Redirected Inbound Call Sampling – An Example of Fit for Purpose Non-probability Sample Design

Burton Levine
Karol Krotki

NISS/WSS Workshop on Inference from Nonprobability Samples

September 25, 2017

RTI International is a registered trademark and a trade name of Research Triangle Institute.

Overview

- Describe Redirected Inbound Call Sampling—RICS
- **NHIS evaluation study—goal is to quantify bias**
  - Data collection metrics
  - Distributions of respondents
  - Compare weighted estimates to “gold standard”
  - Night owls
  - Evaluate primacy effect
- **IVR system implementation study—goal to optimize the instrument**
- **Method to create confidence intervals that reflect bias as well as variance**
- Outstanding RICS issues

Introducing Redirected Inbound Call Sampling (RICS) Surveys

RICS Survey participants come from:

- **Misdials to non-working toll-free numbers**
- Toll-calls that fail to connect to their intendent target

Data can be collected through several methods:

- **Interactive voice response (IVR) system**
- Live interviewer
- Sent to a web-site

Redirected Inbound Toll-free Calls are Commonplace

In a random sample of 100 toll-free phone numbers

- 39 were working and not redirected
- 61 were non-working or redirected
  - 13 (21%) of the these were redirected numbers
    - 6 were offers for free cruises that require listening to a time-share pitch
    - 6 were for either a Medic-Alert devise or auto repair insurance
    - 1 was for erectile dysfunction medication.

RICS Survey We Fielded and Conference Presentations

At RTI we fielded 6 RICS surveys.

Date	Name	Respondents
Sept 2015	BRFSS Evaluation Study	6,799
Oct 2016	National Adult Tobacco Survey	4,302
March 2017	**NHIS Evaluation Study**	**10,469**
April 2017	New York City Sleep Study	1,532
August 2017	National Adult Tobacco Survey	4,630
Still in field	**IVR Evaluation Study**	**TBD**

RICS Conference presentations

- 2016 AAPOR
- 2017 AAPOR—Invited session dedicated to RICS
- 2017 JSM

Paper in progress

Analyze metadata—10 million file

Each inbound calls contains the following metadata

- Inbound telephone number
- Billing ZIP
- State and County (based on billing ZIP)
- Phone type (landline/cell phone)
- Local carrier where call was dropped
- Time the call was placed

Phone type	Percent of calls
Landline	44
Wireless	55
Unknown	1

We analyzed 10,000,000 inbound calls.

Analyze the metadata (continued)

NHIS Evaluation Study—Pilot

We created a 27 question instrument that mimics questions from the:

- American Community Survey (ACS)
- National Health Interview Survey (NHIS)—Adult sample
- Used to test the paradata—not discussed in this talk

We piloted the original questionnaire using collecting 300 respondents.

Problems

- High item nonresponse (up to 40%) for continuous outcomes
- Large number of break-offs (up to 7%) with some questions

NHIS Evaluation Study—Pilot (continued)

Solutions:

- We rewrote continuous questions as categorical
  - Hours of sleep
  - Age
  - Phone calls placed in a day
  - Twitter use in last week
  - Days per month drink alcohol
  - Alcoholic drinks per day, on the days one drinks

> Using the phone keypad, please enter your age___. If you prefer not to answer press the pound key.
>
> How old are you?
> a. If you're 18-24 years-old, press 1
> b. If you're 25-34 years-old, press 2
> c. If you're 35-44 years-old, press 3
> d. If you're 45-54 years-old, press 4
> e. If you're 55-64 years-old, press 5
> f. If you're 65 years-old or older, press 6
> g. If you prefer not to answer press 7

Moved sensitive questions to the end of the questionnaire

- Ever used guns
- ZIP Code

NHIS Evaluation Study—Data collection metrics

Data collection for two separate one-week periods in 2017

- January 6—January 13
- February 24—March 3

Inbound calls	Respondents	Yield rate (%)
139,022	10,469	7.5

Response rate

Respondents	Non-respondents	Unknown response status	Ineligible	e	AAPOR4
10,469	11,378	113,296	3,879	0.85	8.87%

NHIS Evaluation Study—Data collection metrics *(continued)*

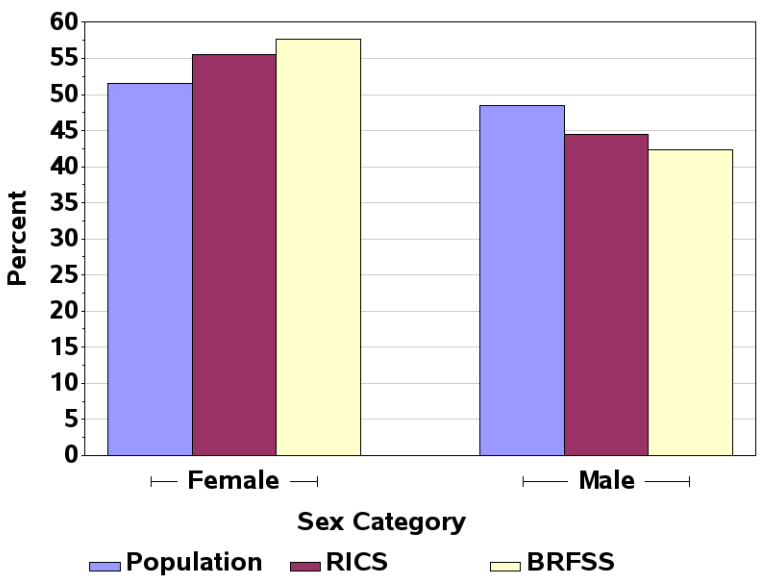
Flow of subjects through the study

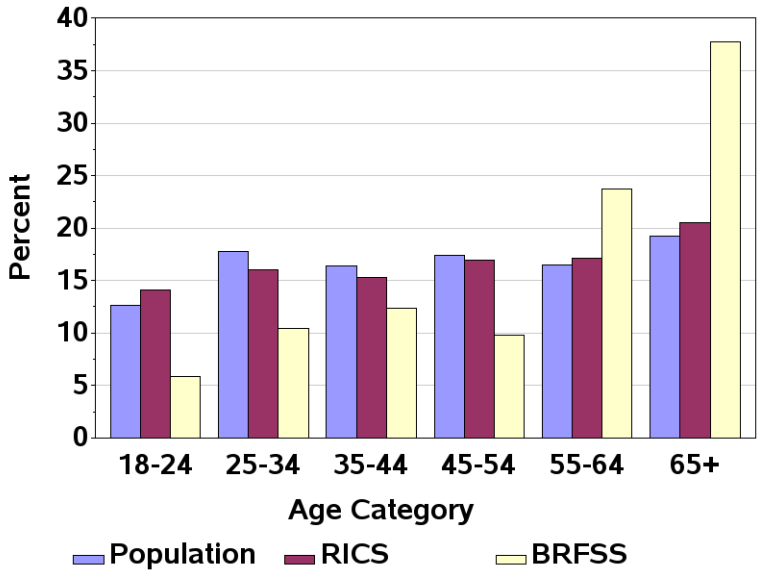
Category	Quantity	Percent of MIDI calls	Percent of previous row
Inbound calls	139,022	100	N/A
Eligible geography	137,840	99	99
Responded to screener	24,735	18	18
Adults	21,998	16	89
Respondents	10,469	8	48
Finished Survey	8,157	6	78

