

Evaluation of Strategies to Improve the Utility of Estimates from a Non-Probability Based Survey

Presenter: Benmei Liu, National Cancer Institute, NIH
liub2@mail.nih.gov

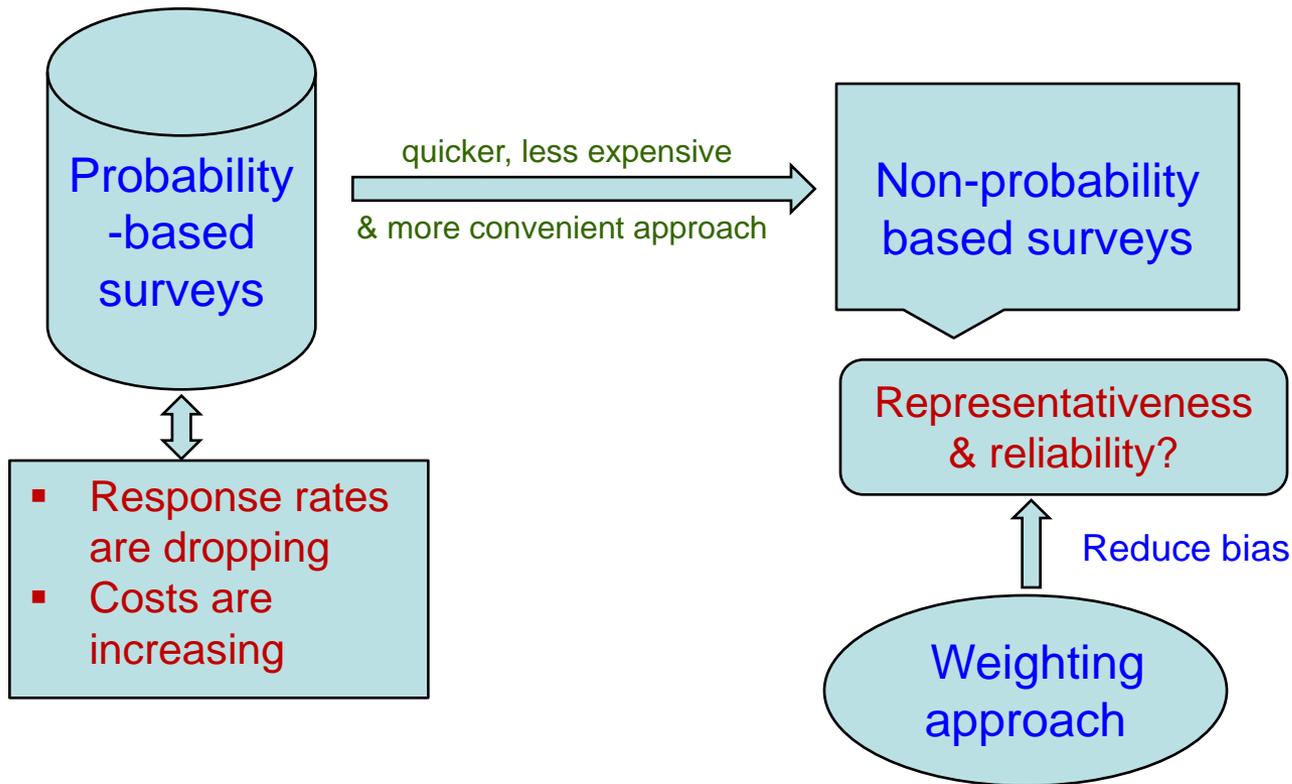
Authors: (Next Slide)

*2015 FCSM Research Conference
December 1, 2015*

Authors (Alphabetic order): A Multi-Agency, Multi-Disciplinary Effort

- ***National Cancer Institute (NCI), Division of Cancer Control and Population Sciences:*** Erin Kent, Benmei Liu, Janet S. de Moor, Gordon Willis, Maggie Wilson, K. Robin Yabroff
- ***Agency for Healthcare Research and Quality (AHRQ):*** Sadeq R. Chowdhury
- ***LIVESTRONG Foundation:*** Stephanie Nutt
- ***Centers for Disease Control and Prevention (CDC), Division of Cancer Prevention and Control:*** Donatus Ekwueme, Juan Rodriguez
- ***Emory University:*** Katherine S. Virgo

Background and Research Goal



- **Research goal: empirically examine the impact of applying weight adjustments to non-probability-based survey data and compare results to those obtained from a probability-based survey based on a similar questionnaire**

Data Source #1: Probability Sample

- ***2011 Medical Expenditure Panel Survey (MEPS) Experiences with Cancer follow-back survey (CSAQ)***
 - Self-administered (paper-based) supplement to the core MEPS
 - Representative of (18+) US non-institutionalized household population of **cancer survivors**
 - MEPS full year RR=54.9%; CSAQ RR=90%
 - Analytic sample: n = 1,203

Data Source #2: Non-probability Sample

□ *2012 LIVESTRONG Survey*

- Same questionnaire as MEPS CSAQ
- **Web-based**, opt-in, available to **cancer survivors** via email, social media and newsletters
- Analytic sample (U.S. residents aged 18+ who had ever been diagnosed with cancer at or after age of 17): n= 5,394

