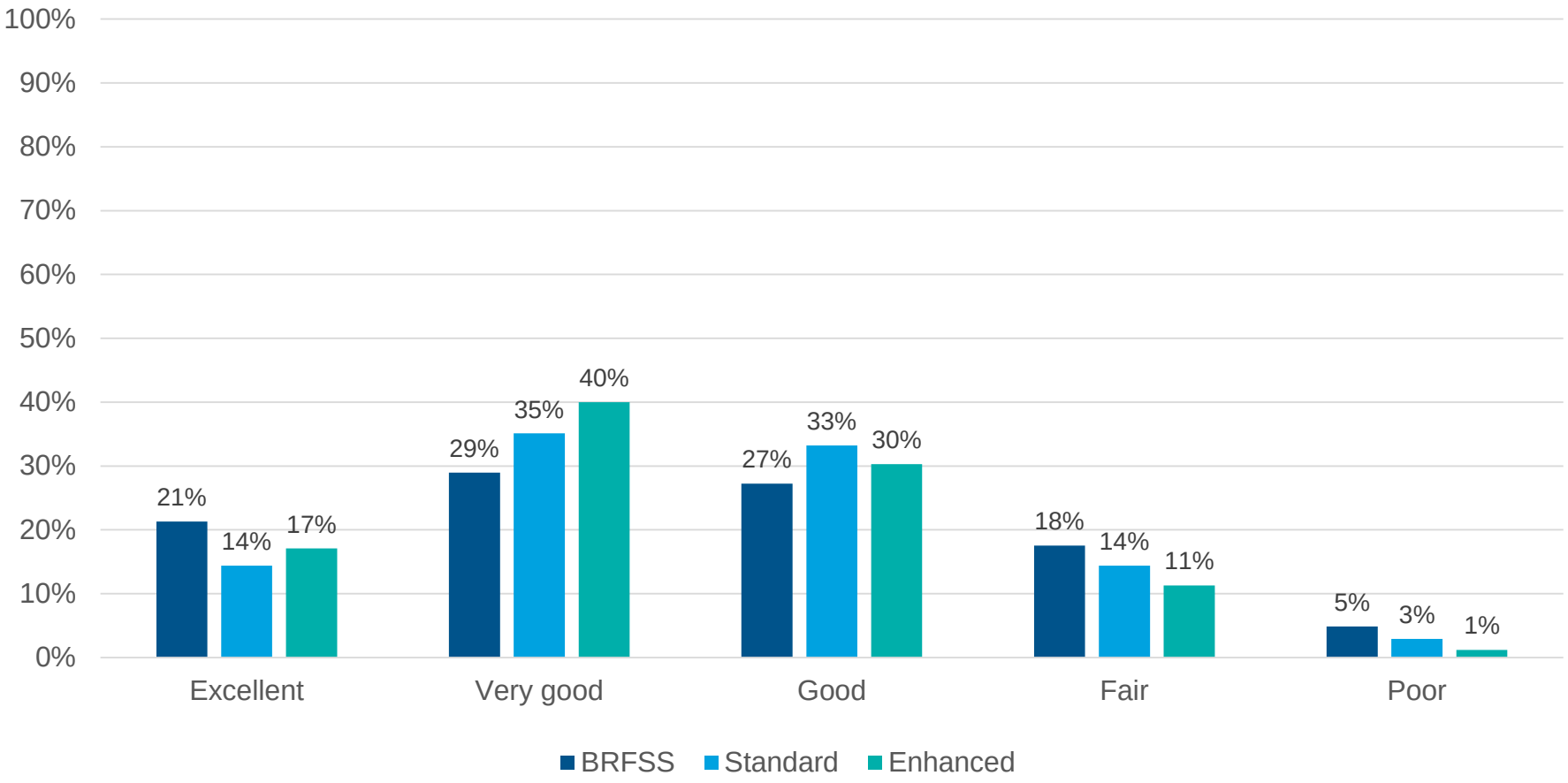


CHINTS Pilots: A few conclusions

- **CHINTS and BRFSS estimates are remarkably close in general**
- **Weighting for both surveys removed almost all potential biases**
- **A few differences remain for outcomes such as smoking**

LA HIS Pilot: Comparing Across Samples
General Health Rating

Would you say that in general your health was …. ?