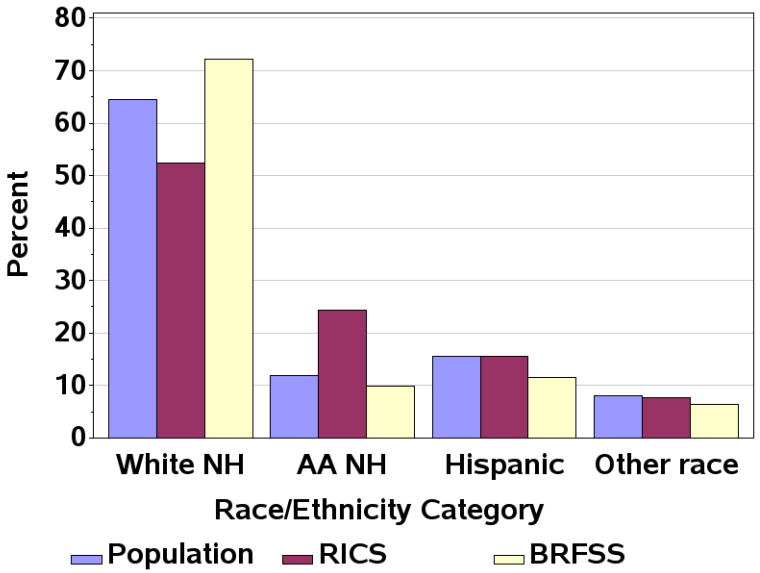
Interview length in minutes

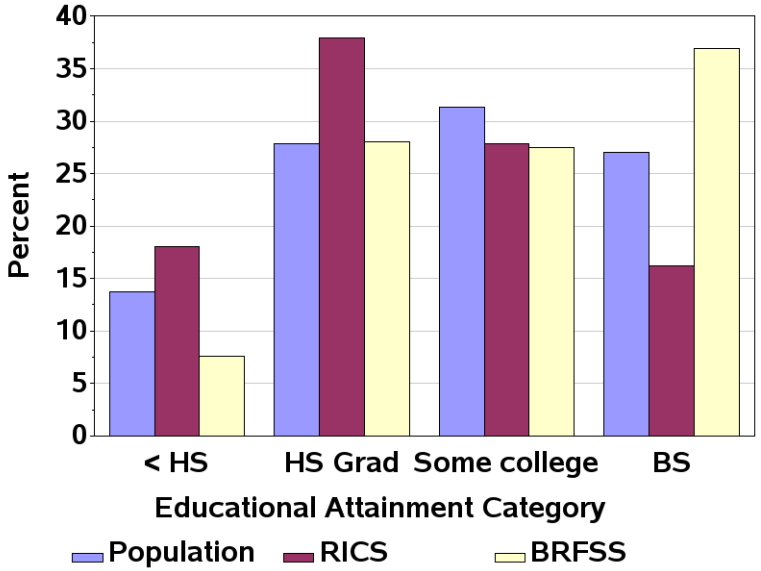
minimum	Percentile					maximum
	10th	25th	50th	75th	90th	
1.2	2.7	4.1	4.8	5.4	6.1	11.3

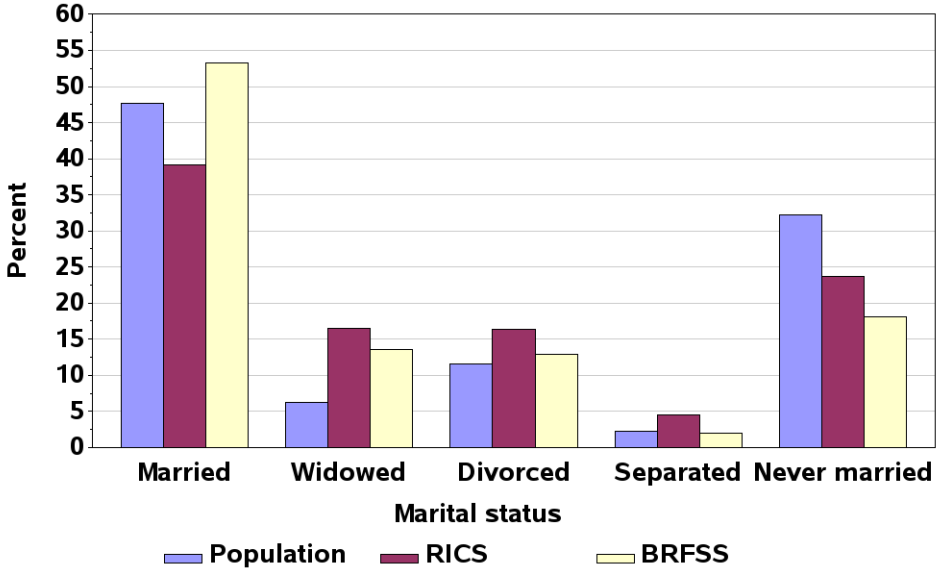
NHIS Evaluation Study— Comparison of demographic distributions among: population, RICS respondents and BRFSS respondents

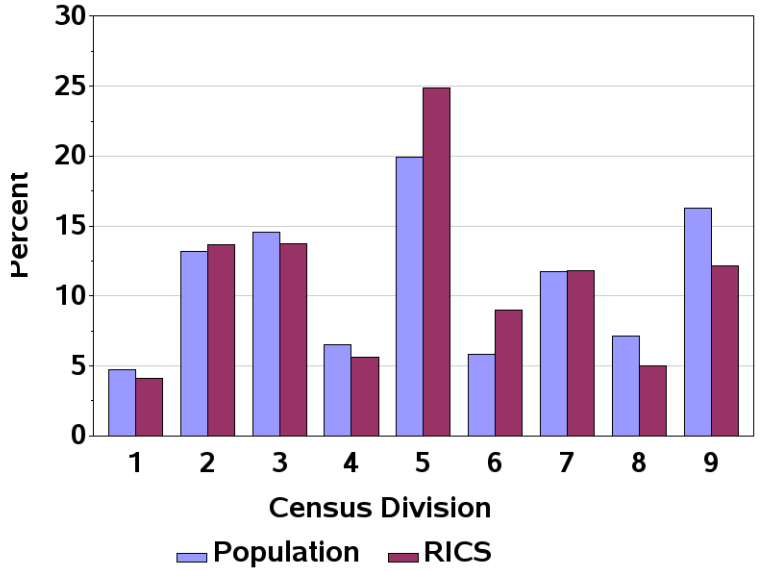








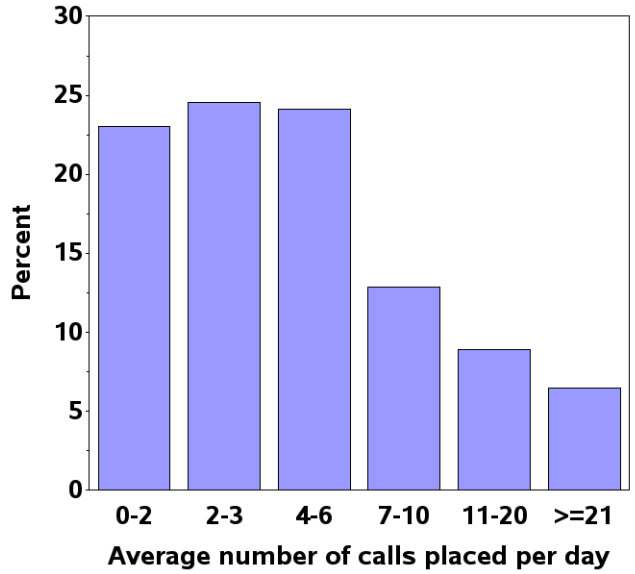




NHIS Evaluation Study— Calculating Sampling Weights

We calculated the weights two ways

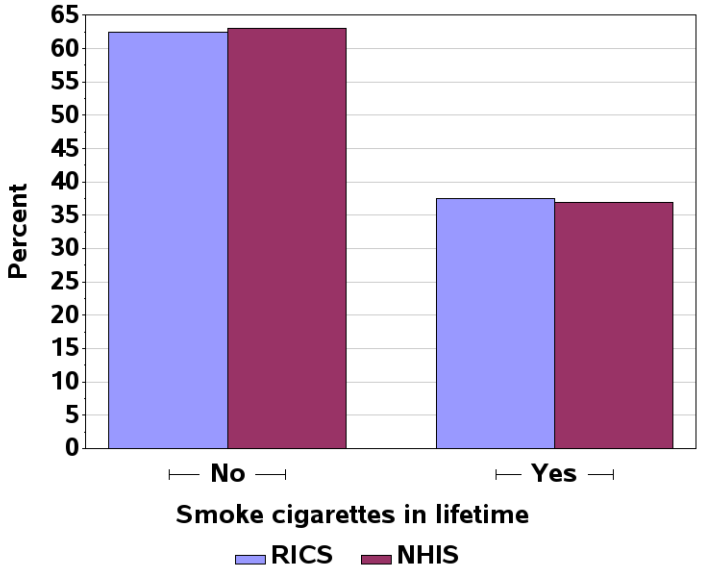
1. The base weight was inversely proportional to the average number of calls placed per day (UWE=2.14)
2. The base weight was equal for all respondents (UWE=1.27)