Response rate is undefined

Different Weighting Approaches for the LIVESTRONG Sample

- Unweighted
- Sample-based Raking: *Adjust LIVESTRONG data to distribution of demographic or other characteristics from MEPS CSAQ*
- Propensity Score Adjustment (PSA): *Weight data by inverse of estimated propensity to be in the LIVESTRONG sample relative to MEPS (Lee, 2006)*
- PSA + Raking: *PSA first, then raking (Lee and Valiant 2009, Brick 2015)*

Choices of Weighting Variables

- **Four** key demographic variables: Age, sex, race/ethnicity, region (same variables used in the core MEPS raking)
- Additional **five** socio-demographic variables: education, marital status, current employment status, **cancer type, years from first diagnosis**
- Raking dimensions were formed using either
 - the four single variables; or
 - the nine single variables; or
 - the intersection of age by other variables (e.g., age*sex, age*race/ethnicity, age*sex*employment, etc.)

Table 1a: Estimates for Variables used in Weighting*

	Probability: MEPS CSAQ		Non-Probability: LIVESTRONG					
	n=1,203	Weighted	n=5,394	Unweighted	A (Raked 4 vars)	B (PSA 4 vars)	C (Raked 9 vars)	D (PSA 9 vars)
	n	%	n	%	%	%	%	%
Age								
18-49	175	13.1	2,075	38.5	13.1	26.0	13.1	27.0
50-64	390	32.3	2,484	46.1	32.3	38.9	32.3	42.4
65+	638	54.6	835	15.5	54.6	35.0	54.6	30.5
Sex								
Male	468	42.5	1,855	34.4	42.5	37.2	42.5	38.1
Female	735	57.5	3,539	65.6	57.5	62.8	57.5	61.9
Region								
Northeast	185	16.9	1,079	20.0	16.9	17.8	16.9	17.1
Midwest	294	23.2	1,181	21.9	23.2	22.4	23.2	22.3
South	475	40.7	1,698	31.5	40.7	36.7	40.7	36.7
West	249	19.2	1,436	26.6	19.2	23.0	19.2	24.0
Race/Ethnicity								
Hispanic, NH black, NH Asian	273	12.9	428	7.9	12.9	10.0	12.9	9.4
Other	930	87.1	4,966	92.1	87.1	90.0	87.1	90.6

* "PSA+Raking" method gave the same estimates as "Raking" alone for the variables used in weighting

Table 1b: Estimates for Variables used in Weighting*

	Probability: MEPS CSAQ		LIVESTRONG					
	n=1,203 n	Weighted %	n=5,394 n	Unweighted %	A (Raked 4 vars)	B (PSA 4 vars)	C (Raked 9 vars)	D (PSA 9 vars)
					%	%	%	%
Education								
High school or less	606	42.8	416	7.7	8.4	8.1	42.8	21.8
Some college or more	597	57.2	4,978	92.3	91.6	91.9	57.2	78.2
Marital Status								
Married	641	57.2	3,802	70.5	70.2	70.2	57.2	64.3
Not married	562	42.8	1,592	29.5	29.8	29.8	42.8	35.7
Current Employment Status								
Full-time	302	27.1	3,022	56.0	39.3	47.5	27.1	45.6
Part-time	105	8.8	474	8.8	8.4	8.6	8.8	9.9
Retired	380	33.8	974	18.1	41.5	29.7	33.8	24.7
Not employed for wages / Other	416	30.4	924	17.1	10.8	14.1	30.4	19.7
Cancer Type								
Breast	235	17.7	1,636	30.3	27.9	29.7	17.7	24.4
Prostate	159	14.2	341	6.3	12.0	8.8	14.2	9.9
Colorectal	59	4.7	328	6.1	6.2	6.1	4.7	4.8
Multiple	86	6.9	589	10.9	14.5	12.6	6.9	9.3
Other single cancers	664	56.5	2,500	46.3	39.5	42.9	56.5	51.6
Years from First Cancer DX								
<2	129	11.1	1,082	20.1	17.2	18.6	11.1	15.9
2-5	291	24.3	2,095	38.8	33.9	36.4	24.3	32.7
6-10	236	18.8	1,137	21.1	22.7	21.8	18.8	20.7
11+	547	45.7	1,080	20.0	26.2	23.1	45.7	30.7

*"PSA+Raking" method gave the same estimates as "Raking" alone for the variables used in weighting

