

Investigating Internet Opt-in Panels for Behavioral Surveillance

Carol A. Gotway Crawford, PhD

Chief, Population Health Surveillance Branch

Catherine Okoro, PhD

Epidemiologist

Satvinder Dhingra, MPH

Epidemiologist
Northrop Grumman

Federal Committee on Statistical Methodology Research Conference

November 4, 2013

Washington, DC

National Center for Chronic Disease Prevention and Health Promotion

Division of Population Health



What Are Internet Opt-in Panels?

- ❑ **Potential panelists are recruited via the Internet**
 - Banner ads, email lists, promotions, and offers
 - Double opt-in process to become a panel member
- ❑ **Panelists become the pool for sample selection**
- ❑ **Panel may or may not be representative of the population**
 - Coverage is limited to Internet users (~ 80% of the population)
 - Respondent selection and motivation

Why Use Internet Opt-in Panels?

- Lower cost than probability-based sampling
- Shorter collection and prep time for data release to the public than current methods (RDD, face-to-face)
- Expands the surveillance and study tool-kit
- Permits longitudinal and in-depth follow-up studies
- Increases administrative and design flexibility and efficiency

Pilot Study

- **4 States**
 - Cooperative agreements in GA, IL, NY, and TX
- **3 Vendors**
 - Different sampling methodologies
 - Cooperating and collaborating
 - De-duplication of respondents
 - Nearly identical questionnaire format
- **3 Levels of Geography**
 - National
 - State
 - Metropolitan Statistical Area

Pilot Objectives

- **Compare sampling methodologies**
 - Sample matching, source blending, and quota
- **Assess feasibility and accuracy for public health**
- **Compare estimates with those from other surveys**
- **Evaluate across a range of parameters:**
 - Cost, geographic granularity, and timeliness

Sampling Methodologies

■ Sample Matching

- Different modes of recruitment are used to ensure representativeness for hard-to-reach populations
- Potential respondents are selected by matching to a random sample from the American Community Survey
- Final responses are weighted to known characteristics in the U.S. using propensity score weighting

■ Sample Blending

- Uses population segments designed to reflect behavioral differences but based on Census data
- Apply the segmentation structure locally to balance, weight, and blend sample

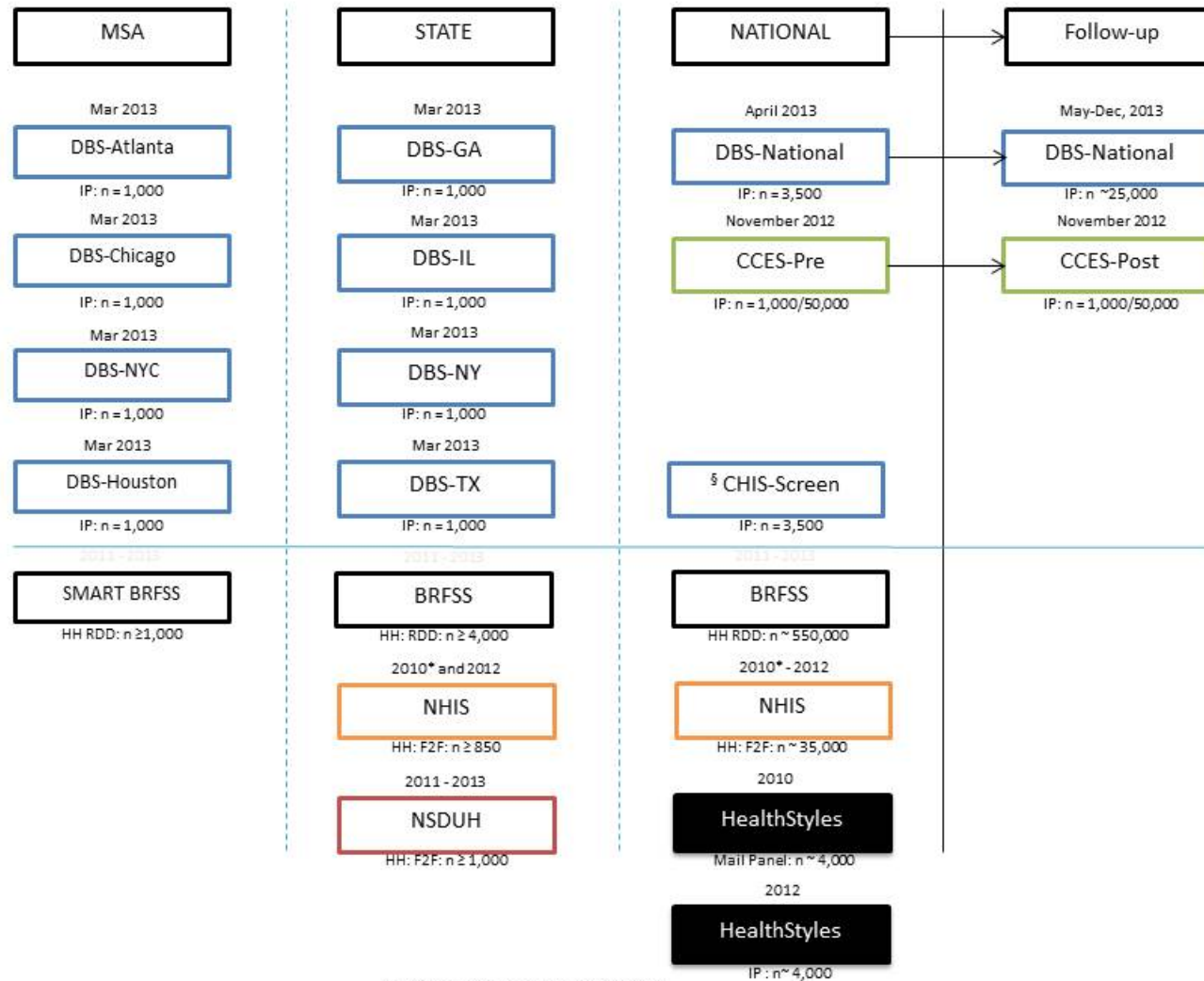
■ Quota Sampling

- A non-probability sample in which respondents take the survey on a first-come, first-served basis according to a fixed quota

Questionnaire Development

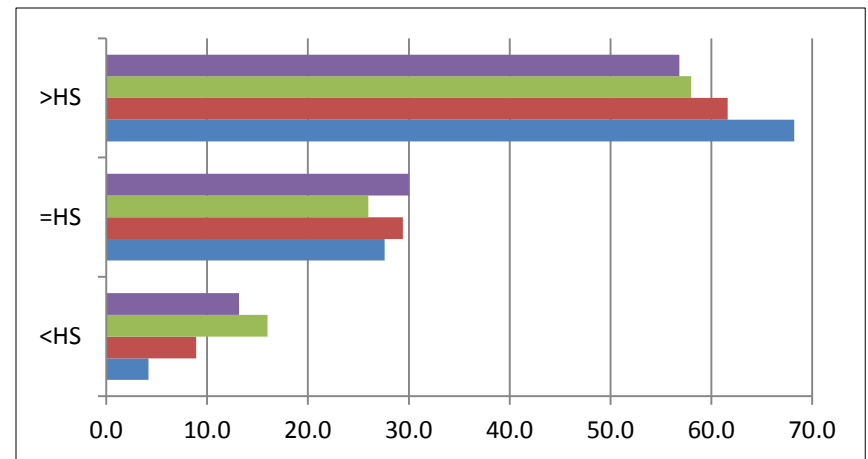
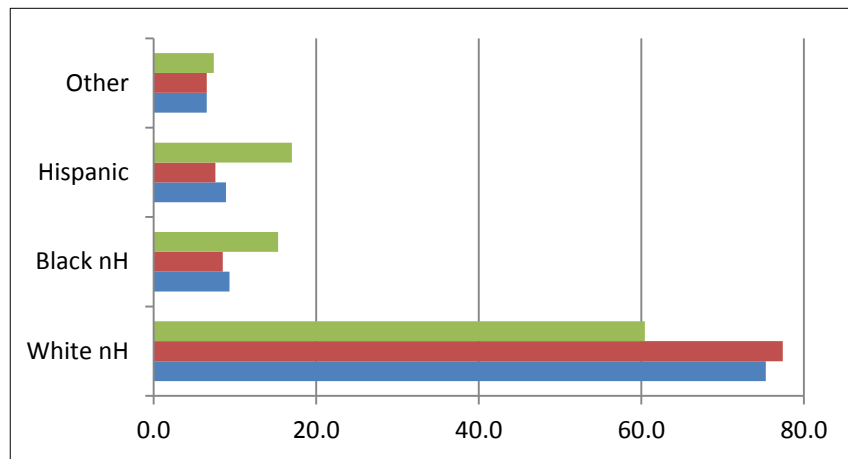
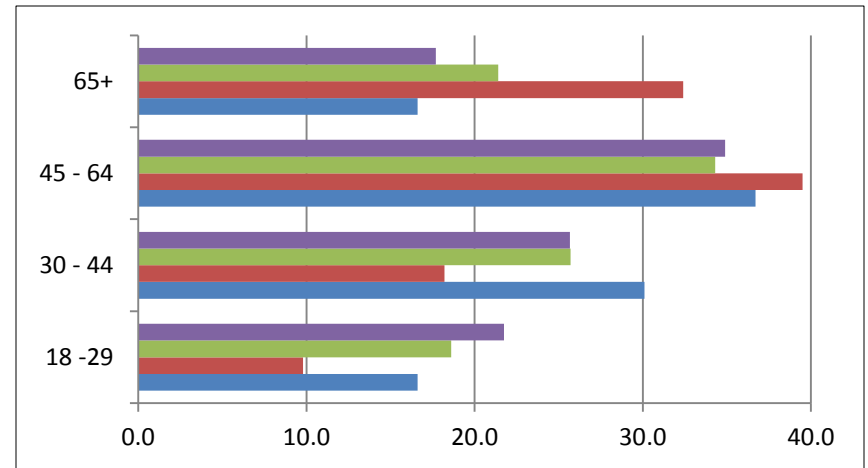
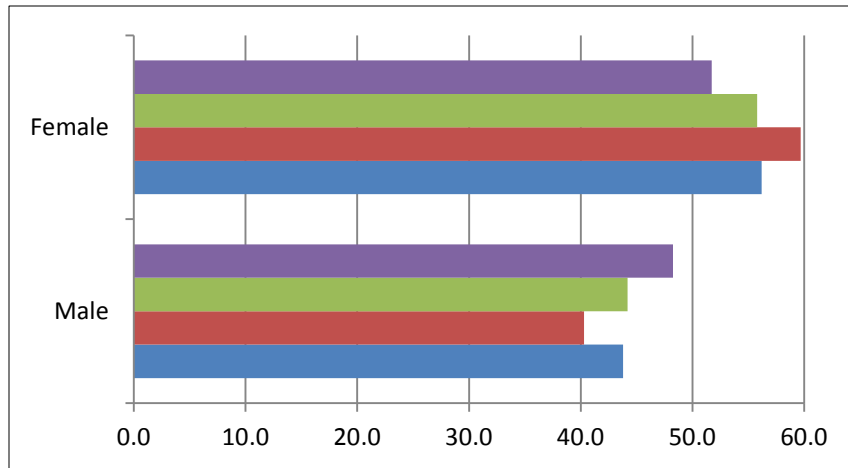
- **Survey consists of ~80 questions (20 minutes)**
- **Questions drawn from:**
 - CDC: BRFSS, NHANES, & NHIS
 - NIH: PROMIS
 - SAMHSA: NSDUH
 - ONC: Consumer Survey of Attitudes Toward the Privacy and Security Aspects of EHR and HIE
 - NPWF (National Partnership for Women and Families)
 - NSF supported Cooperative Congressional Election Study

Benchmarking



*PROMIS-10 questions: NHIS 2010

National: Demographics (Unweighted)