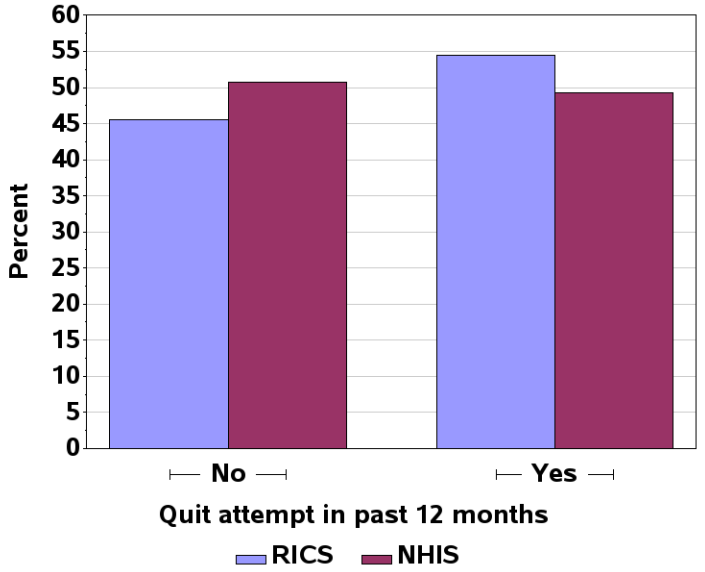


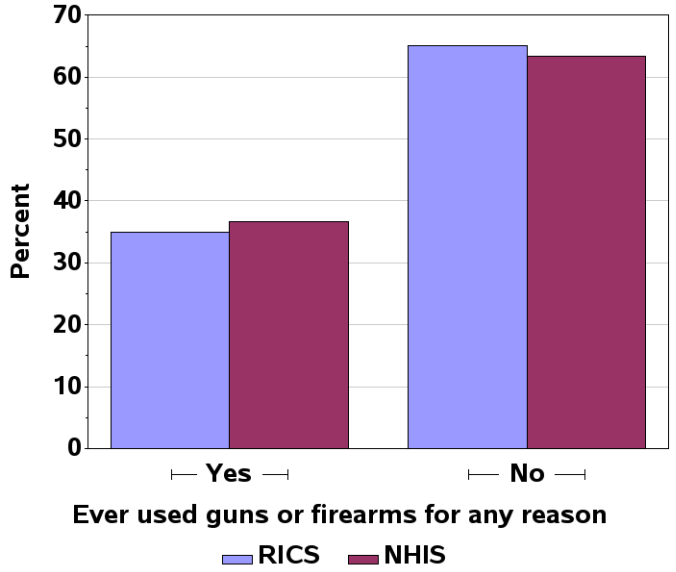
In both methods we calibrated to the following national marginal distributions estimated from the 2015 1-year ACS.

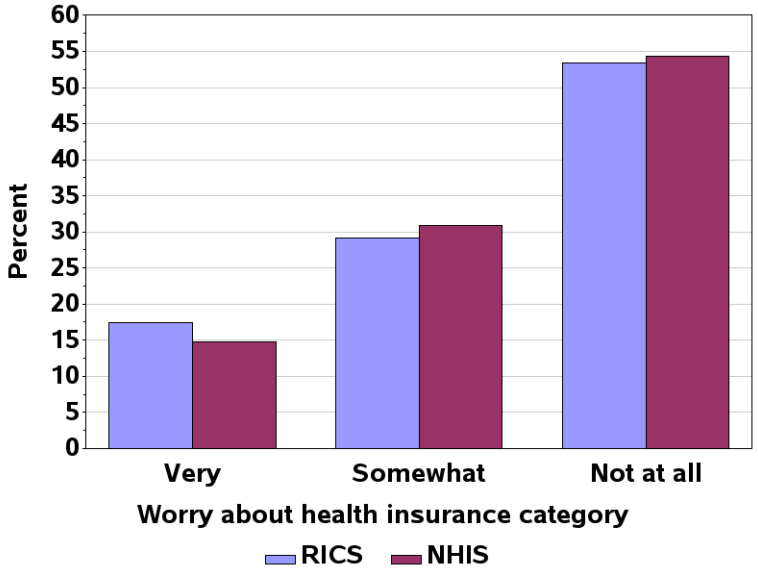
- Sex (2-levels)
- Age category (6-levels)
- Race/ethnicity (5-levels)
- Educational attainment (4-levels)
- Census division (9-levels)

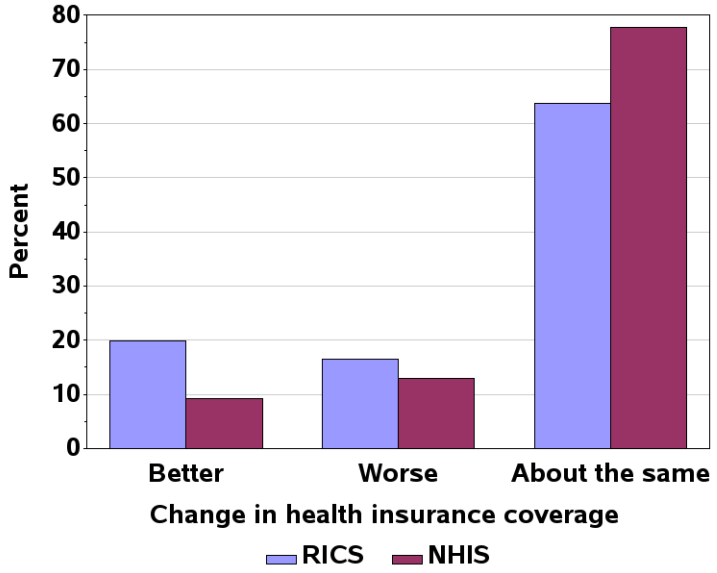
NHIS Evaluation Study— Comparing categorical outcomes