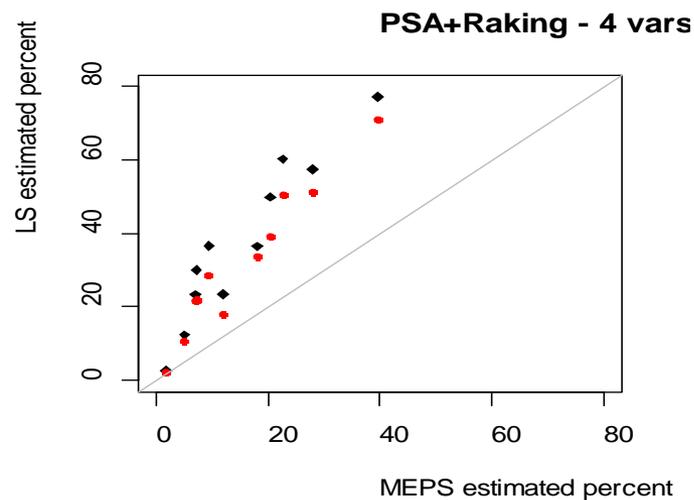
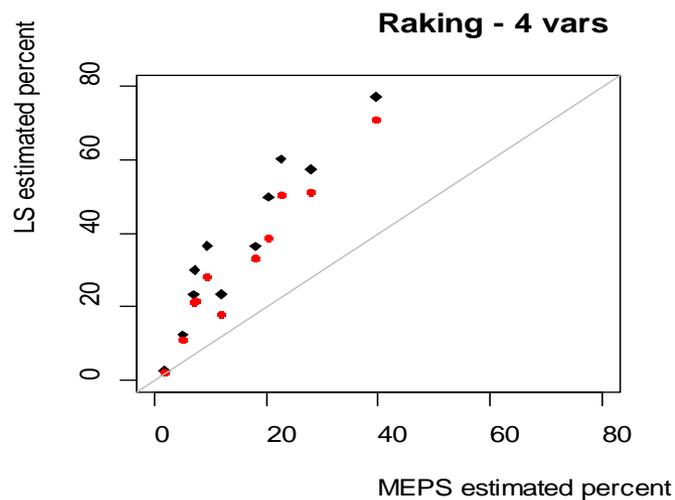
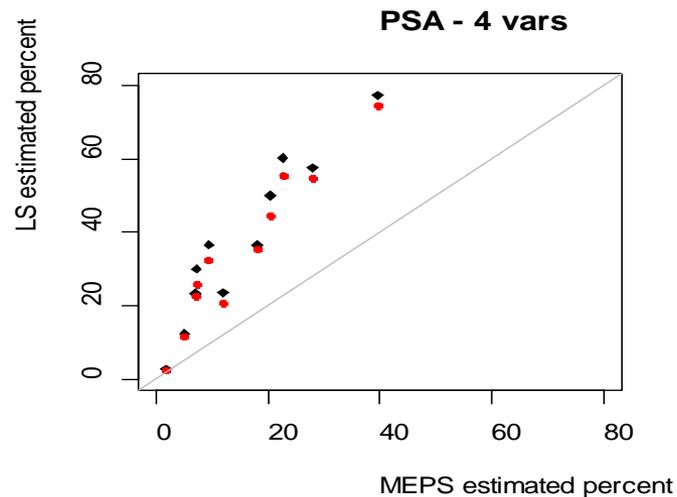
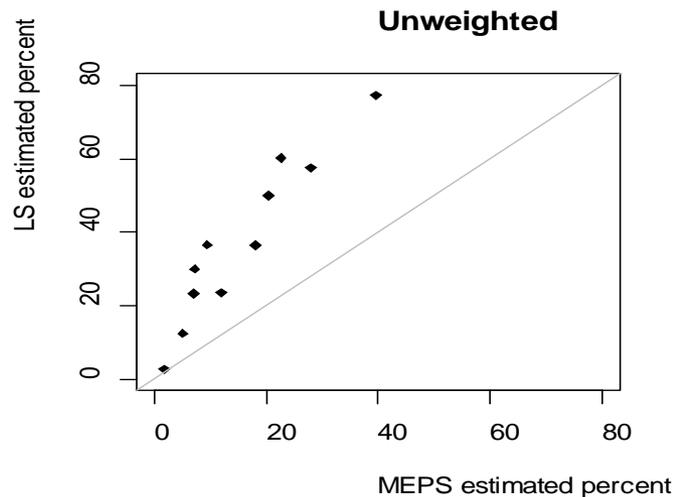
Major Outcomes of Interest

(11 binary outcomes in total)

- Employment changes (5 outcomes)
 - Made work changes since cancer diagnosis (composite measure)
 - Took extended paid time off from work because of cancer
 - Took unpaid time off from work because of cancer
 - Changed from working full-time to part-time because of cancer
 - Changed from working part-time to full-time because of cancer
- Financial Burden (6 outcomes)
 - Financial impact because of cancer (composite measure)
 - Had to borrow money or go into debt because of cancer
 - Ever filed for bankruptcy because of cancer
 - Made other financial sacrifices because of cancer
 - Ever unable to cover share of cancer medical costs
 - Ever worry about paying medical bills related to cancer

Estimates: Employment Changes and Financial Burden (11 outcomes)