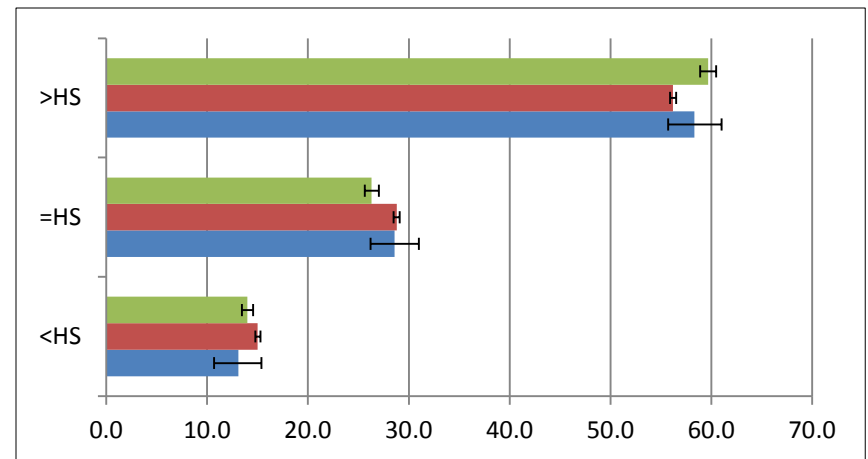
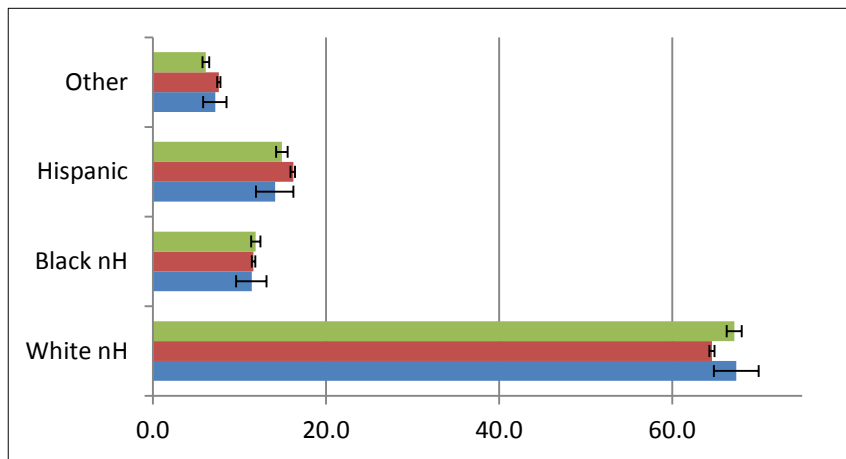
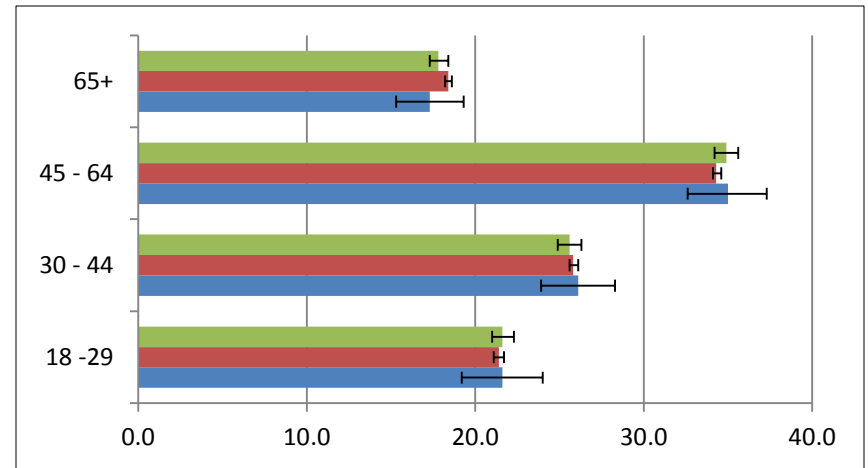
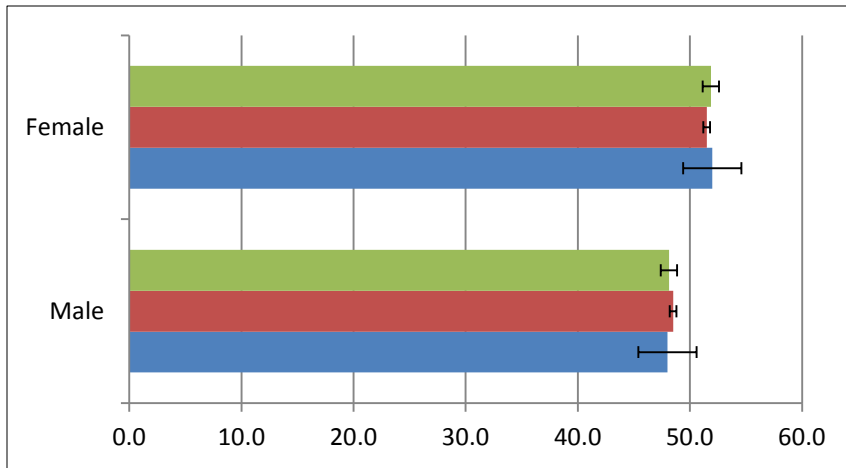
■ YouGov 2013
IPS Matched

■ BRFSS 2012
DF-RDD CATI

■ NHIS 2012
HH CAPI

■ CPS 2012

National: Demographics (Weighted)

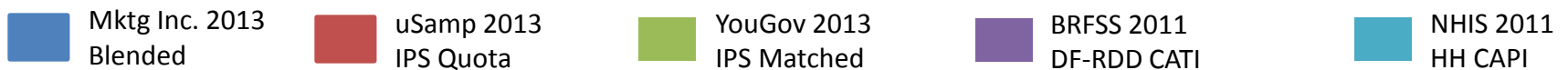
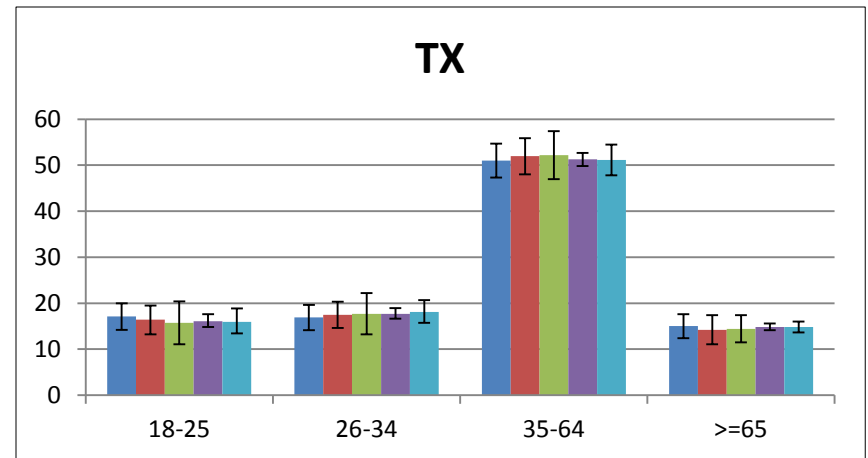
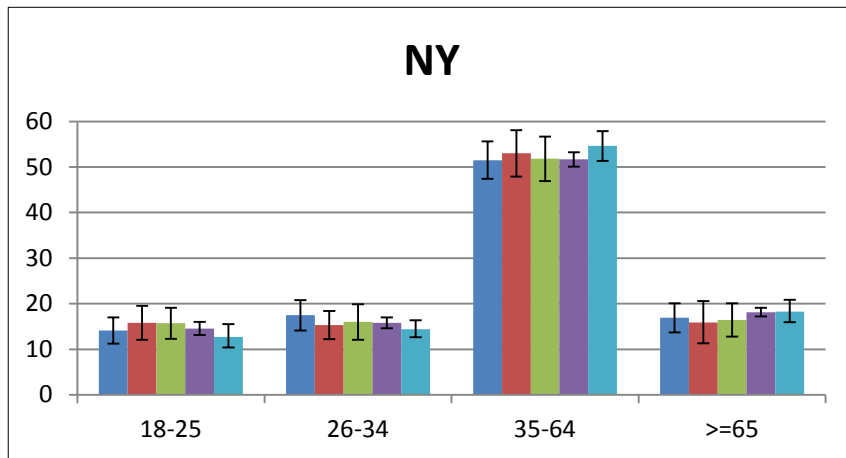
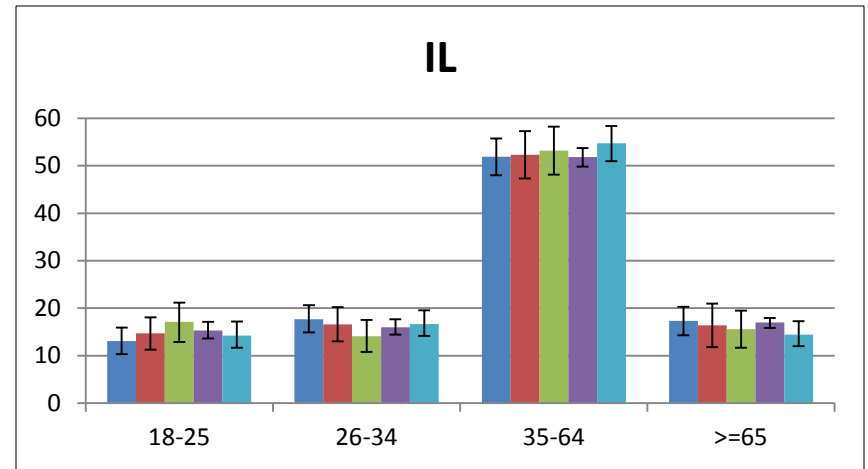
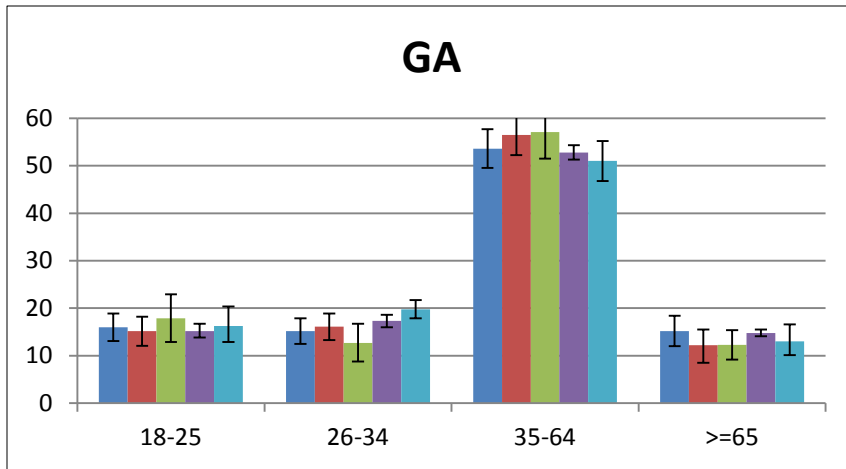