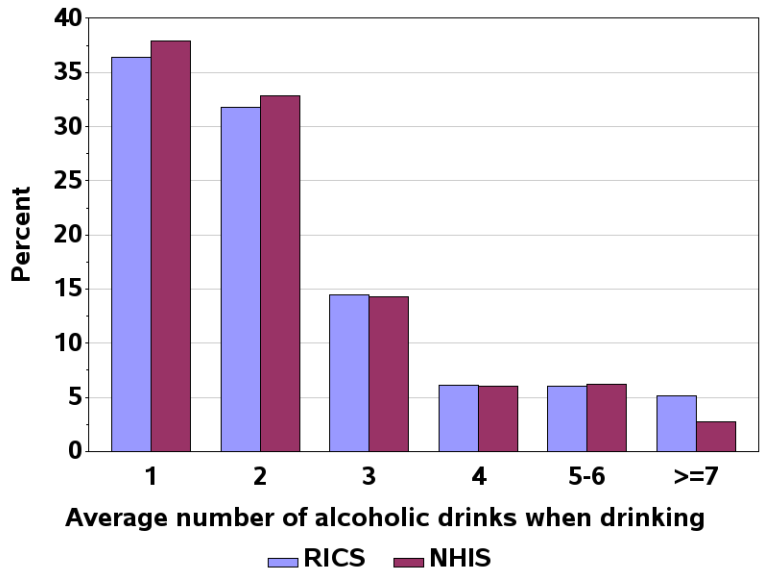


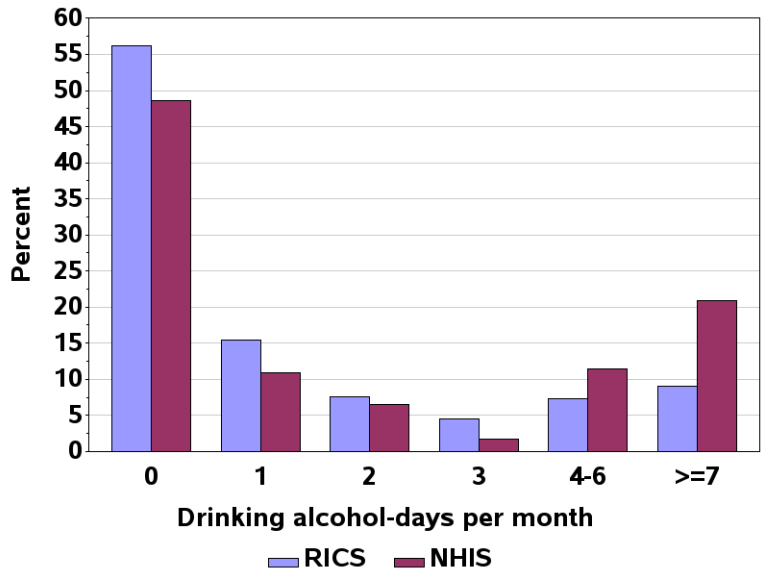


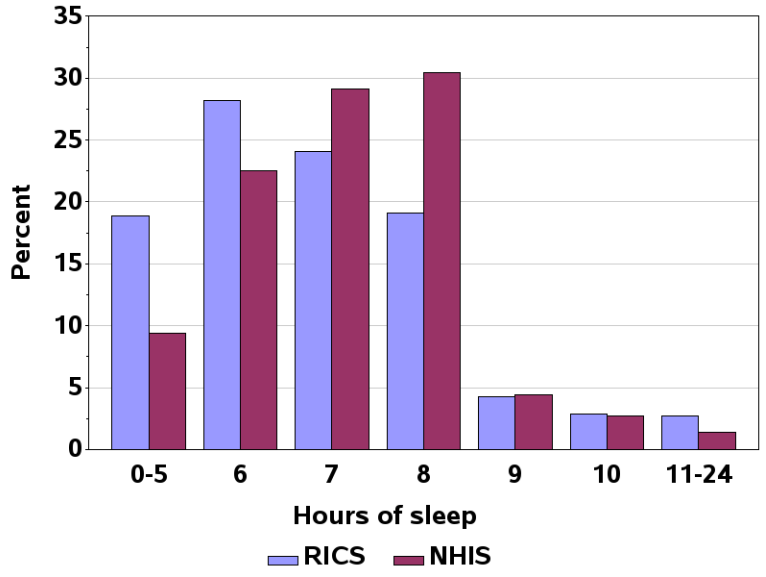


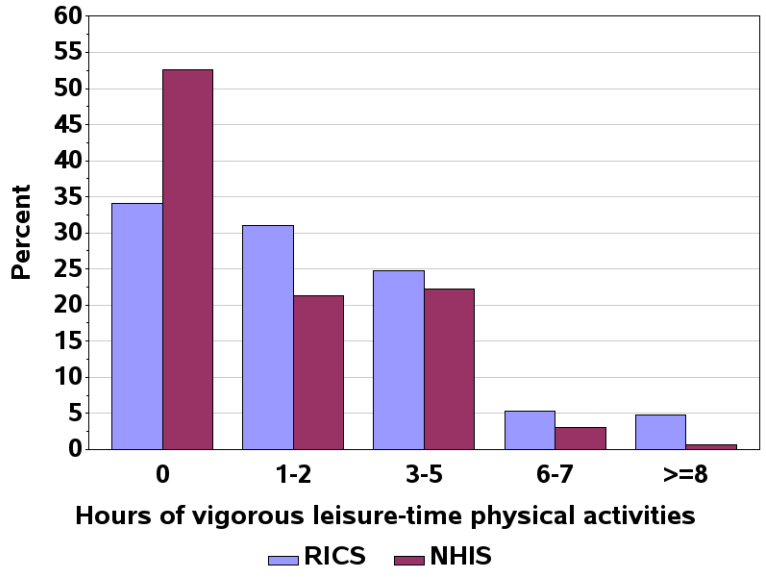


NHIS Evaluation Study— Comparing continuous outcomes





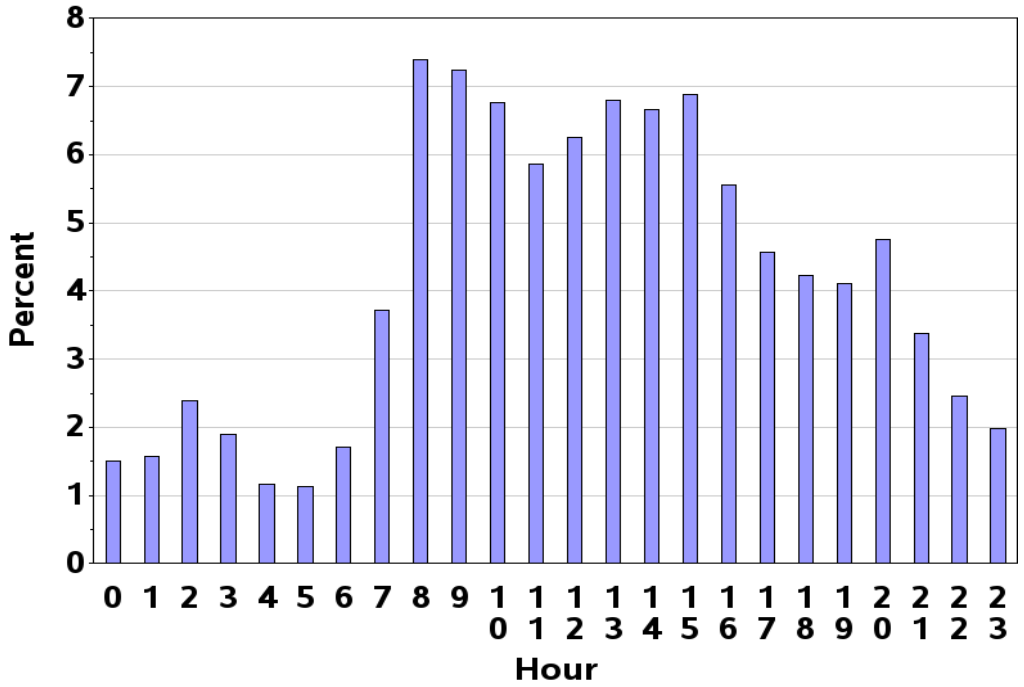




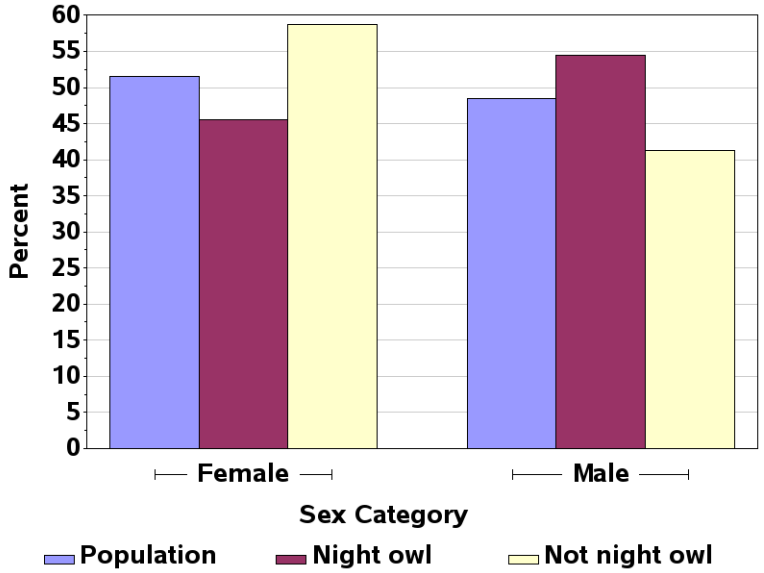
NHIS Evaluation Study—Night owls