Different weighting using Age, Sex, Race/Ethnicity, Region

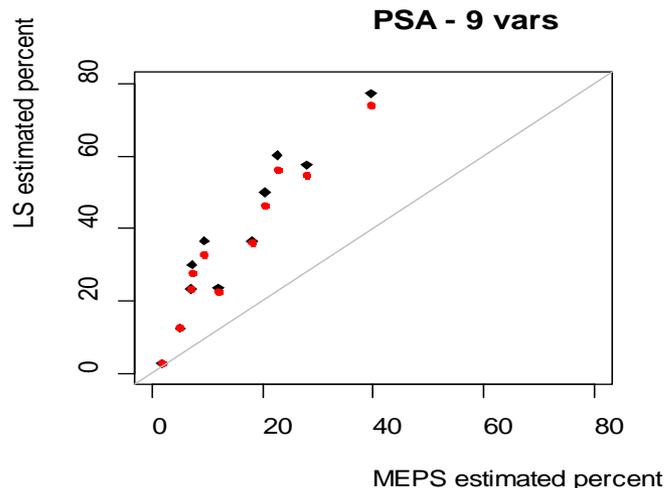
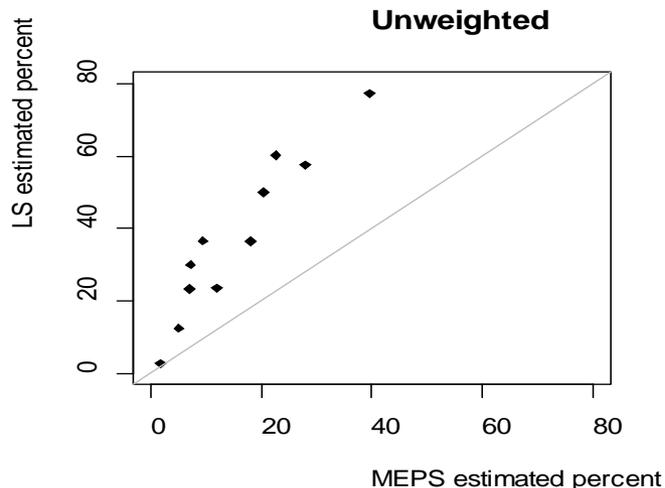


Black diamonds: unweighted; Red dots: weighted

➤ All the red points are expected to be on the diagonal if a weighting method works perfectly

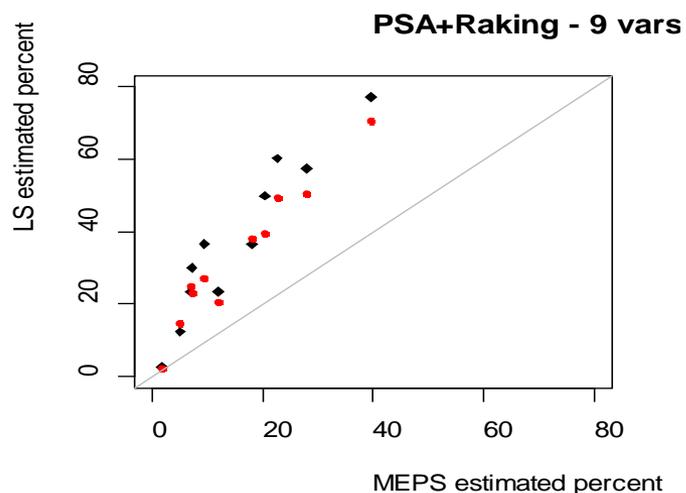
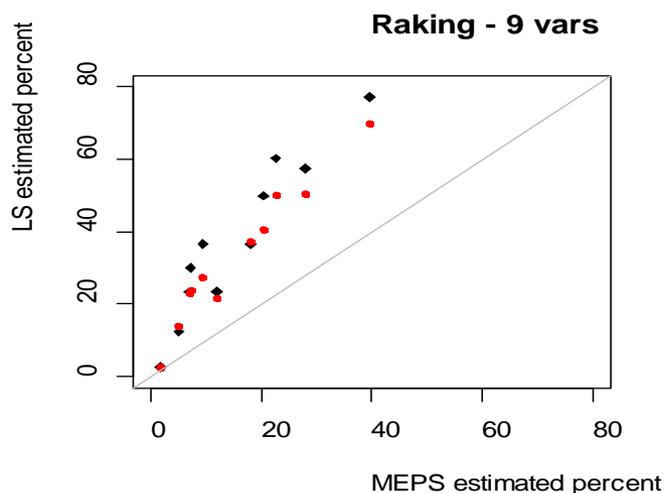
Estimates: Employment Changes and Financial Burden (11 Outcomes)

Different weighting using Age, Sex, Race/Ethnicity, Region + Five Other Variables*



*Five variables:

- Education
- Marital status
- Employment status
- Cancer type
- Years from 1st DX