YouGov 2013
IPS Matched

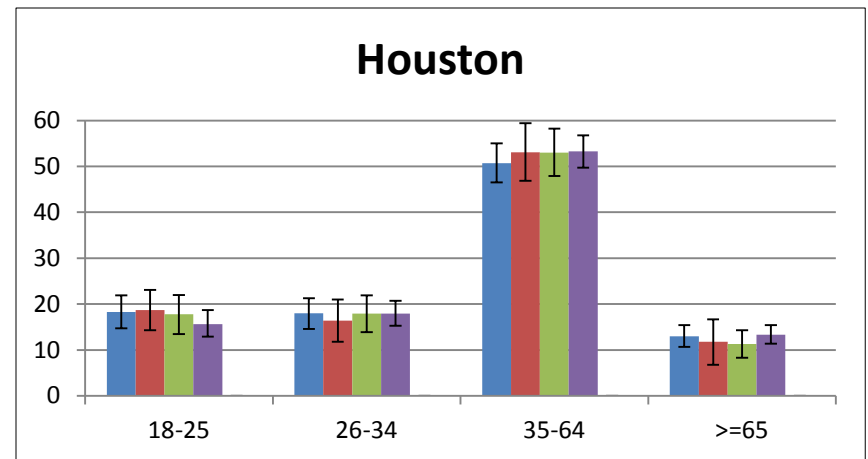
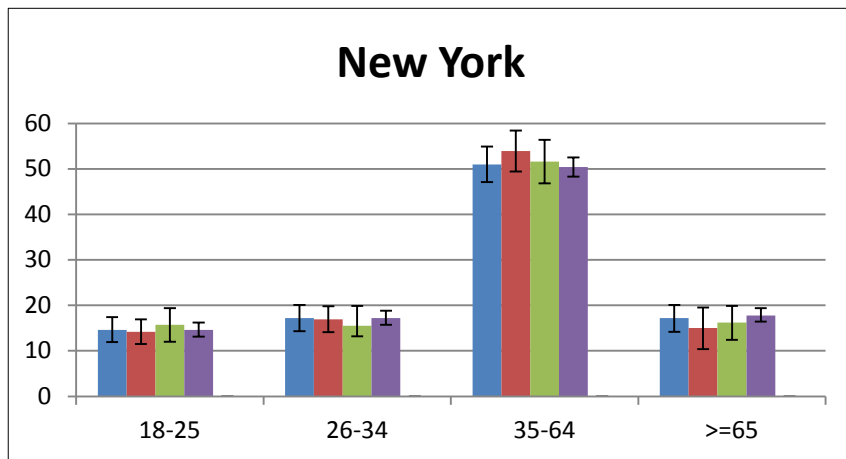
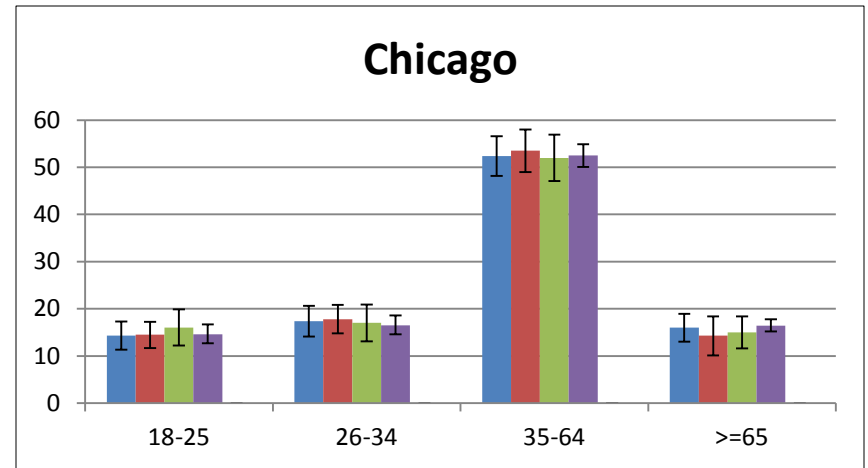
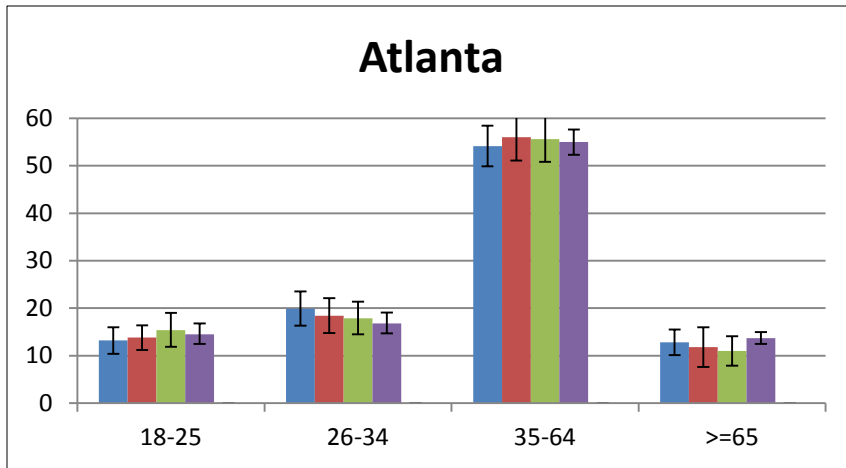
BRFSS 2012
DF-RDD CATI

NHIS 2012
HH CAPI

State: Age



MSA: Age



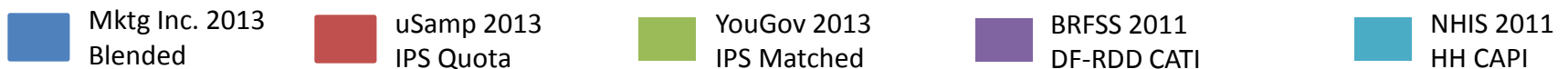
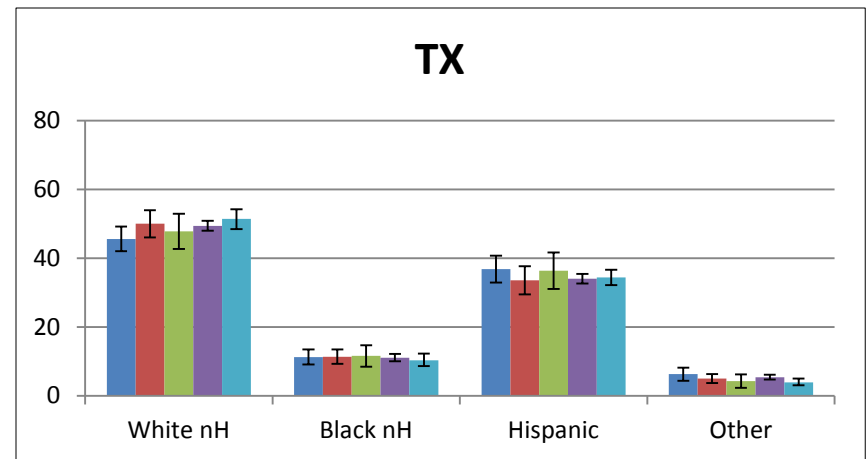
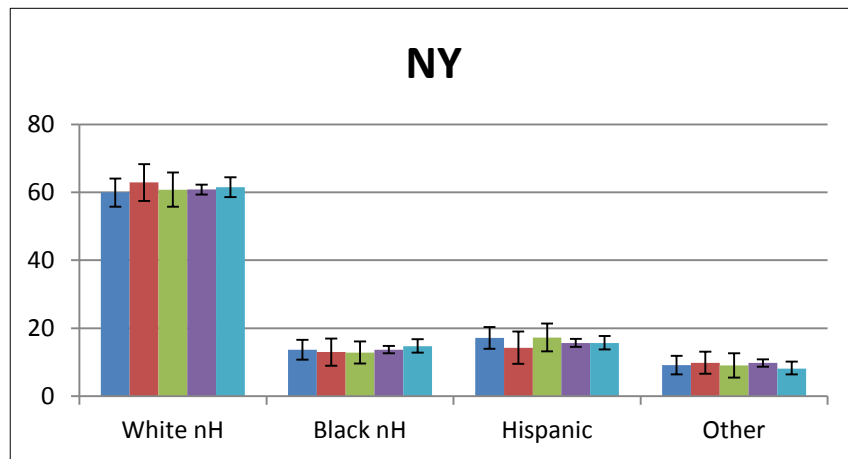
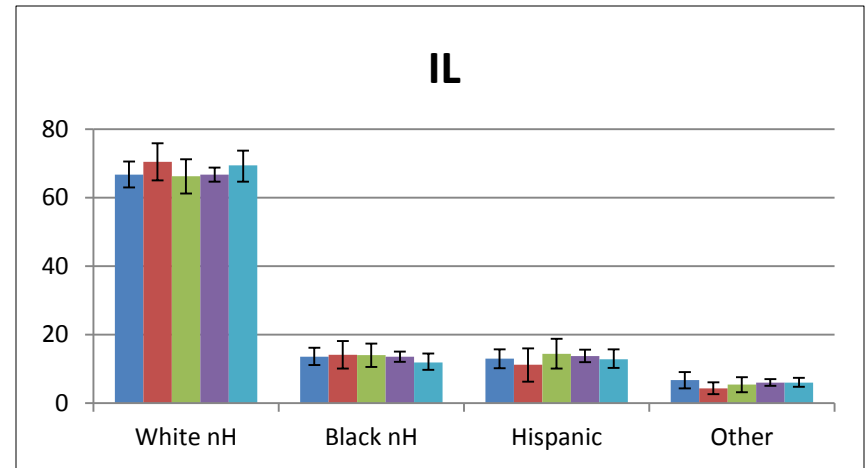
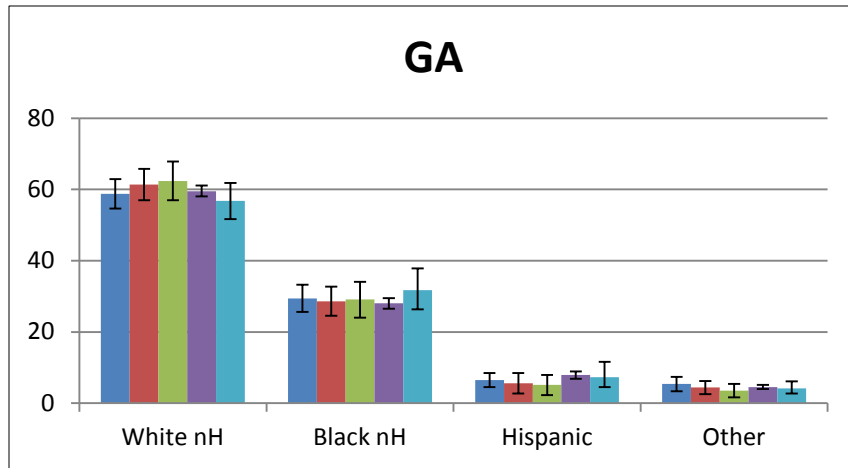
Mktg Inc. 2013
Blended

uSamp 2013
IPS Quota

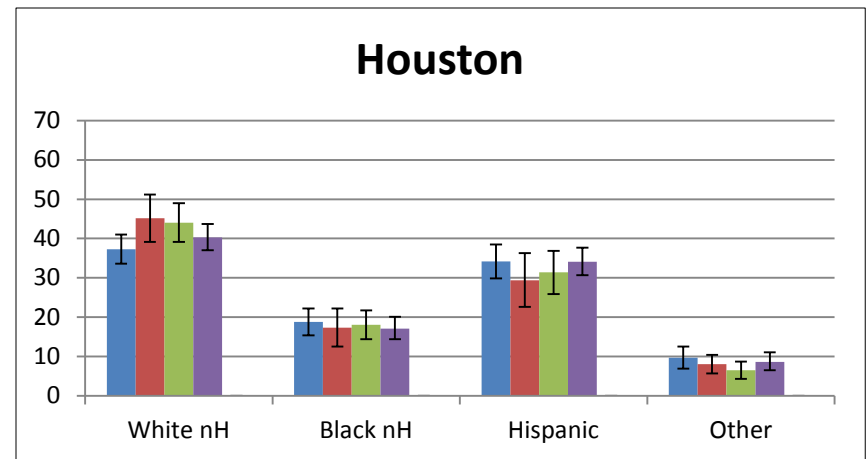
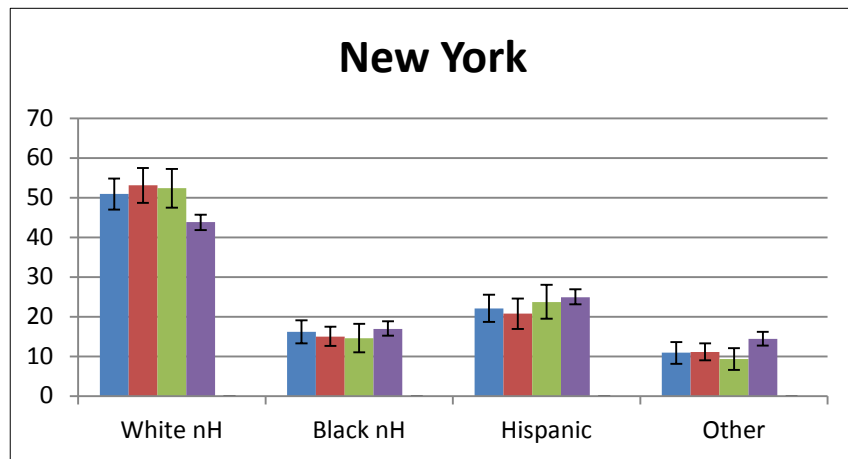
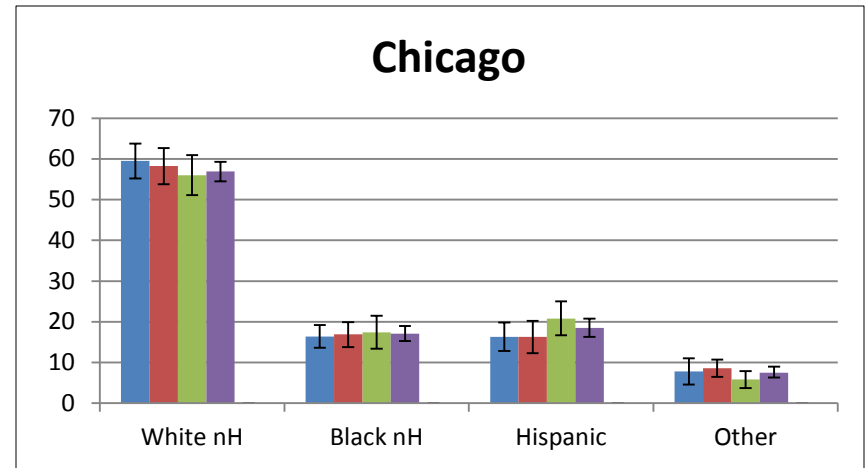
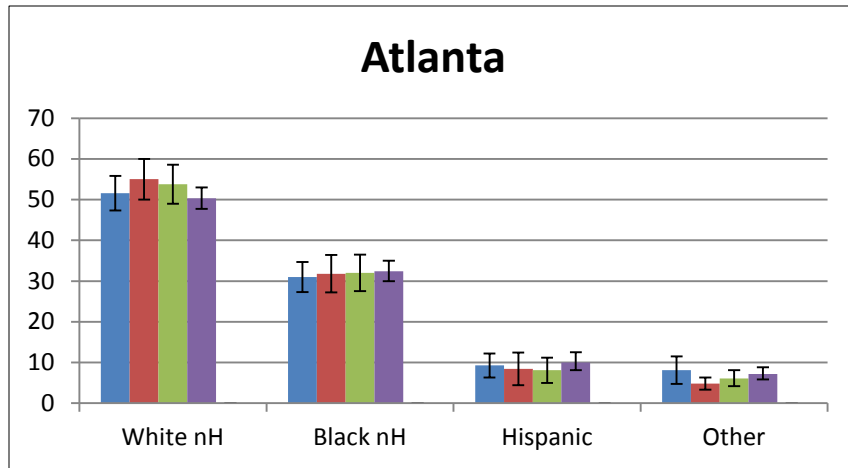
YouGov 2013
IPS Matched