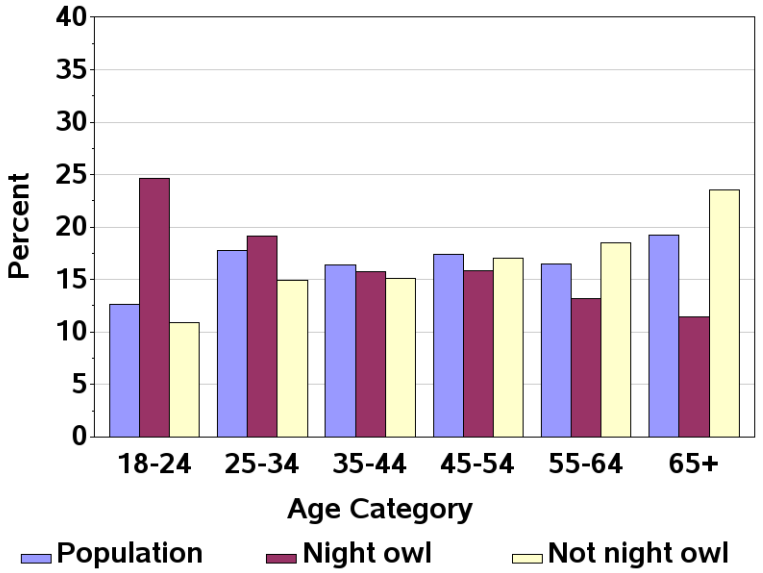
We define *night owls* as individuals that respond to the survey between 10pm and 8am.

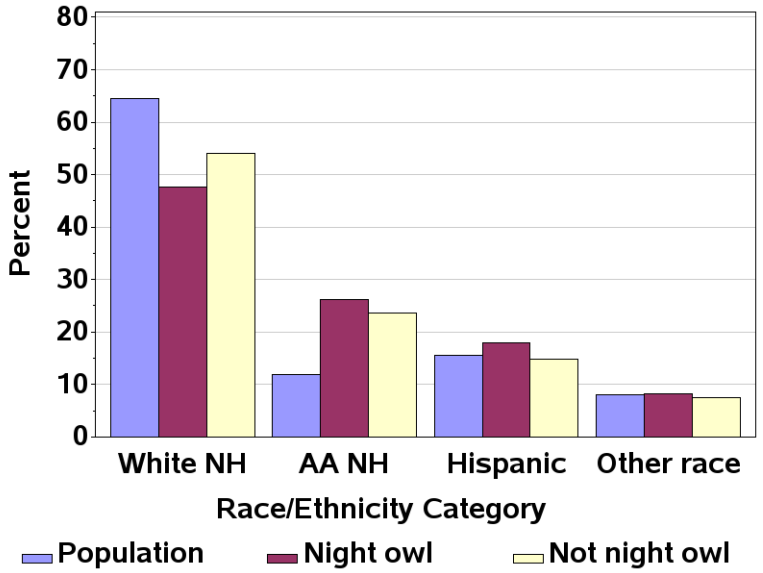
24% of the respondents are night owls

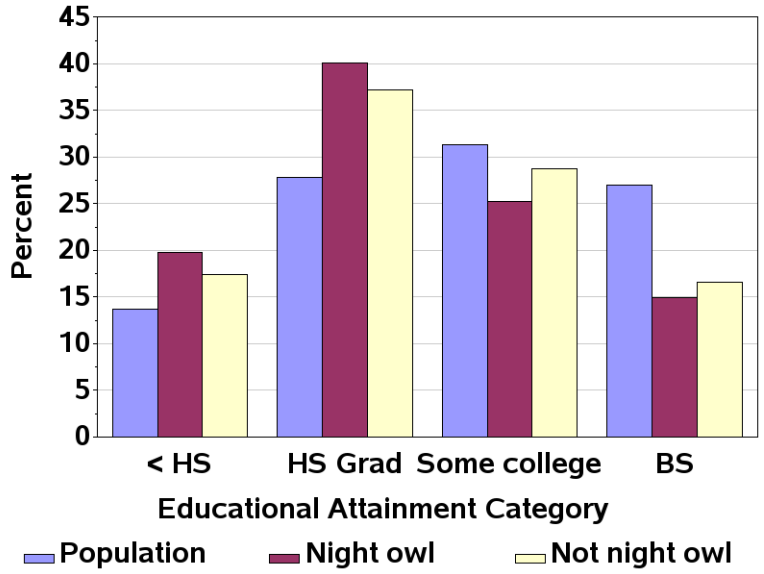


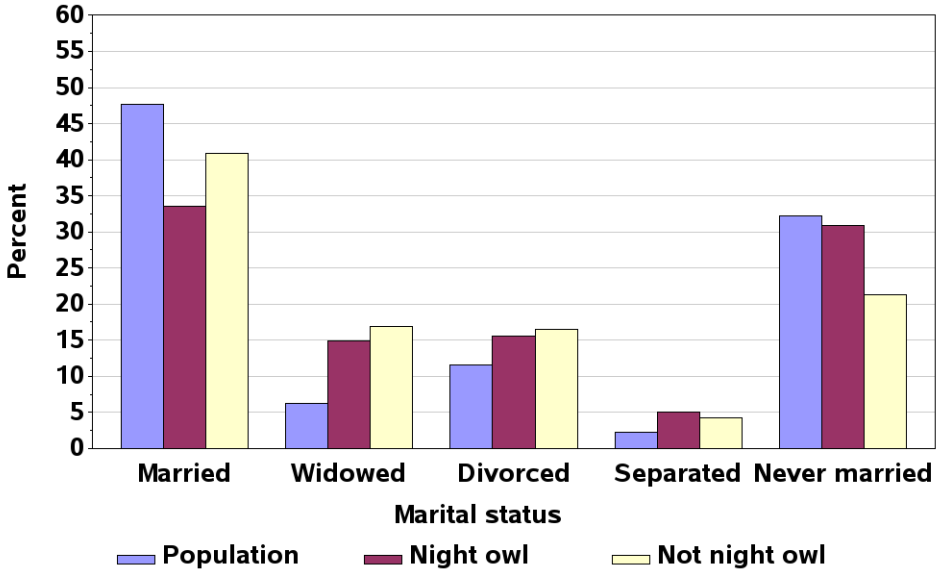
NHIS Evaluation Study—Night owls (continued)