LA HIS Pilot: Variances

- **We found that the variances due to unequal weighting effects are larger for the enhanced method which does not balance along the way.**
  - The standard protocol adjusted distribution along the sampling to conform to population so weight adjustments did not need to be large and variable

ICF's Experience with Non-Probability Samples (Continued)

- **National Immunization Survey (NIS)**
  - Immunization rates monitored with National Immunization Survey (NIS)
    - samples and screens households
    - conducts household interviews
    - collects medical records on immunization from providers
  - NIS Challenges
    - Low incidence population+ lack of appropriate frame -> Large sample size required
    - Expensive and time consuming to conduct
    - Low response rates--an increasing problem in public health surveillance
- **Childhood Immunization Mobile Pilot Survey (CHIMPS) - Exploring possible solutions for NIS challenges**

Childhood Immunization Mobile Pilot Survey (CHIMPS): Exploring possible solutions for NIS challenges

- **The approach involves both a mobile web survey and panel sample methodology**
- **CHIMPS questionnaire is similar to the NIS**
- **Benefits of the CHIMPS methodology:**
  - Timeliness
  - Flexibility
  - Cost-effectiveness
- **Two Weighting Methodologies**
  - Typical Poststratification
  - Propensity Score Matching

CHIMPS Weighting: Overview of Propensity Matching Methodology

- **Concatenate NIS and CHIMPS datasets**
- **Assign weights equal to 1 for CHIMPS records**
- **Build weighted logistic model**
  - **Dependent variable: y = 1 for those records from CHIMPS, else y = 0**
  - **Predictors: respondent's gender, maternal marital status, household income categories, maternal age group, maternal education level and rent/own home status**
- **Output the propensity scores, then use the inverse of propensities as new weights**

Assessment of the two methods: variation in the weights

Variable	Minimum	Mean	Median	Maximum	CV
Weights - Propensity	19,98.12	17,646.3	12,958.74	83,960.15	78.64
Weights - Poststratification	18,229.18	20,994.69	19,834.61	35,084.55	22.37

CHIMPS Pilot: Conclusions

- **Poststratification weighting method**
  - **Pros: lower variations and less limitation with size of datasets**
  - **Cons: may not have a good estimates to match with NIS**
- **Propensity Matching weighting method**
  - **Pros: give better estimations which are closer to NIS's outputs**
  - **Cons: higher variation due to small amount of observations (272 cases)**

The Big Question Around Non-Probability Samples

- **In the absence of a statistical theory supporting non-probability sampling, is there a method or reasonable decision rule that allows a non-probability samples to stand alone?**

An Empirical Method to Establish Usability of Nonprobability Surveys for Inference

Non-Probability Samples

- **Not Inferential - Accepted in market research, several academic disciplines but no accepted statistical theory**
- **Fast (500 interviews, nationwide, with parents in households with 19 – 35 month old children in 24 hours, 200 interviews in NYC for correlational study in 12 hours)**
- **Low cost, relatively, even when paying an incentive**
- **Hard to reach to survey (19 – 35 month children)**

An Empirical Method to Establish Usability of Nonprobability Surveys for Inference

- **This is a proposed method to push beyond just comparing NPS to PS and to allow for use of NPS for inference, i.e., in manner of a PS**
- **Motivated by risk tolerance, as in design based surveys, where we design a survey and select a sample with the risk α (generally = 0.05) of getting a bad sample, that is, in 1 out of 20 surveys, using predefined (a priori) decision rule**

An Empirical Method to Establish Usability of Nonprobability Surveys for Inference

Assumptions

- **The NPS is from a panel "quota sample" (NOT a river sample, or other convenience sample),**
  - The sample design that is repeatable
- **A successful comparison to PS on the first occasion the NPS stands alone at later times if**
  - 1. Panel demos only change marginally (user decides acceptable level of change)
  - 2. The same quota sample design is used

Empirical Method Rules

The organization that is responsible for making these estimates, selects the level of risk they are willing to accept by deciding on what to compare

1. Make overall population estimates, PE, or
2. Make sub-population estimates, SPE, or
3. Conduct multivariate analysis, MA
4. Include post stratification adjustment, PSW

If the organization

I. Only want **overall estimates** then a rule using comparisons at the overall level and defined a priori.

II. wants **overall estimates** and **sub-population estimates** then a rule covering overall comparisons and sub-population comparisons and defined a priori.