SMART BRFSS 2012
DF-RDD CATI

State: Race/Ethnicity



MSA: Race/Ethnicity



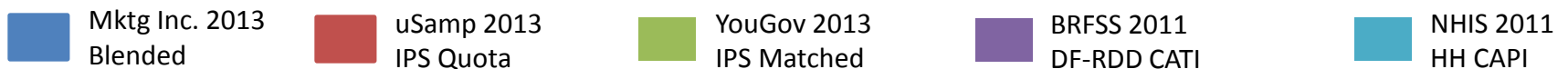
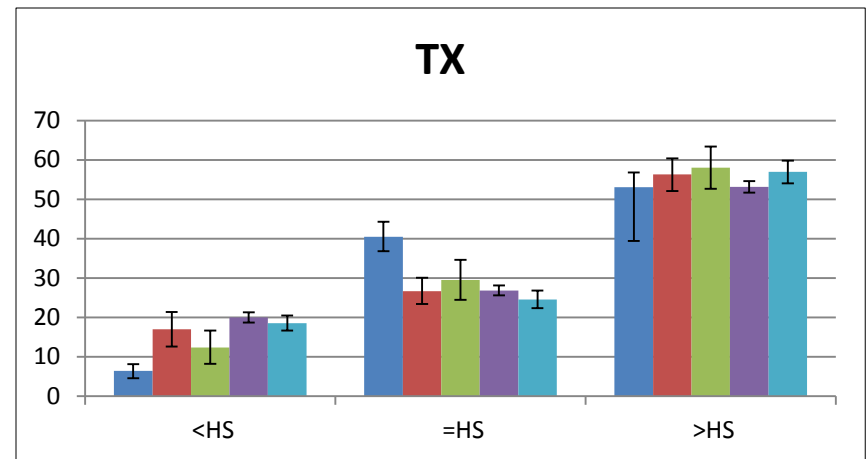
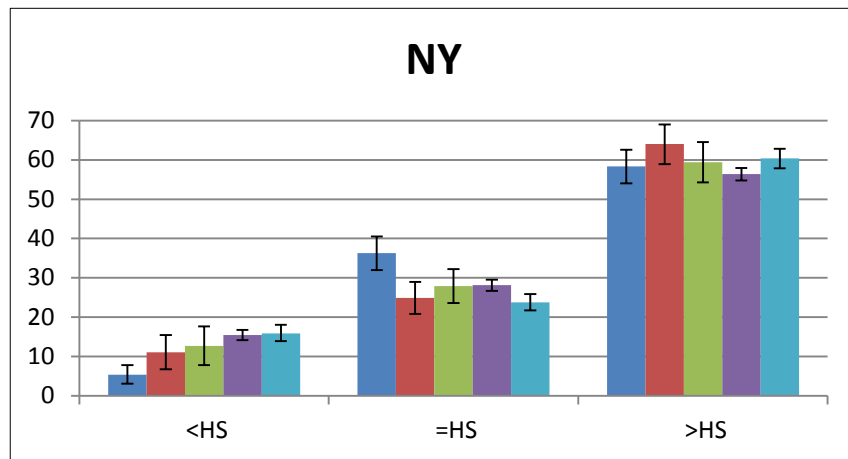
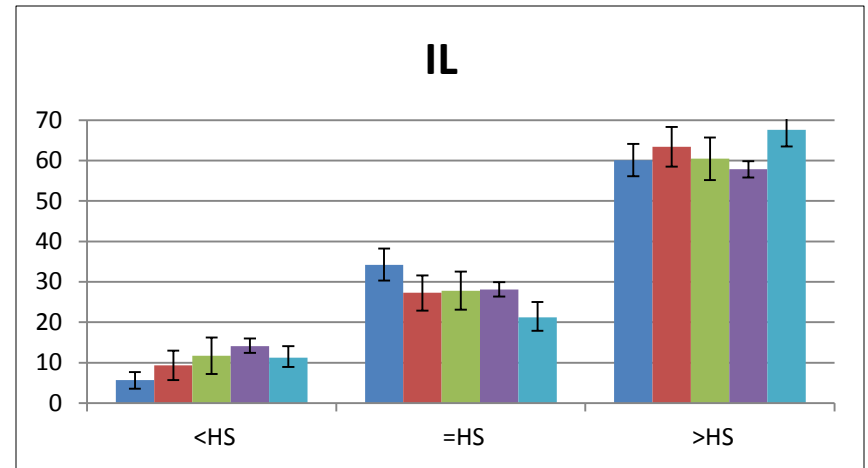
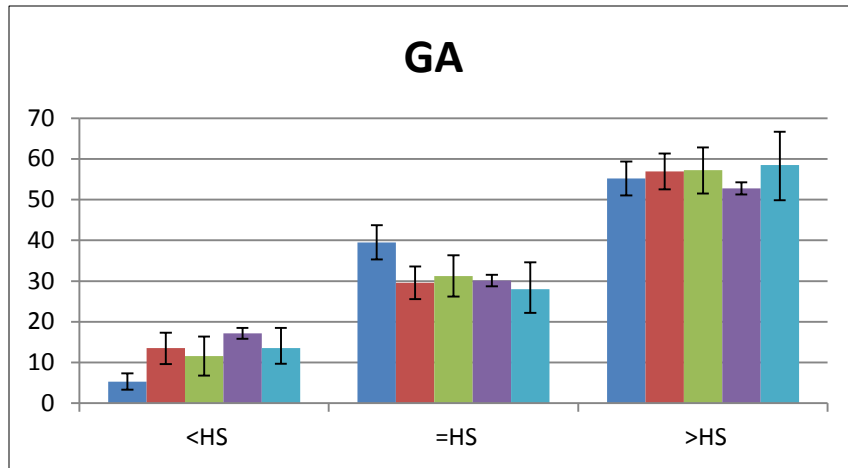
Mktg Inc. 2013
Blended