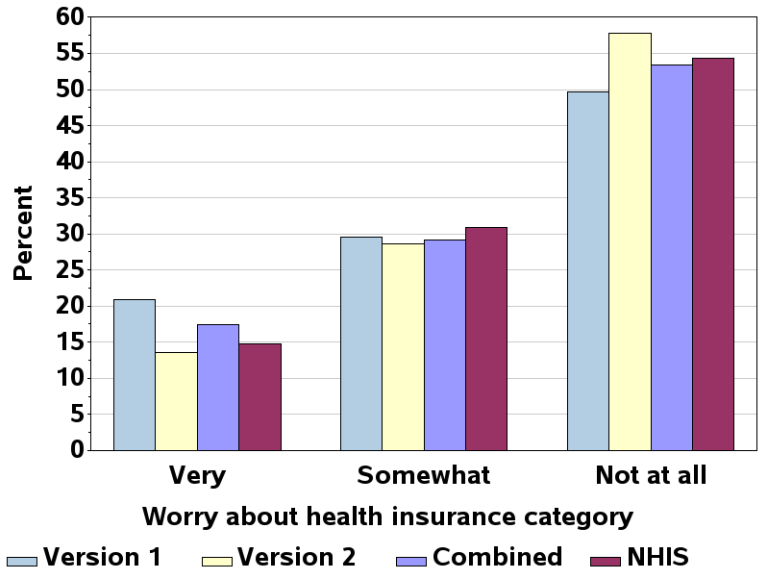




NHIS Evaluation Study—Investigating the ordering of the categories

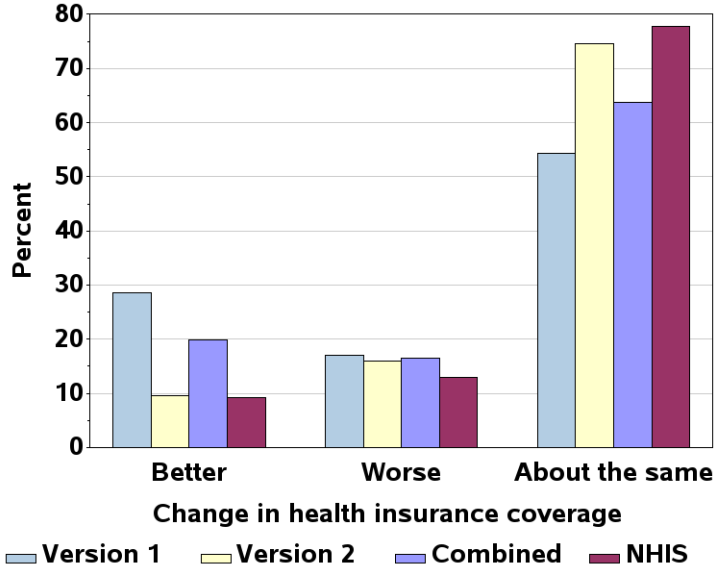
If you get sick or have an accident, how worried are you that you will be able to pay your medical bills?

Order	Version 1	Version 2
1	Very worried	Not at all worried
2	Somewhat worried	Somewhat worried
3	Not at all worried	Very worried



In regard to your health insurance or health care coverage, how does it compare to a year ago?

Order	Version 1	Version 2
1	Better	About the same
2	Worse	Worse
3	About the same	Better



IVR Evaluation Survey

2 by 2 experiment—Categorical/continuous by global/local prefer not to answer prompt

- Categorical vs continuous response options for 4 questions
  - Hours of sleep
  - Physical activity per week
  - Alcohol per month
  - Number of drinks
- Global vs local prefer not to answer prompt
- One second delay in the prefer not to answer prompt for the test condition continuous response and local prefer not to answer prompted.
- Also tested
  - Softball question
  - No break-in (NBI)

IVR Evaluation Survey *(continued)*

	5 different implementations of the IVR				
Prefer-not-to-answer	Global	Local	Local	Local	Global
Continuous outcomes coded	Categorical	Categorical	Continuous	Continuous	Continuous
Delay prefer-not-to-answer	No	No	No	Yes	No
Inbound calls recruited	6,324	6,287	5,511	5,507	10,197
Respondents	505	615	465	528	748
Yield rate (%)	8.0	9.8	8.4	9.6	7.3
Break off (%)	15.1	21.3	21.7	20.5	13.7
Question	Item nonresponse (%)				
Smoke 100 cigarettes	14.5	4.6	2.2	2.1	(SB-3.9) 12.7
Smoke every day, some days, or not at all	0.4	0.5	0.0	0.4	0.1
Hours of sleep	7.5	4.7	19.6	13.1	8.2
Sex	2.8	1.6	2.2	1.7	1.1
Age	3.2	2.3	2.6	1.7	4.0
Hispanic	4.0	13.8	7.7	7.2	2.9
Race	3.4	11.2	7.5	5.7	(NBI-2.4) 3.9
Educational attainment	5.0	4.7	4.5	3.2	(NBI-2.8) 2.8
Physical activity	0.4	7.2	32.9	22.6	(NBI-3.0) 6.2
Drink alcohol—days per month	0.2	5.9	20.7	13.9	(NBI-2.7) 3.9
Number of alcoholic drinks	17.0	16.4	28.6	18.9	12.6
Last time you worked	8.2	10.8	12.6	12.5	5.6
Average item nonresponse	5.6	7.0	11.8	8.6	5.3

SB—Softball questions; NBI—No barge-in

IVR Evaluation Survey *(continued)*

		5 different implementations of the IVR					Population
Prefer-not-to-answer		Global	Local	Local	Local	Global	
Continuous outcomes coded		Categorical	Categorical	Continuous	Continuous	Continuous	
Delay prefer-not-to-answer		No	No	No	Yes	No	
Characteristic	Category	Percentage of respondents in each demographic category					
Sex	Male	40.3	44.0	45.1	45.1	46.8	48.7
	Female	59.7	56.0	54.9	54.9	53.2	51.3
Age	18-24	9.8	10.0	11.3	12.1	10.4	12.4
	25-34	14.9	17.1	16.6	16.0	15.3	17.9
	35-44	19.0	15.5	13.7	14.8	15.3	16.2
	45-54	20.9	16.5	18.8	16.8	16.3	17.1
	55-64	17.4	19.1	16.3	19.8	18.8	16.6
	65 or older	18.0	21.8	23.4	20.4	23.8	19.7
Race	White NH	55.2	49.9	56.9	52.7	54.6	64.3
	Black NH	**19.3**	**25.4**	**22.3**	**22.6**	**22.6**	**12.1**
	Hispanic	18.1	15.8	14.1	16.2	13.6	15.7
	Other race	7.3	8.9	6.6	8.5	9.2	8.0
Educational attainment	Less than high school	18.1	16.9	17.6	16.4	15.1	11.7
	High school grad	34.2	36.2	35.4	38.0	39.1	29.0
	Some college	29.2	27.0	28.2	26.0	26.8	28.7
	BS or higher	18.5	20.0	18.9	19.6	19.0	30.7
UWE		**1.24**	**1.33**	**1.24**	**1.30**	**1.25**	N/A