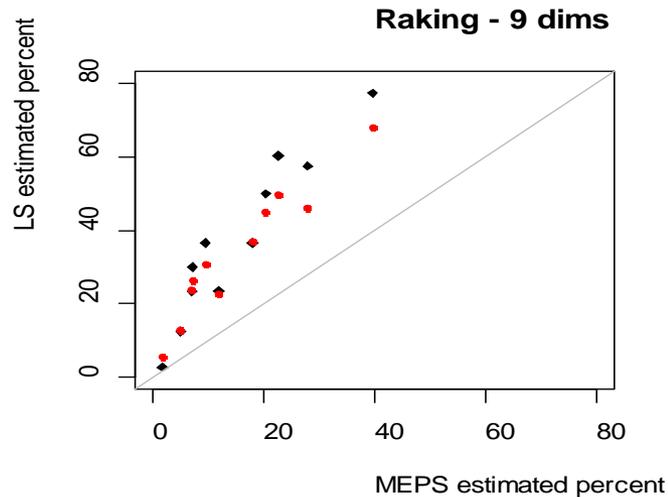
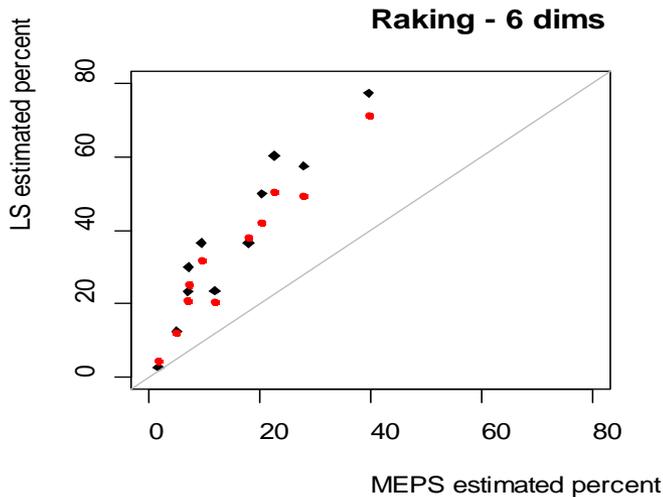
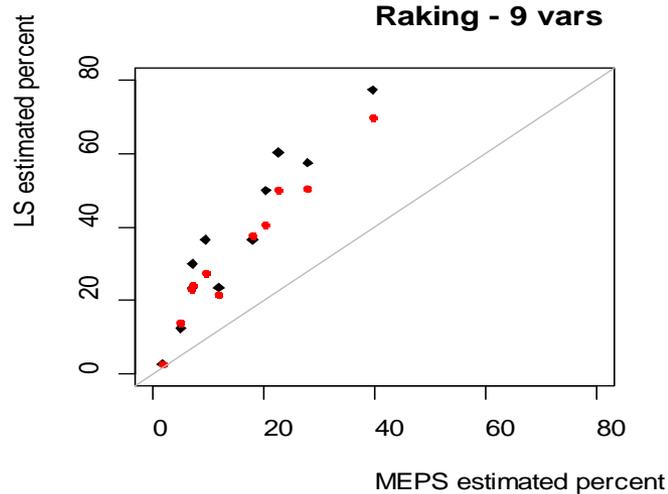
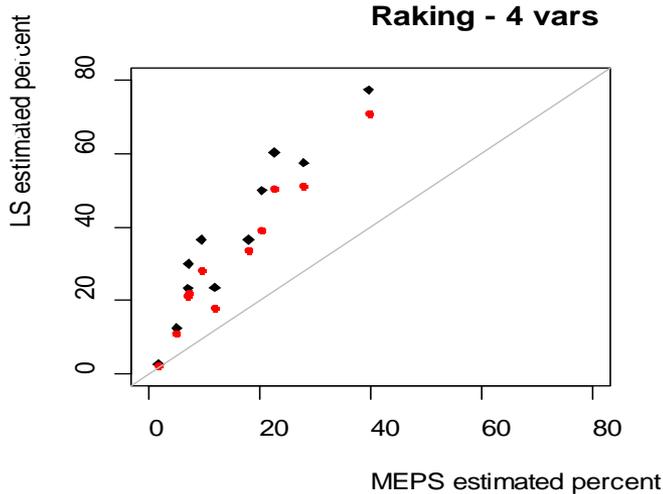


Black diamond: unweighted; Red dots: weighted

➤ All the red points are expected to be on the diagonal if a weighting method works perfectly

Estimates: Employment Changes and Financial Impacts (11 Outcomes)

Different Raking Variables/Dimensions



6 dimensions:

- age*sex
- Age*raceethnic
- Age*region
- Age*employment
- Age*sex*employment
- Age*marital status

Additional 3 dimensions:

- Age*education
- Age*cancer type
- Age*yrs since DX

Black diamond: unweighted; Red dot: weighted

➤ All the red points are expected to be on the diagonal if a weighting method works perfectly

Association Analyses

Run Multivariate logistic regression models

- Two outcomes:
 - Any financial Impact due to cancer
 - Any work change due to cancer
- Predictors included:
Age(3), Sex(2), Education(2), Race/ethnicity(2), Marital Status(2), Region(4), Years from Cancer Diagnosis(4)
 - Degrees of freedom: 12
- Unweighted and weighted using different set of weights

Table 2: Association Between Variables: Adjusted ORs
Dependent Variable: Any financial Impact due to cancer

Respondent characteristics	MEPS (n=1,203)	LIVESTRONG (n=5,394)	
	Weighted	Unweighted	Weighted (Raked 4 vars)
	OR (95% CI)	OR (95% CI)	OR (95% CI)
Age group			
18-49	3.41 (2.14 - 5.41)	4.13 (3.44 - 4.96)	3.75 (2.98 - 4.72)
50-64	1.73 (1.19 - 2.54)	2.68 (2.24 - 3.19)	2.53 (2.07 - 3.08)
65+	REF	REF	REF
Sex			
Male	REF	REF	REF
Female	1.38 (0.93 - 2.05)	1.27 (1.13 - 1.43)	1.40 (1.15 - 1.69)
Education			
High school graduate or less	1.04 (0.72 - 1.51)	1.44 (1.16 - 1.77)	1.27 (0.93 - 1.75)
Some college or more	REF	REF	REF
Race/Ethnicity			
Hispanic, NH black, NH Asian	2.21 (1.51 - 3.23)	1.23 (1.00 - 1.53)	1.52 (1.03 - 2.25)
Other	REF	REF	REF