uSamp 2013
IPS Quota

YouGov 2013
IPS Matched

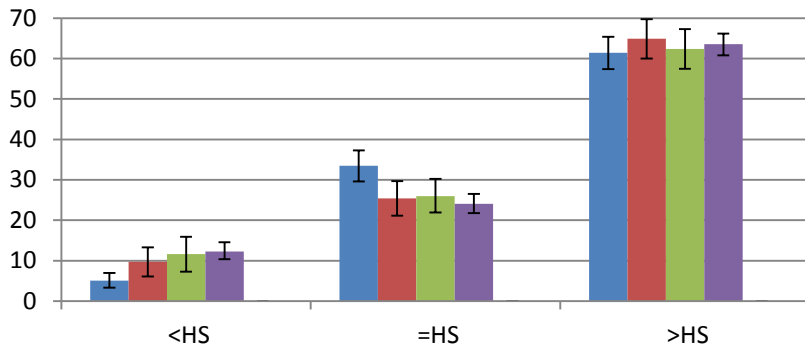
SMART BRFSS 2012
DF-RDD CATI

State: Education

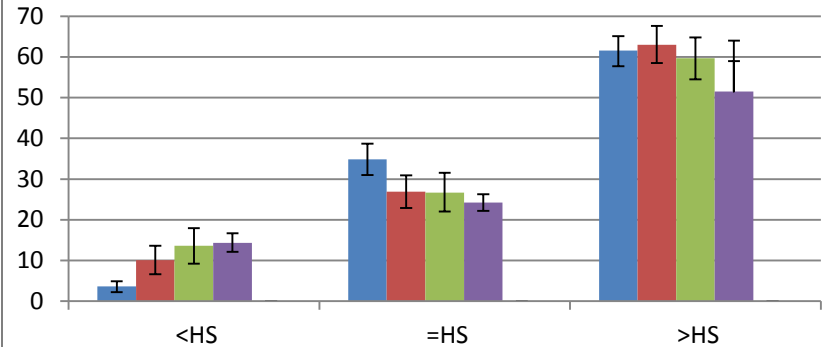


MSA: Education

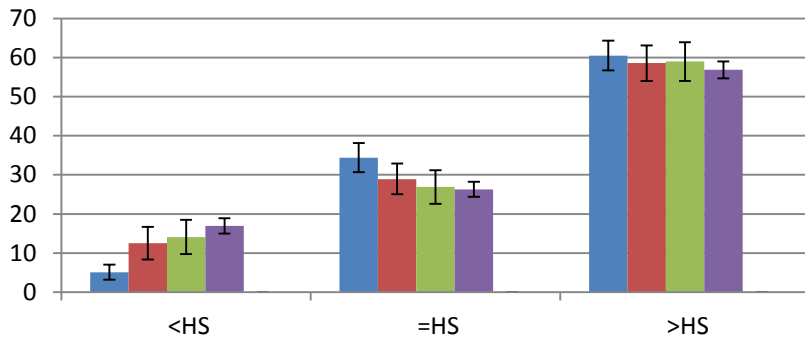
Atlanta



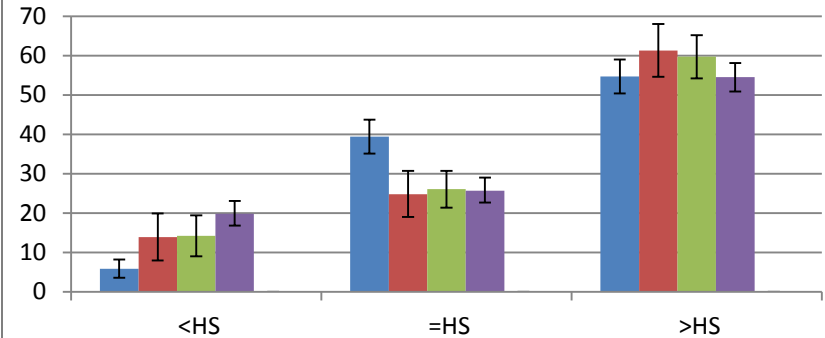
Chicago



New York



Houston



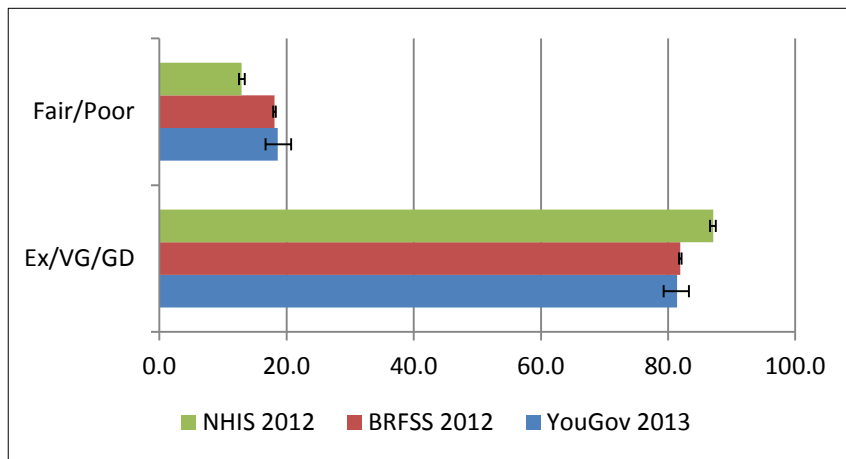
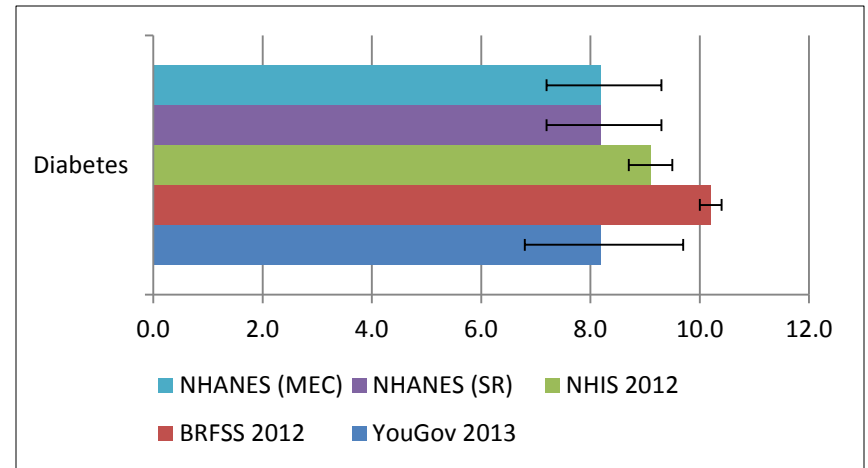
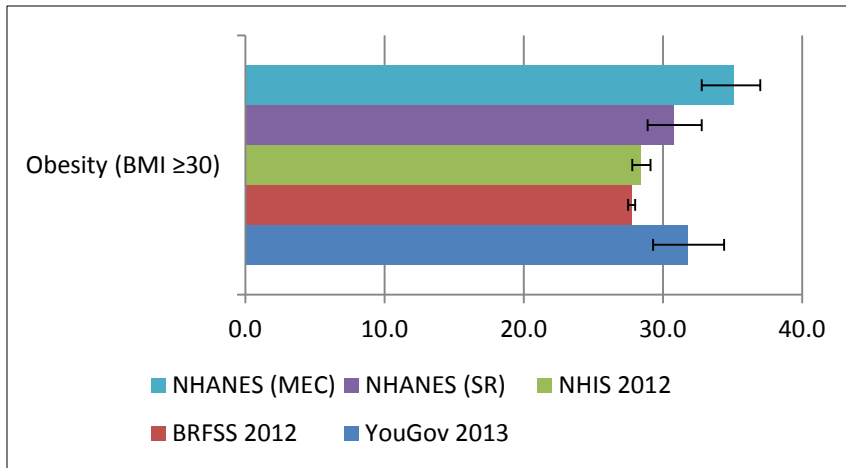
Mktg Inc. 2013
Blended

uSamp 2013
IPS Quota

YouGov 2013
IPS Matched

SMART BRFSS 2012
DF-RDD CATI

National: Outcomes

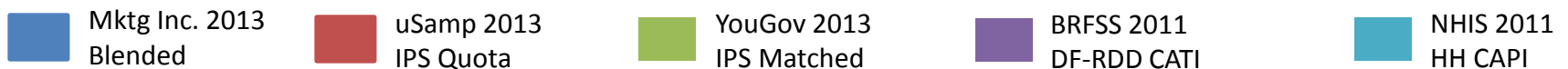
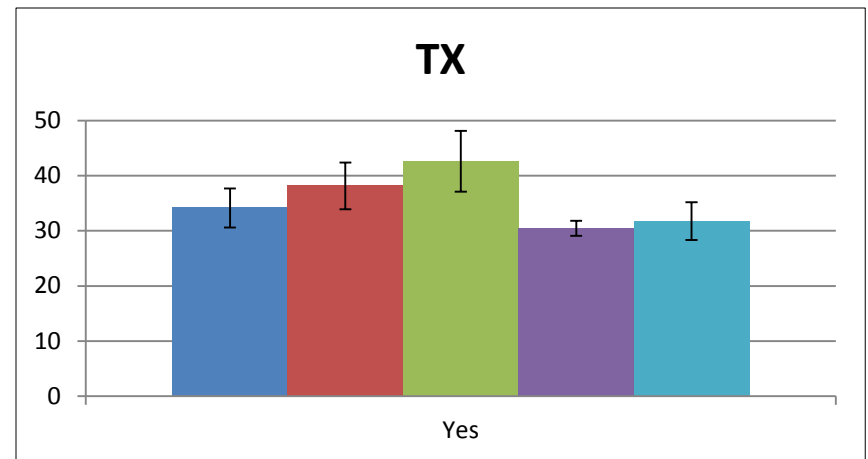
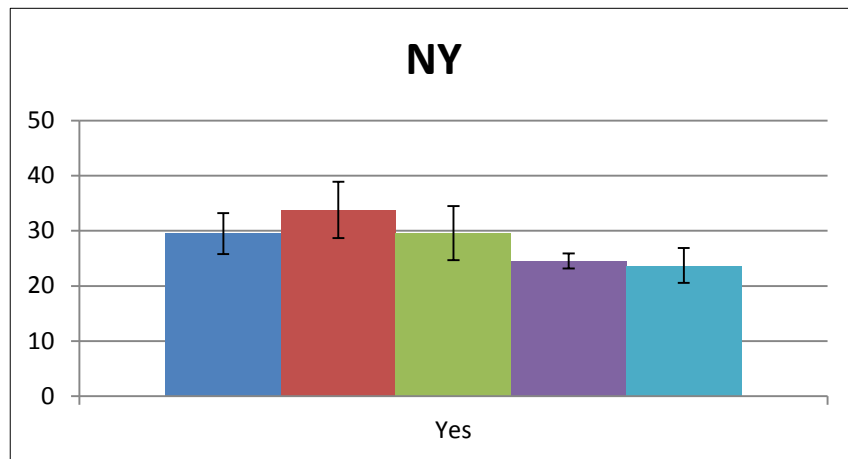
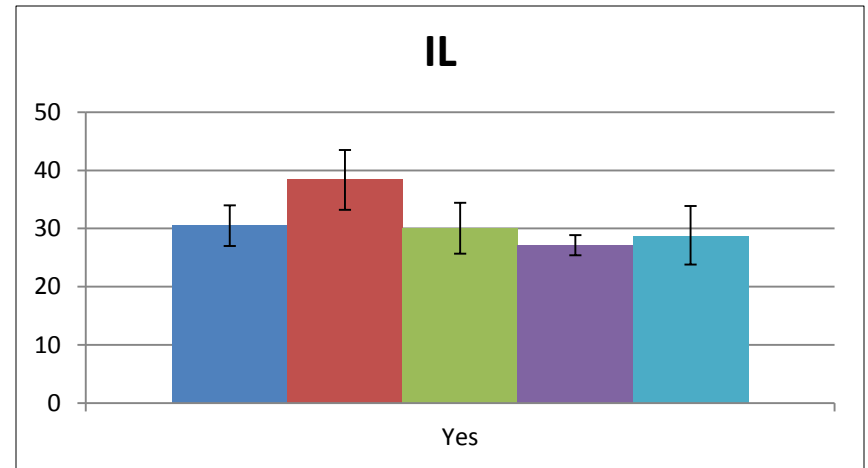
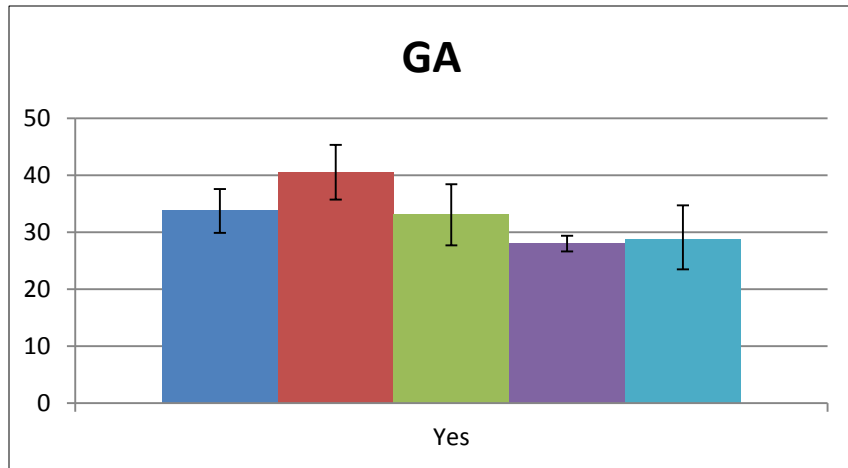


YouGov 2013
IPS Matched

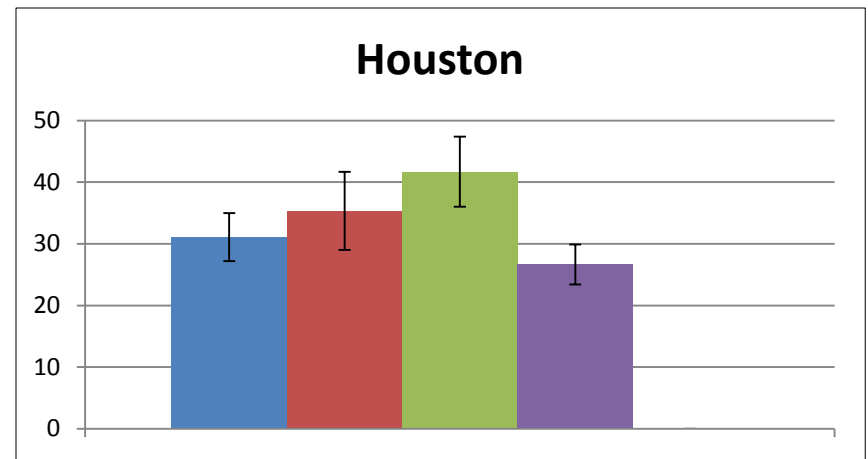
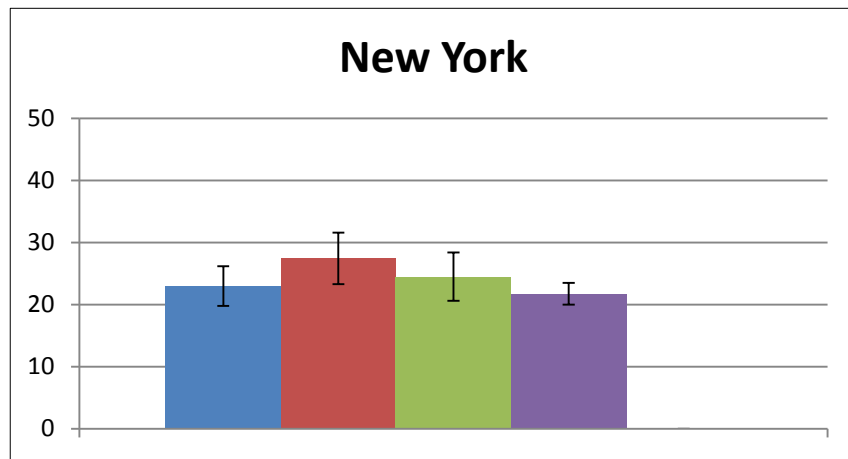
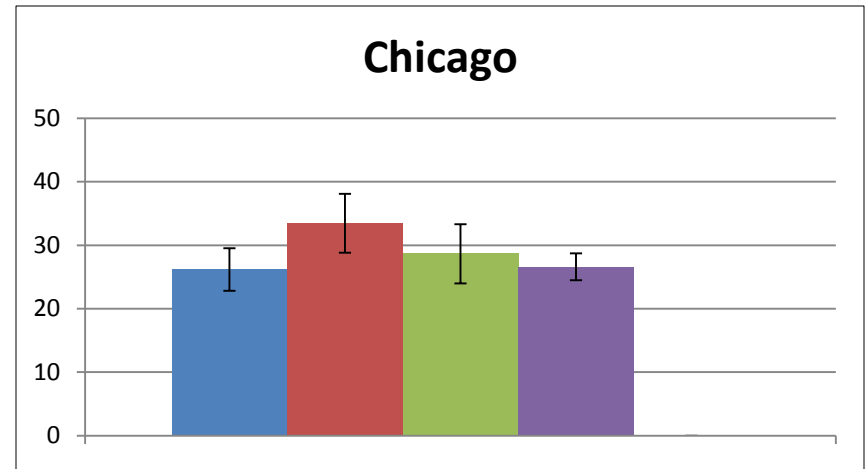
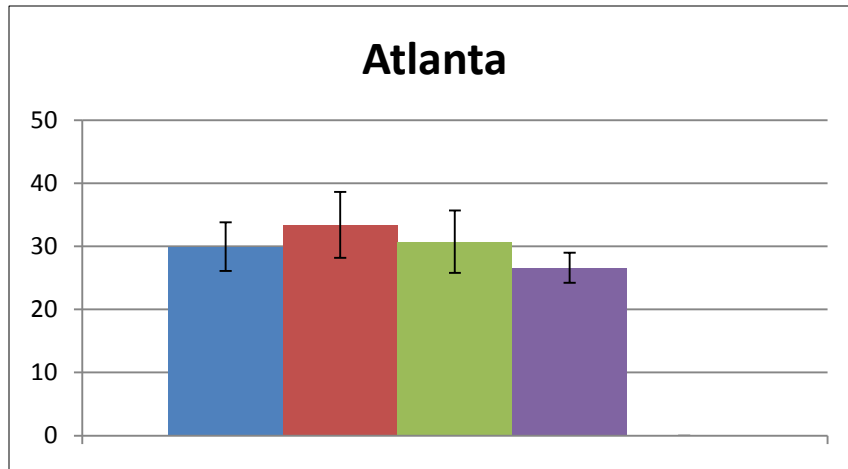
BRFSS 2012
DF-RDD CATI