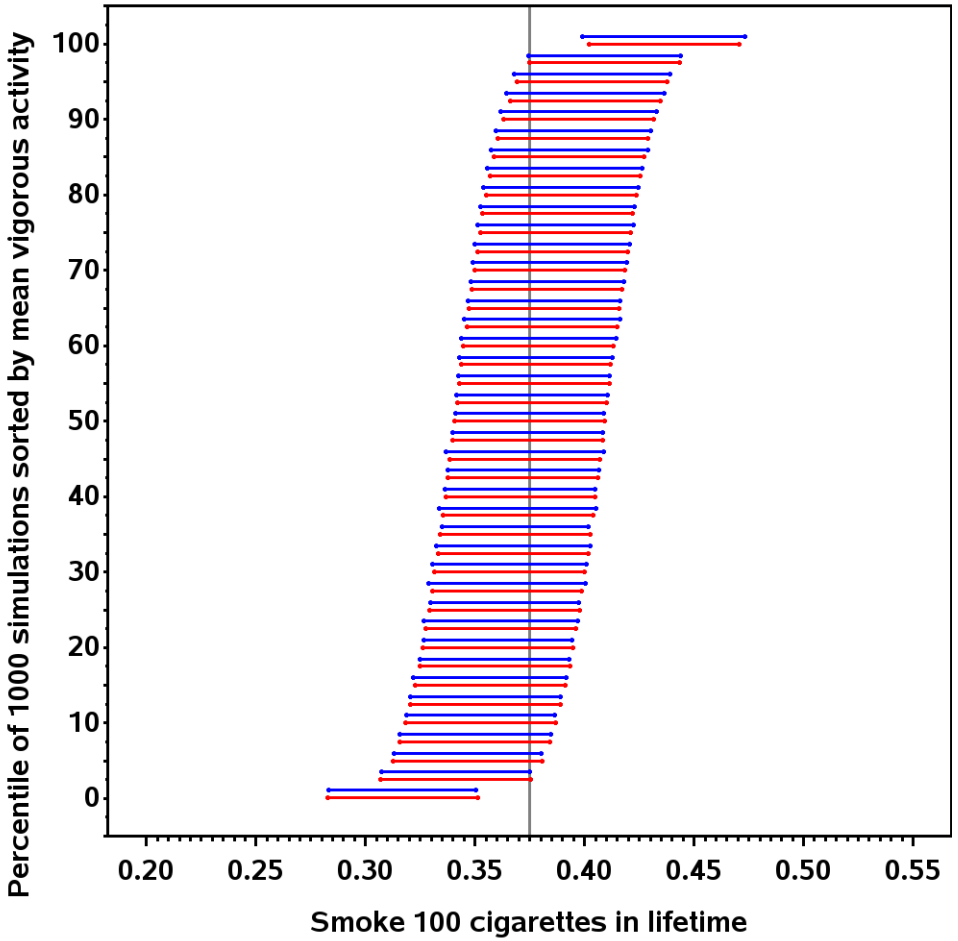
IVR Evaluation Survey *(continued)*

		Categorical		Continuous			
Health outcome	Category	Global	Local	Local	Local-delay	Global	NHIS
Hours of sleep	6 hours or less	35.0	33.4	**51.4**	**50.6**	**43.4**	**32.3**
	7 hours	33.4	38.1	15.7	14.1	15.4	30.3
	8 hours	22.6	20.9	25.2	26.2	27.6	29.7
	9 hours	4.8	4.3	3.6	4.5	4.4	4.1
	10 hours or more	4.2	3.2	4.2	4.6	9.2	3.6
Physical activity per week	**0 times**	**23.8**	**27.5**	**8.4**	**13.3**	**14.8**	**53.1**
	1 or 2 times	27.2	33.3	24.5	19.2	24.0	14.8
	3-5 times	34.7	28.7	40.3	39.9	40.8	21.3
	6 or 7 times	7.4	4.7	10.3	7.3	10.0	8.9
	8 or more times	6.9	5.9	16.6	20.3	10.4	1.9
Alcohol per month	0 times	51.1	52.5	46.7	45.0	48.5	46.2
	1 time	15.8	17.8	11.1	11.8	9.3	11.6
	2 or 3 times	15.6	10.1	13.9	15.5	17.9	8.4
	4 to 6 times	6.9	8.2	9.4	9.8	9.2	11.7
	7 or more times	10.6	11.3	18.8	17.9	15.1	22.0
Number of drinks	1	57.6	55.1	35.8	31.9	29.3	38.3
	2	21.3	23.3	25.9	20.4	31.0	32.8
	3 or 4	11.8	12.9	12.9	16.0	16.7	20.0
	5 or more	9.2	8.7	25.5	31.7	23.0	8.9
Mean absolute deviation		6.7	6.9	6.8	7.8	6.7	N/A

Margin of error needed for 95% of the simulated studies to contain the true value

Percent smoke 100 cigarettes in lifetime

1,000 simulated studies with 1,000 respondents



Blue: normal confidence interval

Red: confidence interval needed for 95% of the simulations to contain the true value

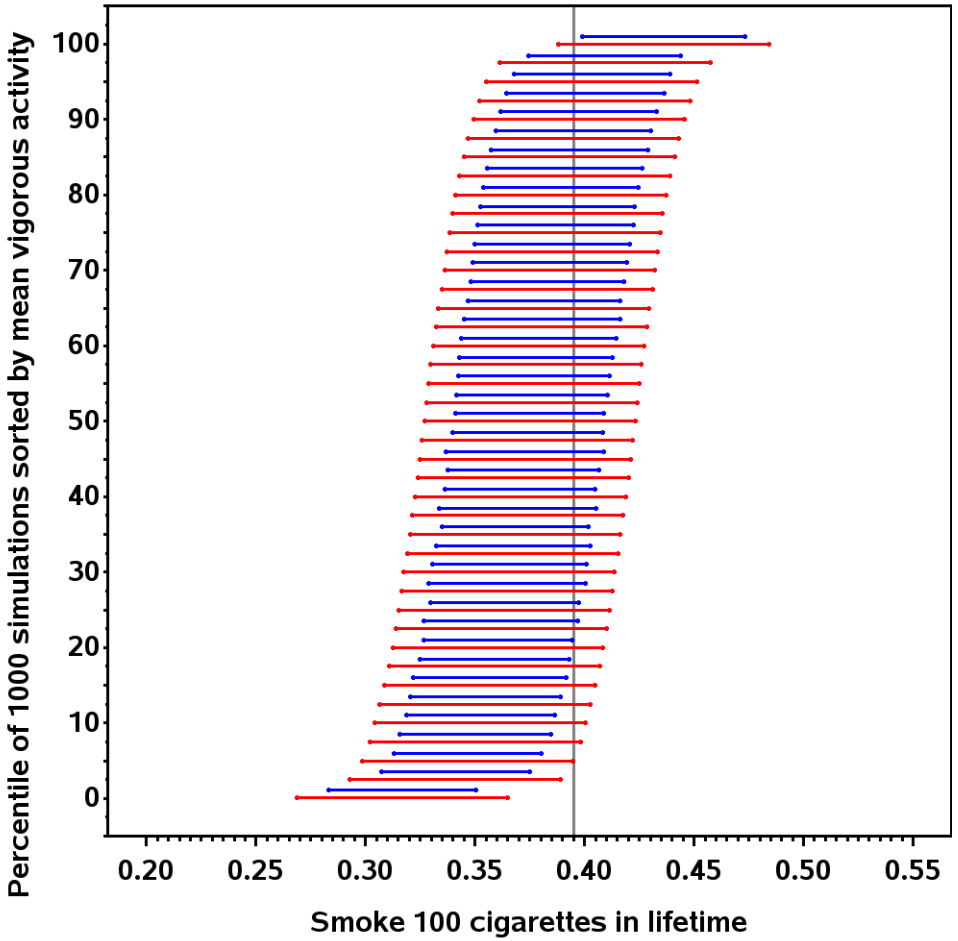
Bias 0%

Sample size	Margin of error		
	Mean MOE over 1000 simulations	Extra	Accounting for bias and variance
500	5.00	0.00	5.00
1,000	**3.49**	**-0.07**	**3.42**
2,000	2.45	-0.17	2.28
4,000	1.73	-0.13	1.60
6,000	1.41	-0.03	1.38
8,000	1.22	-0.08	1.14
10,000	1.09	-0.07	1.02

Margin of error needed for 95% of the simulated studies to contain the true value

Percent smoke 100 cigarettes in lifetime

1,000 simulated studies with 1,000 respondents



Blue: normal confidence interval

Red: confidence interval needed for 95% of the simulations to contain the true value