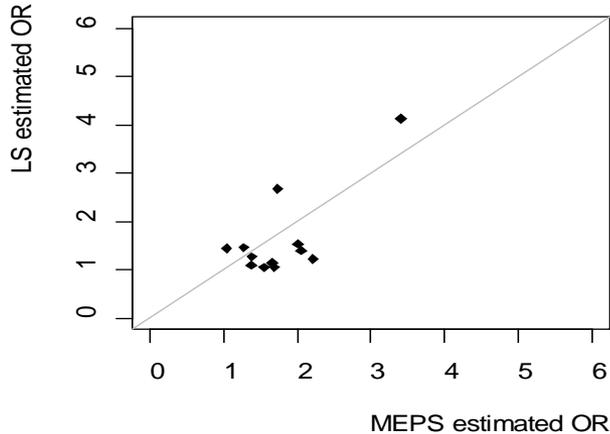
Respondent characteristics	MEPS (n=1,203)	LIVESTRONG (n=5,394)	
	Weighted	Unweighted	Weighted (Raked 4 vars)
	OR (95% CI)	OR (95% CI)	OR (95% CI)
Marital Status			
Married	REF	REF	REF
Not married	1.27 (0.90 - 1.79)	1.46 (1.29 - 1.65)	1.33 (1.09 - 1.63)
Region			
Northeast	REF	REF	REF
Midwest	1.66 (0.85 - 3.26)	1.14 (0.96 - 1.35)	1.06 (0.80 - 1.40)
South	2.01 (1.11 - 3.65)	1.54 (1.31 - 1.81)	1.43 (1.11 - 1.85)
West	2.05 (1.05 - 4.02)	1.40 (1.19 - 1.65)	1.36 (1.04 - 1.78)
Years from First Cancer DX			
<2	1.69 (0.99 - 2.88)	1.05 (0.88 - 1.25)	1.24 (0.93 - 1.67)
2-5	1.55 (1.06 - 2.27)	1.05 (0.90 - 1.23)	1.15 (0.90 - 1.47)
6-10	1.37 (0.86 - 2.20)	1.10 (0.93 - 1.31)	1.27 (0.96 - 1.67)
11+	REF	REF	REF

Association Between Variables: Adjusted ORs

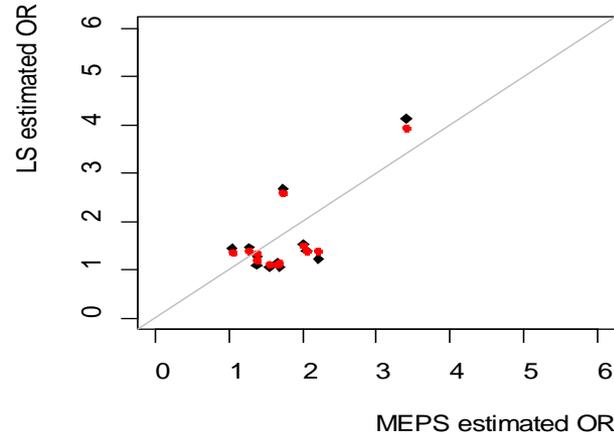
Dependent Variable: Any financial Impact due to cancer