NHIS 2012
HH CAPI

State: Obesity (BMI ≥ 30)



MSA: Obesity (BMI ≥ 30)



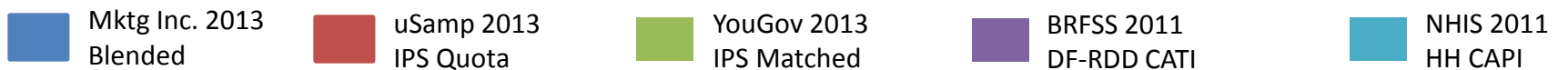
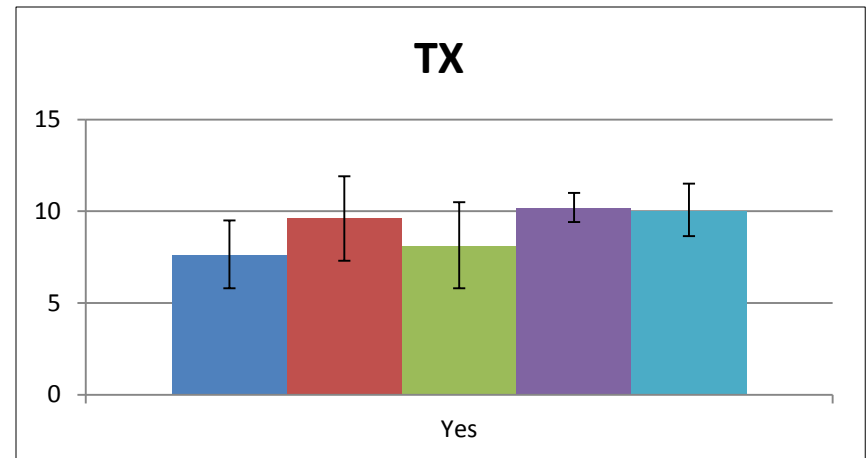
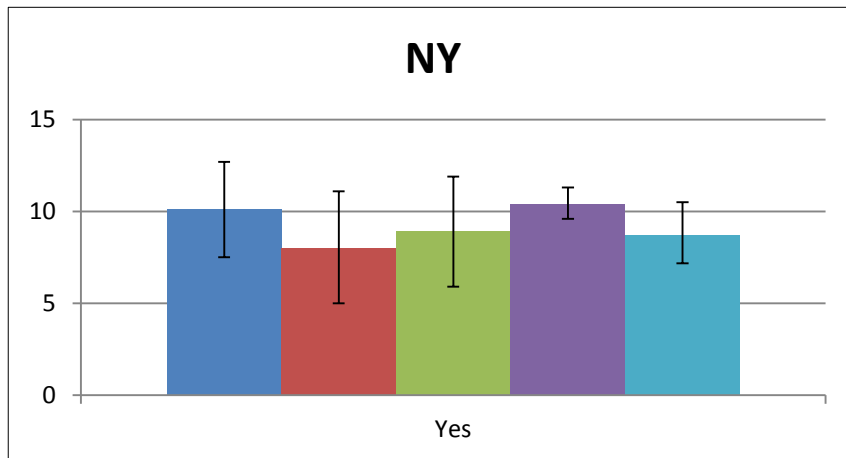
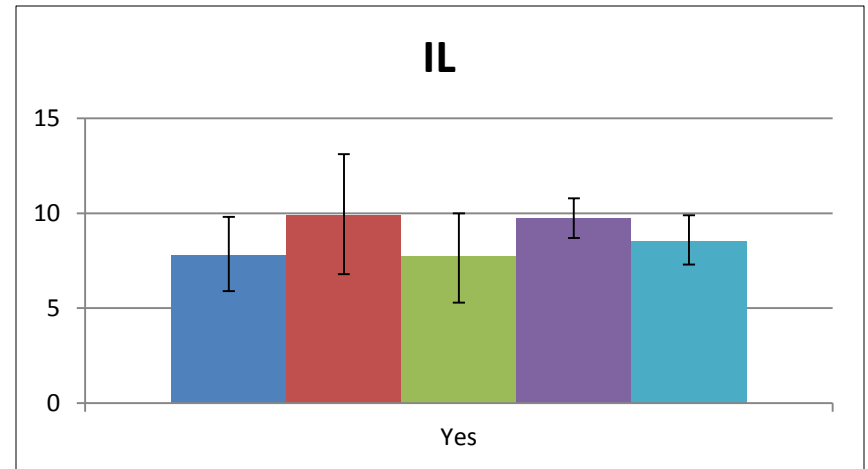
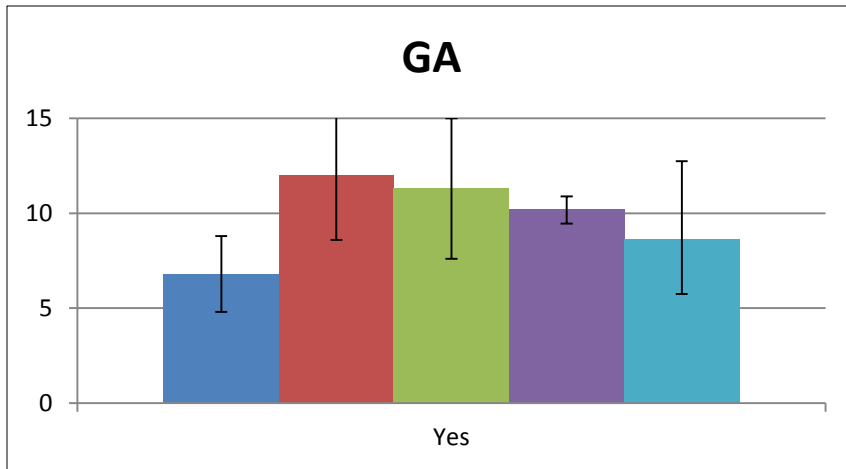
Mktg Inc. 2013
Blended

uSamp 2013
IPS Quota

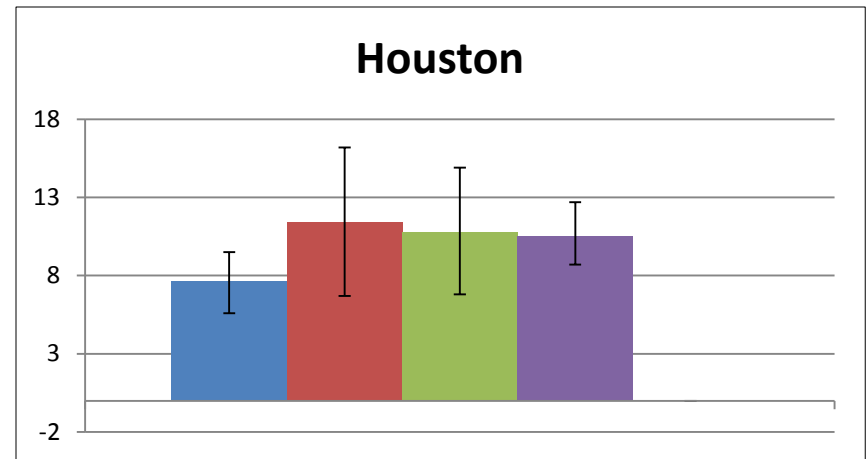
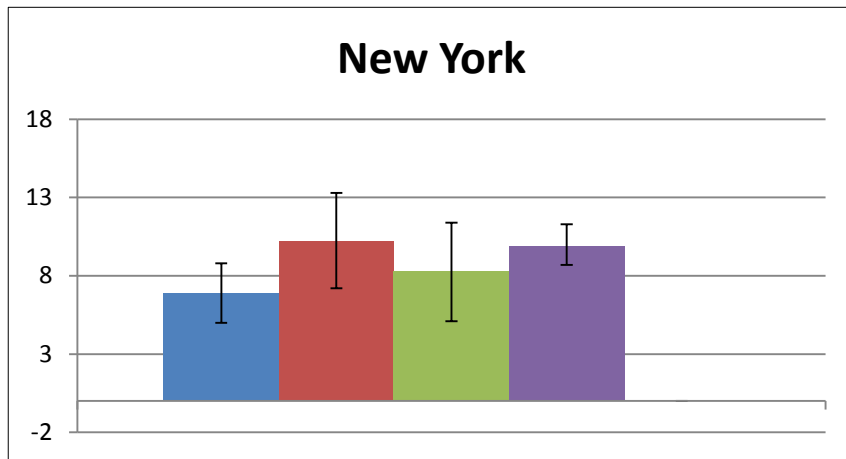
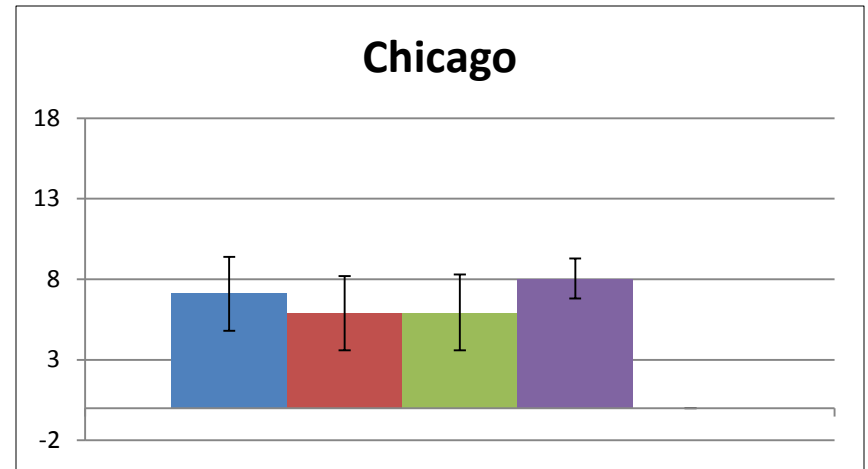
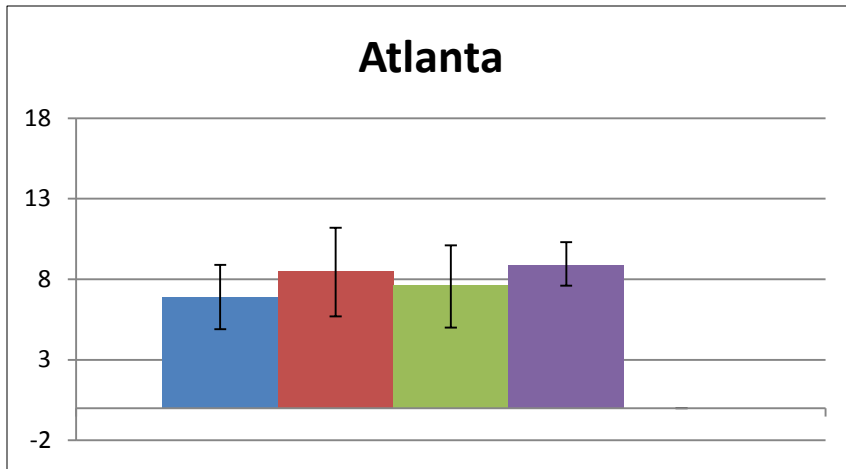
YouGov 2013
IPS Matched