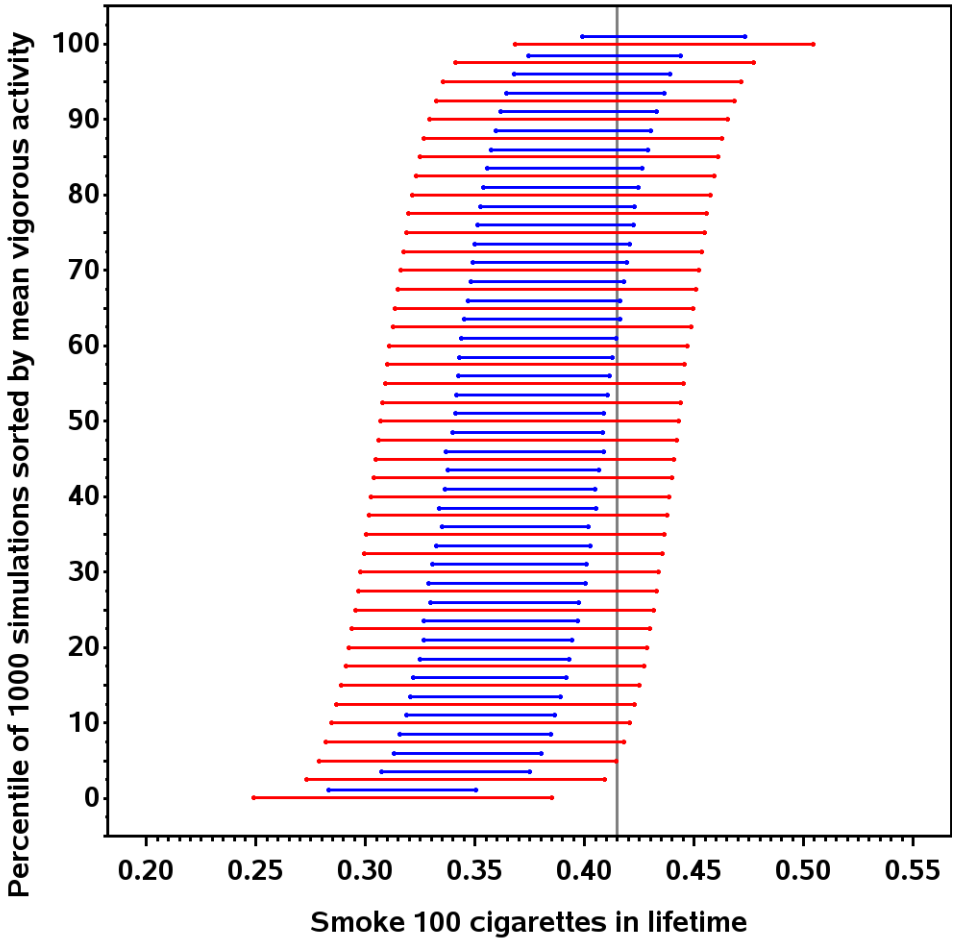
Bias: 2%

Sample size	Margin of error		
	Mean MOE over 1000 simulations	Extra	Accounting for bias and variance
500	5.00	1.00	6.00
1,000	**3.49**	**1.31**	**4.80**
2,000	2.45	1.53	3.98
4,000	1.73	1.61	3.34
6,000	1.41	1.80	3.21
8,000	1.22	1.79	3.01
10,000	1.09	1.79	2.88

Margin of error needed for 95% of the simulated studies to contain the true value

Percent smoke 100 cigarettes in lifetime

1,000 simulated studies with 1,000 respondents



Blue: normal confidence interval

Red: confidence interval needed for 95% of the simulations to contain the true value

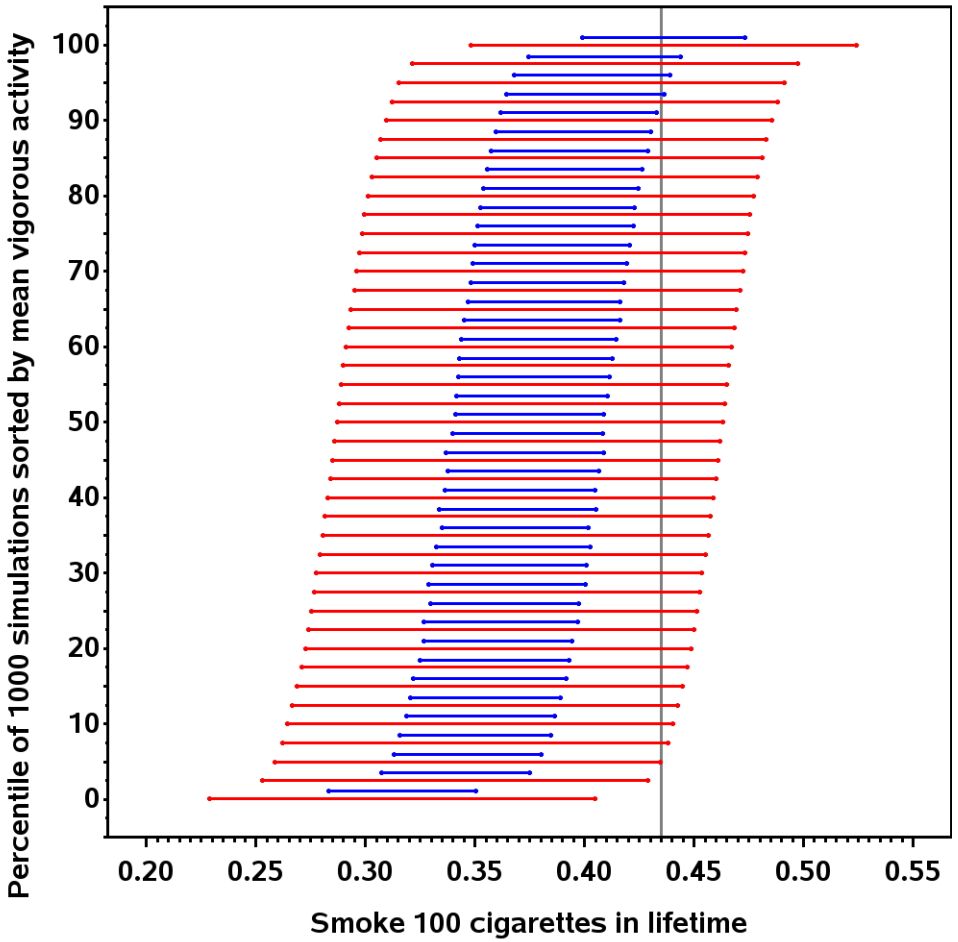
Bias: 4%

Sample size	Margin of error		
	Mean MOE over 1000 simulations	Extra	Accounting for bias and variance
500	5.00	2.98	7.98
1,000	**3.49**	**3.31**	**6.80**
2,000	2.45	3.53	5.98
4,000	1.73	3.62	5.35
6,000	1.41	3.77	5.18
8,000	1.22	3.79	5.01
10,000	1.09	3.80	4.89

Margin of error needed for 95% of the simulated studies to contain the true value

Percent smoke 100 cigarettes in lifetime

1,000 simulated studies with 1,000 respondents



Blue: normal confidence interval

Red: confidence interval needed for 95% of the simulations to contain the true value

Bias: 6%

Sample size	Margin of error		
	Mean MOE over 1000 simulations	Extra	Accounting for bias and variance
500	5.00	4.98	9.98
1,000	**3.49**	**5.31**	**8.80**
2,000	2.45	5.51	7.96
4,000	1.73	5.62	7.35
6,000	1.41	5.76	7.17
8,000	1.22	5.79	7.01
10,000	1.09	5.80	6.89

Outstanding issues

- Develop a strategy to use incentives successfully and cost effectively
- How best to use the IVR system
  - Voice activated data collection
- Test other modes
  - Recruit to a web-instrument
  - Send to live interviewer (CATI)
- Developing and evaluating different weighting approaches
- Calculating variance estimates in nonprobability samples is still a matter of debate
  - Resampling approach—Bootstrap or jackknife
  - Bayesian credible interval
  - SRS formula with or without adjustment for the approximate design effect
- Evaluate the stability of the estimates in a repeated cross-sectional studies

delivering **the promise of science**
for global good

Burton Levine
blevine@rti.org
(919) 541-1252