Different weighting using Age, Sex, Race/Ethnicity, Region

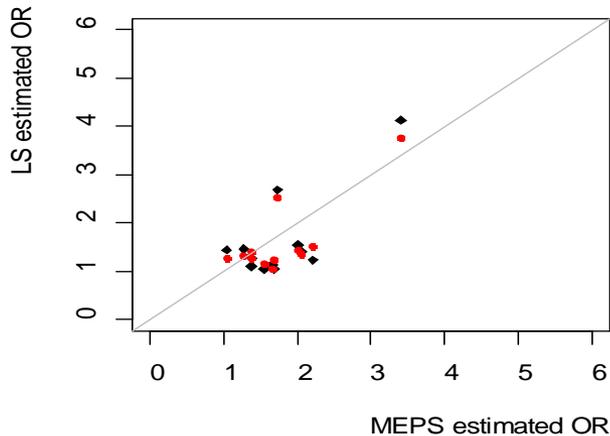
Unweighted



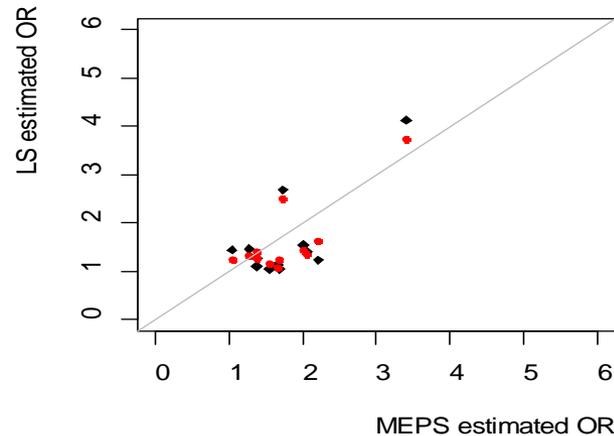
PSA - 4 vars



Raking - 4 vars



PSA+Raking - 4 vars



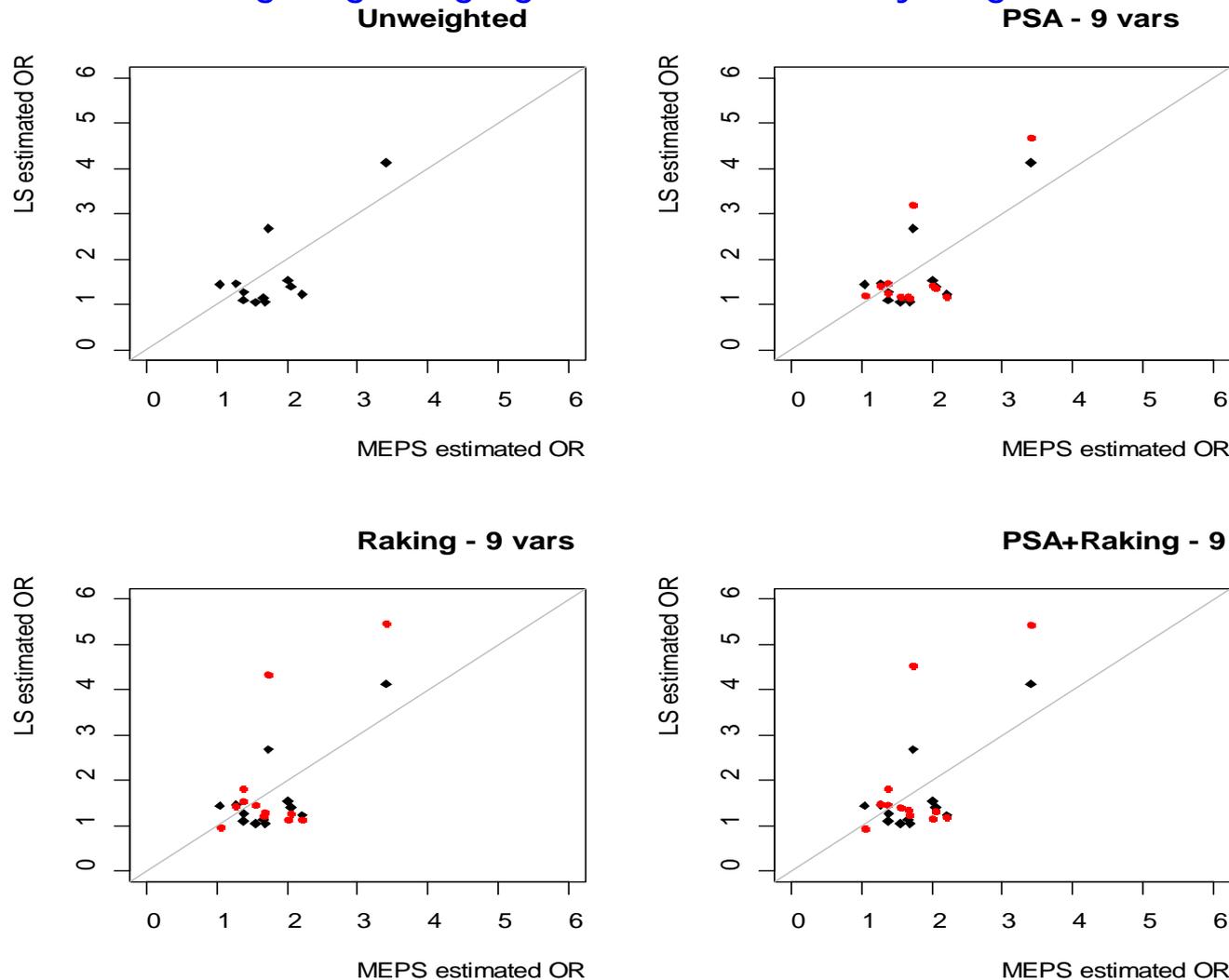
Black diamond: unweighted; Red dot: weighted

➤ All the red points are expected to be on the diagonal if a weighting method works perfectly

Association Between Variables: Adjusted ORs

Dependent Variable: Any financial Impact due to cancer

Different weighting using Age, Sex, Race/Ethnicity, Region + Five Other Variables*



*Five variables:

- Education
- Marital status
- Employment status
- Cancer type
- Years from 1st DX

Black diamond: unweighted; Red dot: weighted

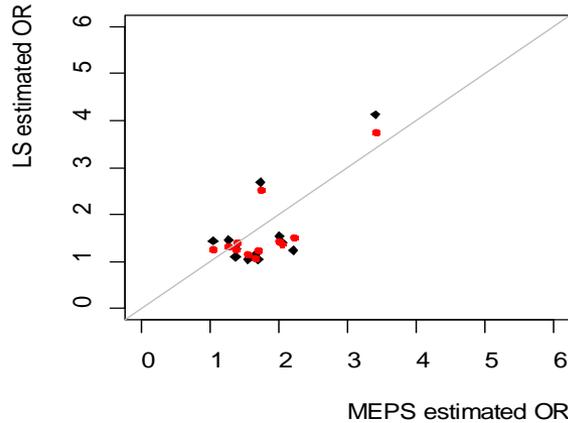
➤ All the red points are expected to be on the diagonal if a weighting method works perfectly