III. wants **overall estimates**, **sub-population estimates** and **multivariate relationships** then a rule covering overall estimate comparisons, sub-population comparisons and “correlational” comparisons and defined a priori.

IV. Considers the overall impact of adjusting – how much

Empirical Method Rules (Continued)

Rules are developed in the form of indices I_k, k = PE,SPE, MA and PSW

- I_k is calculated based on comparisons where a “good” comparison results in a 0 added to the index and a “bad” comparison results in some positive number added to the index.
- Since the rule is defined a priori the organization knows in advance the maximum possible “bad” score, say I_{MAX} and can assign the level of risk at some cutoff, say I_C , where if $I_k <= I_C$ the NPS is acceptable for inference.
- The organization is free to decide on the risk that is acceptable, if I_C near 0 then the organization is not willing to tolerate much risk and when I_C nears I_{MAX} the organization is wiling to tolerate more risk.
- Determining level of risk may include factoring in mode differences, timing, etc. This may increase the level of risk willing o tolerate

Empirical Method Rules (Continued)

Decision Rules

- Points assign as individual comparisons within the predefined rule(s)
- Create index(s) and every time a comparison fails add to the index. If the index score is over a predefined acceptable level of risk the comparison of the NPS to the PS is not successful
- Assume data user chooses rules based on: comparing ever asthma, ever diabetes, ever cancer, ever smoker, current smoker, excellent/very good health, flu shot last year and visited doctor in past year
  - Overall Comparison, 95% confidence intervals (Stephan and McCarty (1958), Sudman (1966)) adding 1 for each unsuccessful comparison
  - Comparison by gender, 95% confidence intervals adding 1 for each unsuccessful comparison
  - Ratio of CV of poststratification weights, if ≤ 1.2, 0 added to index, if ≥ 1.21 added 1 to index
- Max score for index is 25 if add 1 for each failed comparisons, user decides a priori cut off - if I_C > k NPS not acceptable

Overall Comparisons to a Probability Sample

Overall Estimates	PS			NPS Census			NPS Quota	
	Estimate	95% C. I.		Estimate	95% C. I.		Estimate	95% C. I.
Asthma Ever	0.14	0.13		0.12	0.08		0.15	0.10
		0.14			0.16			0.19
Cancer Ever	0.12	0.12		0.07	0.04		0.09	0.07
		0.13			0.10			0.12
Diabetes Ever	0.13	0.12		0.12	0.06		0.10	0.06
		0.13			0.17			0.13
Excellent/Very Good Health	0.51	0.50		0.52	0.41		0.57	0.52
		0.51			0.64			0.63
Flu Shot Last Year	0.29	0.29		0.41	0.32	v	0.44	0.38
		0.30			0.50			0.50
Smoke Ever	0.36	0.36		0.32	0.22		0.34	0.29
		0.37			0.42			0.38
Smoking Now	0.09	0.09		0.10	0.03		0.07	0.03
		0.09			0.16			0.10
Checkup Last 12 Months	0.67	0.66		0.64	0.47		0.69	0.63
		0.67			0.80			0.74

Overall Comparisons to a Probability Sample

Sub-population estimates by gender: Census NPS and Quota NPS both have total score of 4 out of 16.

Census NPS	**Quota NPS**
Male Flu Shots	Male Flu Shots
Female Flu shoots	Female Flu shoots
Male ever cancer	Male ever cancer
Male smoker ever	Female ever diabetes

Ratio of cv of post-stratification weights

Census NPS/PS = 0.03

So we add 0 to index

Quota NPS/PS = 2.54

So we add 1 to index

An Empirical Method to Establish Usability of Nonprobability Surveys for Inference

Index score for Quota NPS and Census NPS is 6 (1 + 4 + 1) and (2 + 4 + 0), respectively

1. For later occasions compare panel demos from time 1 based on a priori decision rule
2. If not substantial change, again user determined, no need to have a comparison PS, conduct NPS using same quota sample design – data is acceptable for use
3. For even later use repeat 1 and 2.
4. When panel demos change too much repeat NPS and PS comparison.

Conclusions

- **Our Comparisons of NPS estimates to probability based estimates have been comparable**
- **Questions around weighting and variance estimation in NPS**
- **Developing Rules to use NPS without a comparison probability study**

Thank You

R. Lee Harding

Richard.l.harding@icf.com