SMART BRFSS 2012
DF-RDD CATI

State: Diabetes



MSA: Diabetes



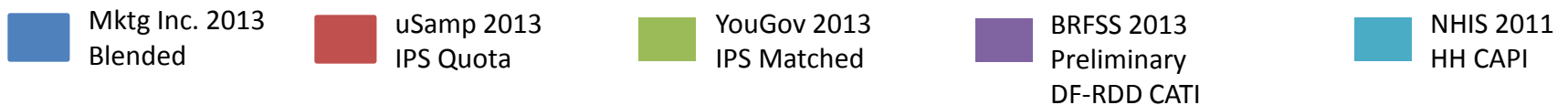
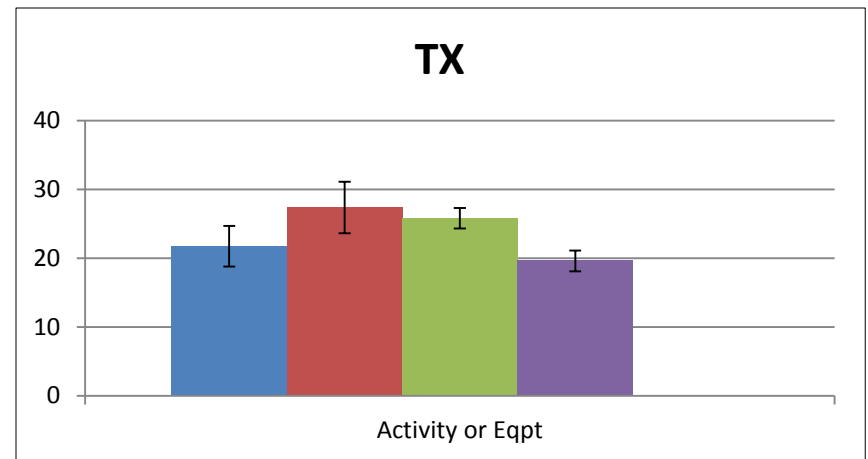
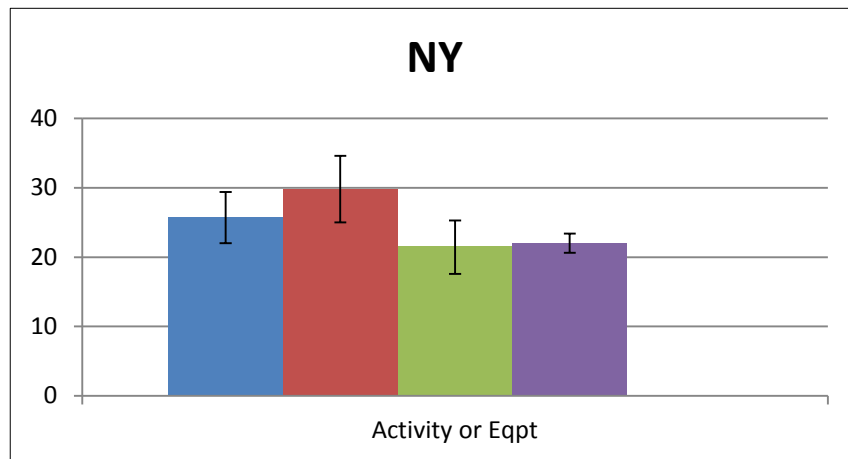
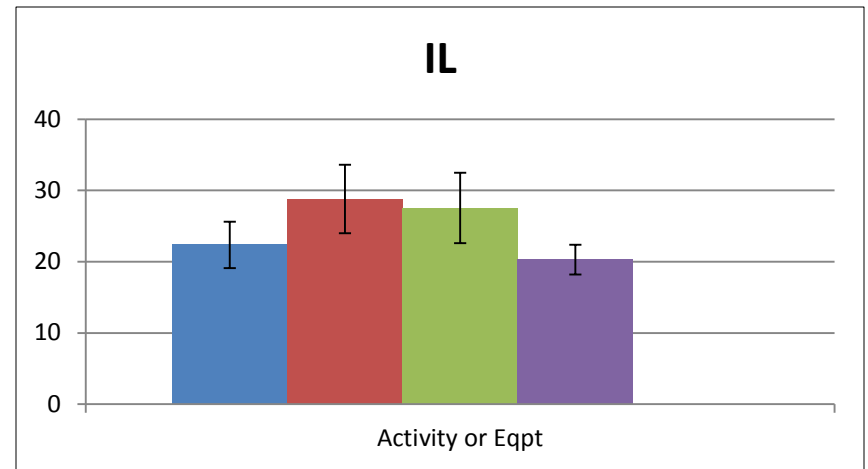
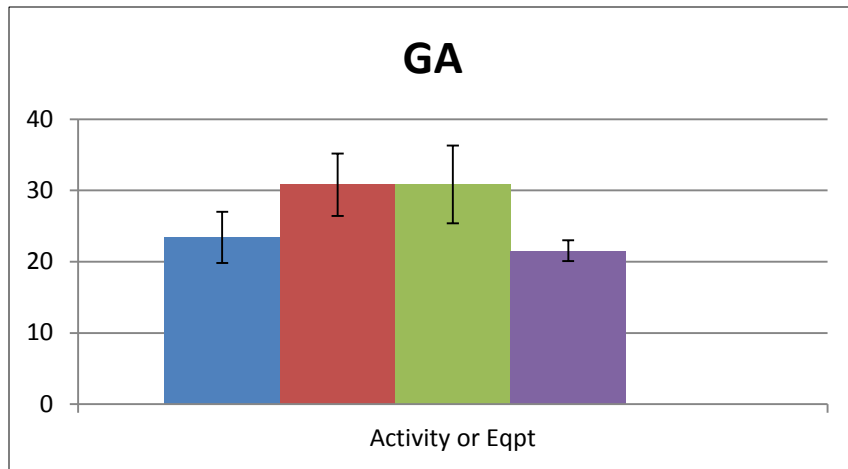
Mktg Inc. 2013
Blended

uSamp 2013
IPS Quota

YouGov 2013
IPS Matched

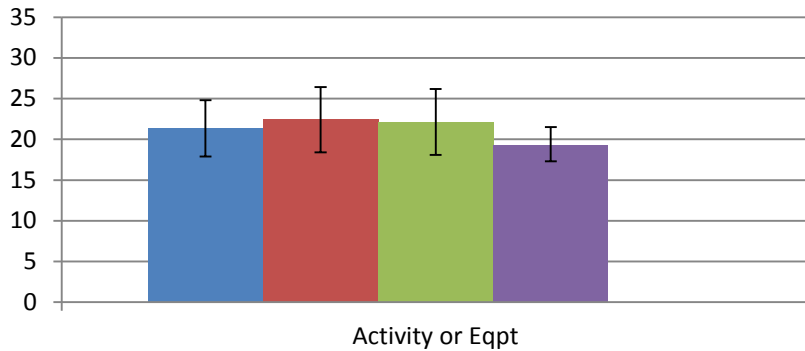
SMART BRFSS 2012
DF-RDD CATI

State: Disability

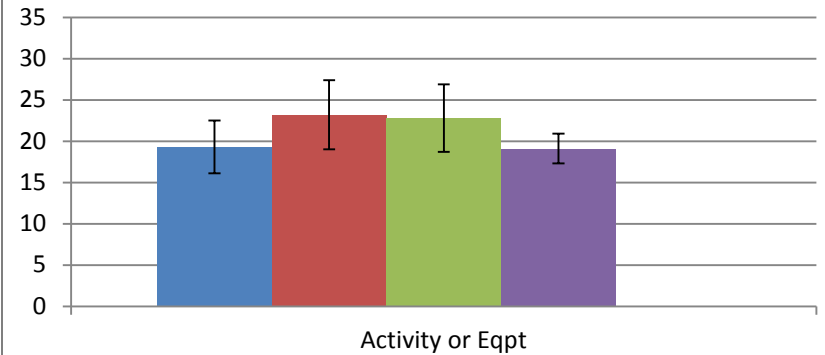


MSA: Disability

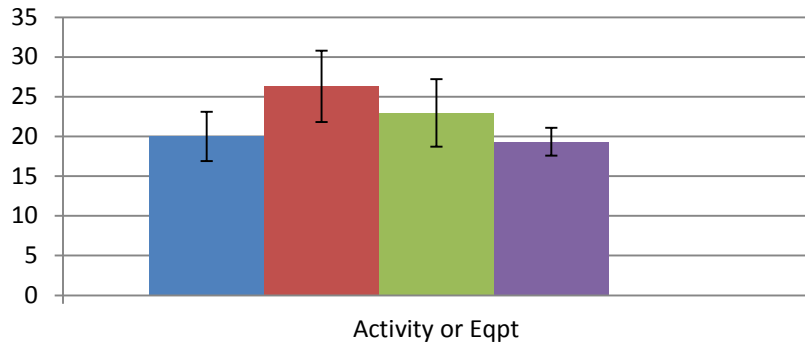
Atlanta



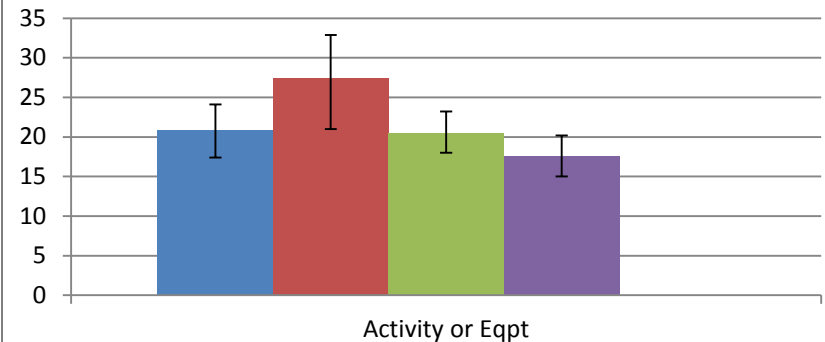
Chicago



New York



Houston



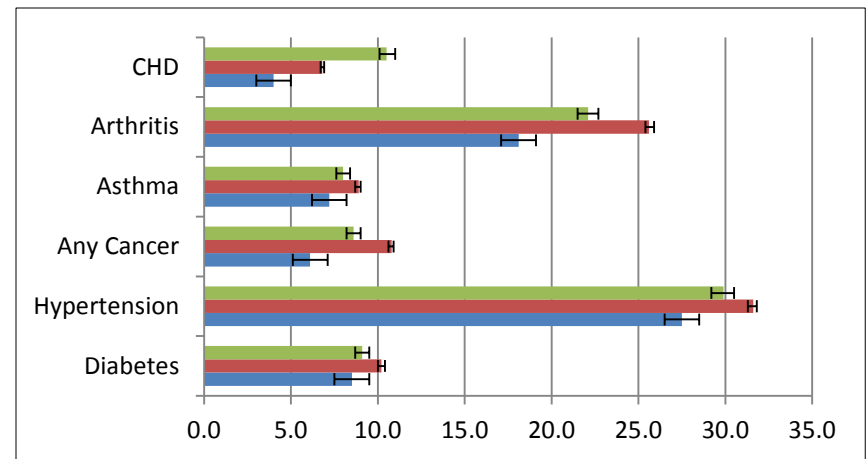
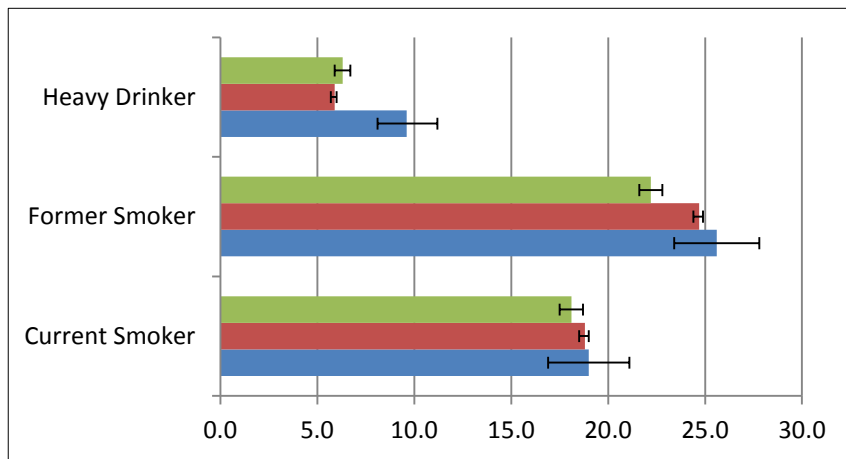
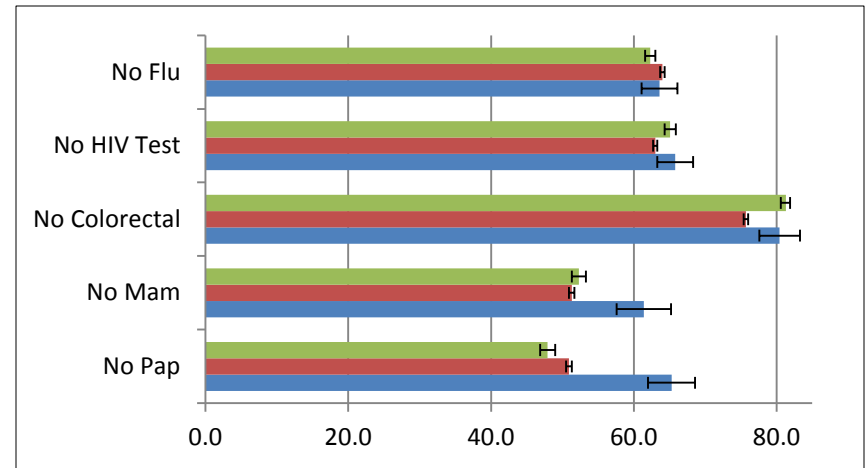
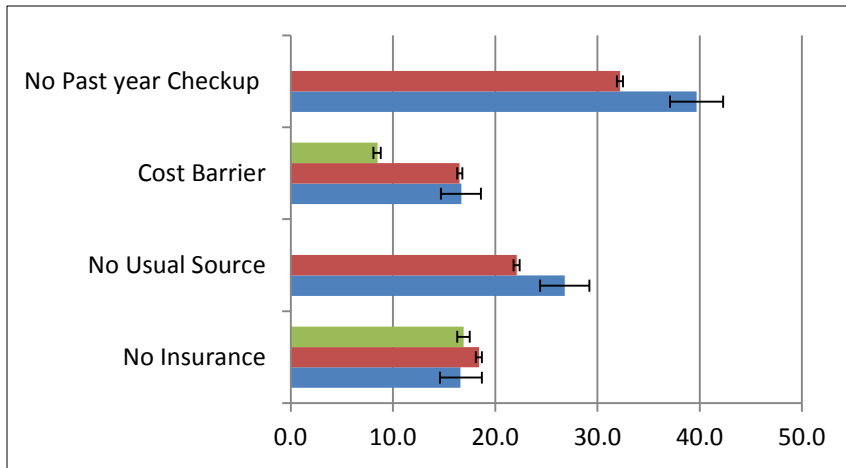
Mktg Inc. 2013
Blended