Association Between Variables: Adjusted ORs

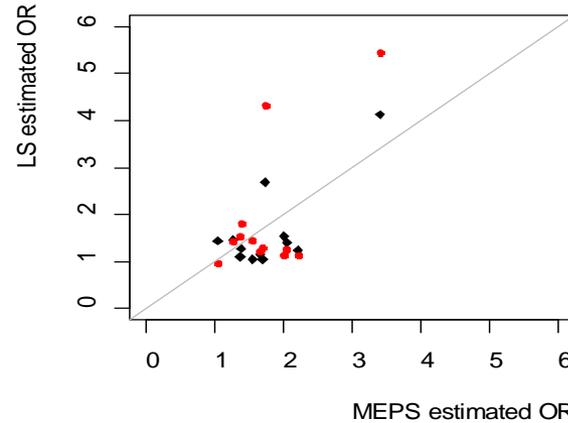
Dependent Variable: Any financial Impact due to cancer

Different Raking Variables/Dimensions

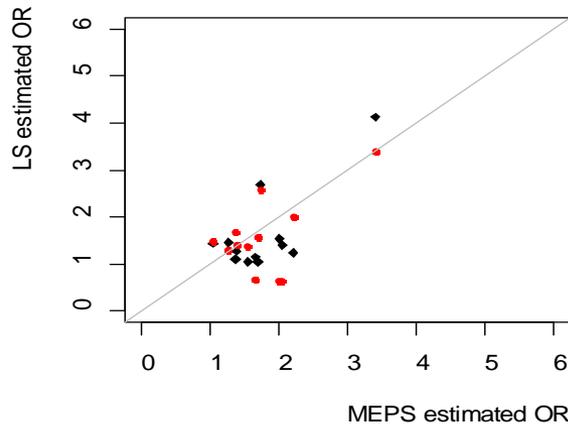
Raking - 4 vars



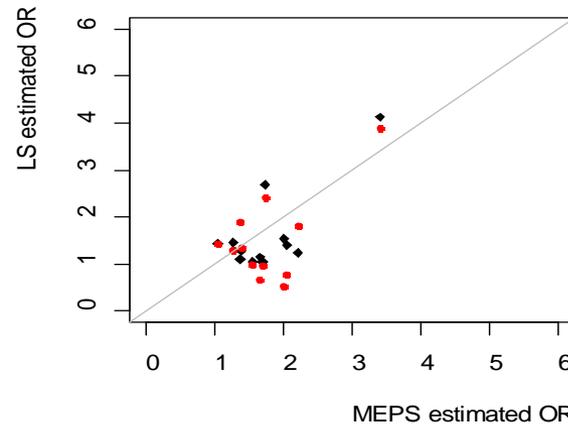
Raking - 9 vars



Raking - 6 dims



Raking - 9 dims



6 dimensions:

- age*sex
- Age*raceethnic
- Age*region
- Age*employment
- Age*sex*employment
- Age*marital status

Additional 3 dimensions:

- Age*education
- Age*cancer type
- Age*yrs since DX

Black diamond: unweighted; Red dot: weighted

➤ All the red points are expected to be on the diagonal if a weighting method works perfectly

Summary and Discussion

For estimation of (absolute) population quantities:

- *For our measures of financial burden and employment, estimates from **LIVESTRONG** non-probability sample, even weighted, were generally not 'close' to those of the **MEPS-CSAQ** probability sample*

For associations (relative measures):

- *Analysis of associations, via regression analysis, illustrated more similarity between surveys irrespective of weighting methods or no weighting*

Overall:

- *Bias due to non-probability sampling may be more of a problem for quantity estimation*

Summary and Discussion

- Raking is more efficient than the propensity score weighting approach in terms of reducing bias
- The composite approach (PSA first then raking) may give similar results as raking alone
- Weighting variables and raking dimensions need to be carefully chosen, weighting may introduce more bias depending on the set of weighting variables used
- Raking with carefully chosen variables helps reduce some bias, but not a lot

Limitations

- **Mode confounding?** MEPS-CSAQ was paper-based, LIVESTRONG a web survey
- **MEPS contains sampling error**
 - Implication for control totals (adding additional variances to the LIVESTRONG weighted estimates)
 - Some cell sizes are very small
 - Challenges in variance estimation

References

1. Lee, Sunghee (2006). Propensity score adjustment as a weighting scheme for volunteer panel web surveys. *Journal of Official Statistics* 22(2):329–49.
2. Lee, Sunghee, and Richard Valliant (2009). Estimation for volunteer panel web surveys using propensity score adjustment and calibration adjustment. *Sociological Methods and Research* 37(3):319-43.
3. Brick, J. Michael (Sept, 2015). Non-Probability sampling assumptions and methods. Presented at the Washington Statistical Society Seminar on the topic of Non-probability Samples.

Any Questions?

Thank you!

Contact info:

Benmei Liu

liub2@mail.nih.gov