uSamp 2013
IPS Quota

YouGov 2013
IPS Matched

SMART BRFSS 2012
DF-RDD CATI

National: Health Care Access, Utilization, Behaviors & Outcomes

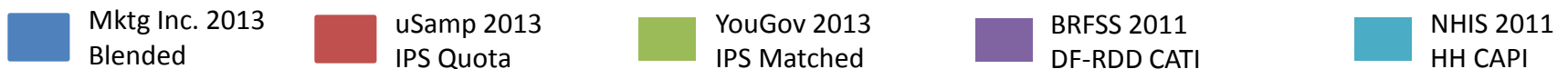
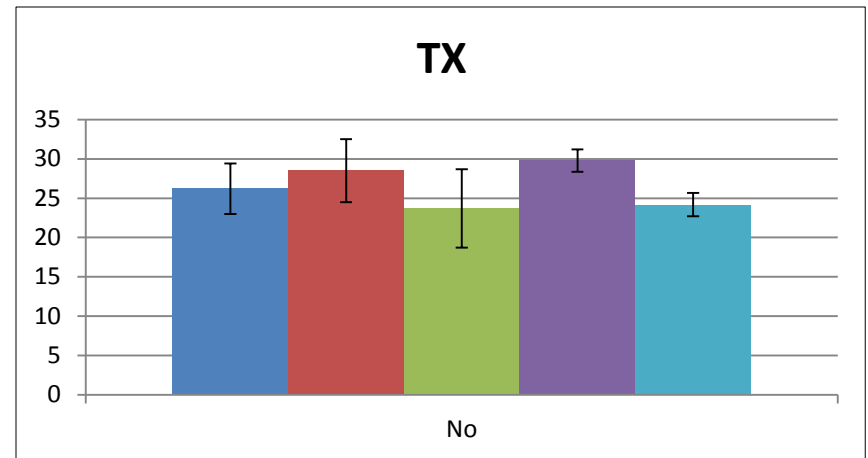
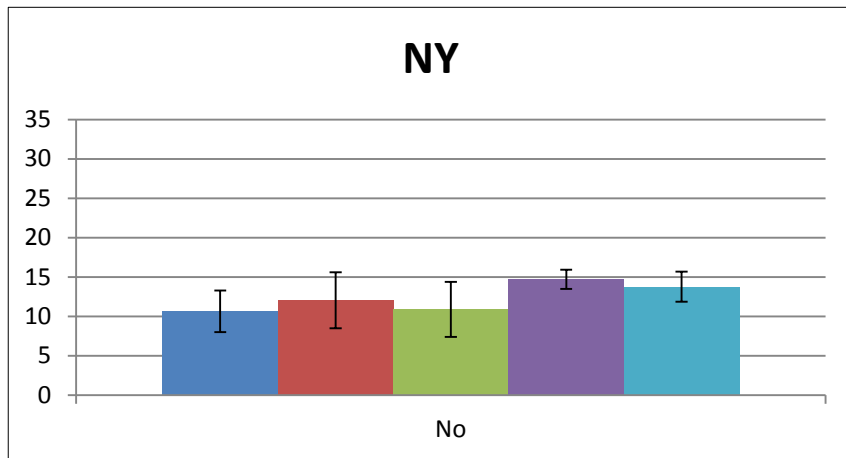
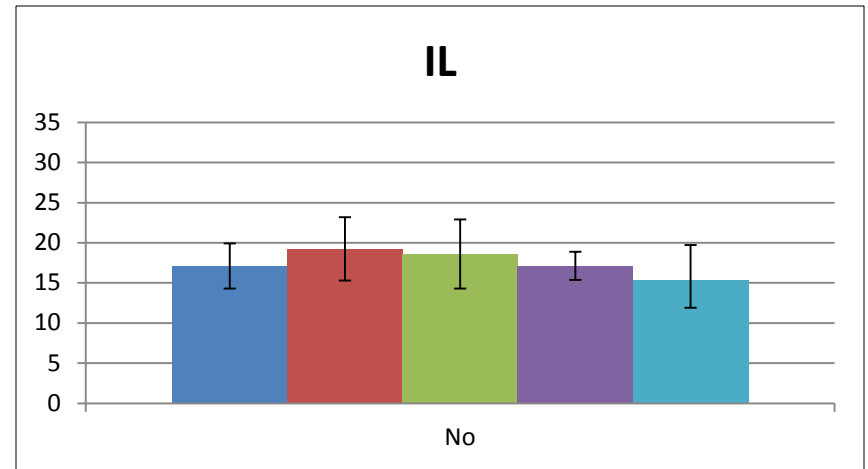
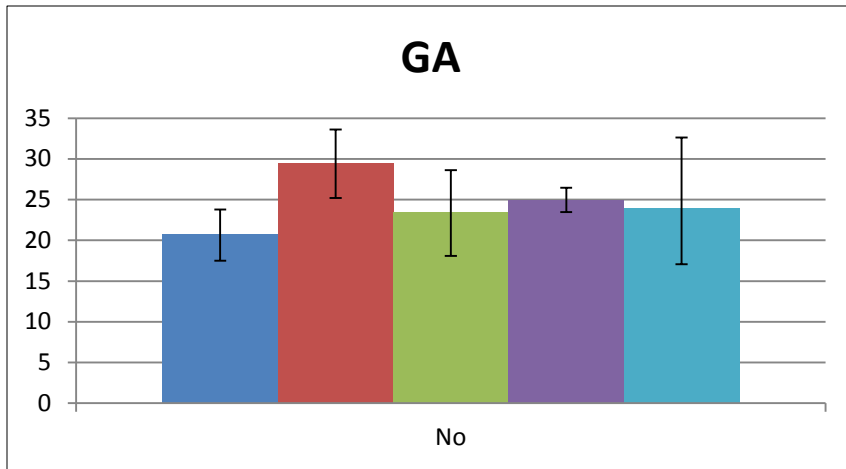


YouGov 2013
IPS Matched

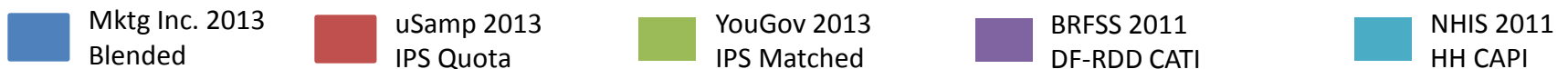
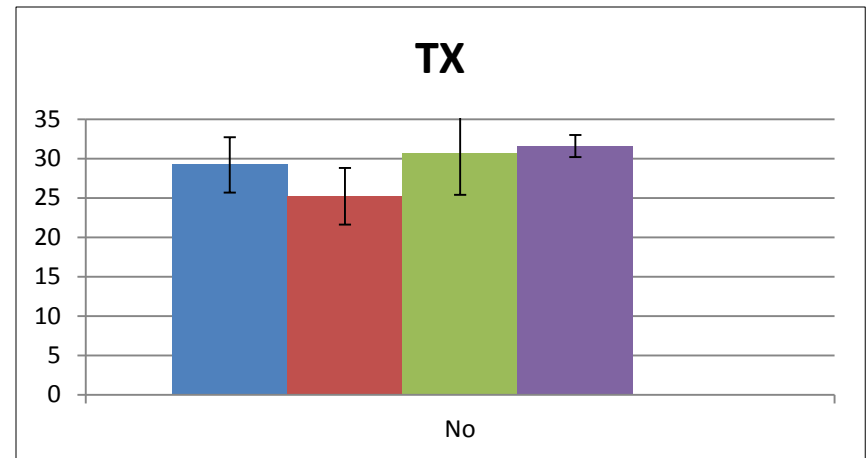
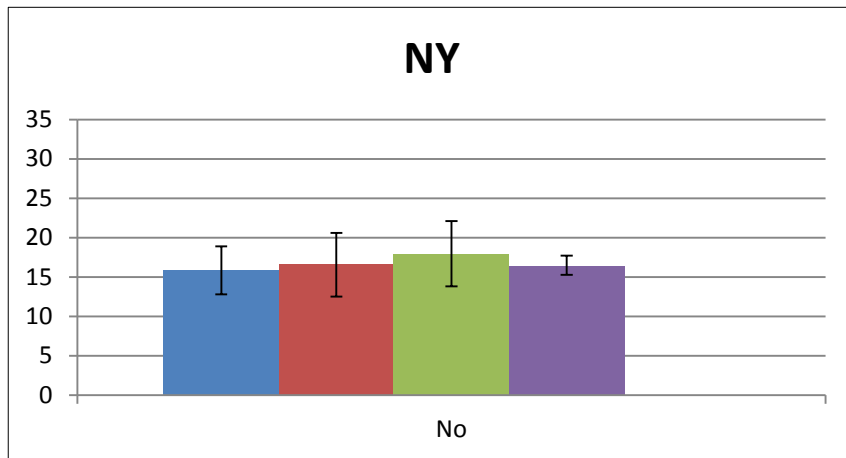
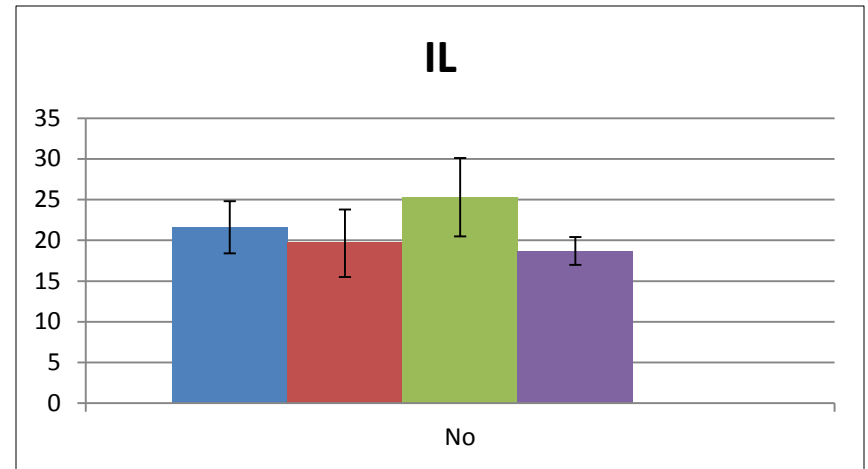
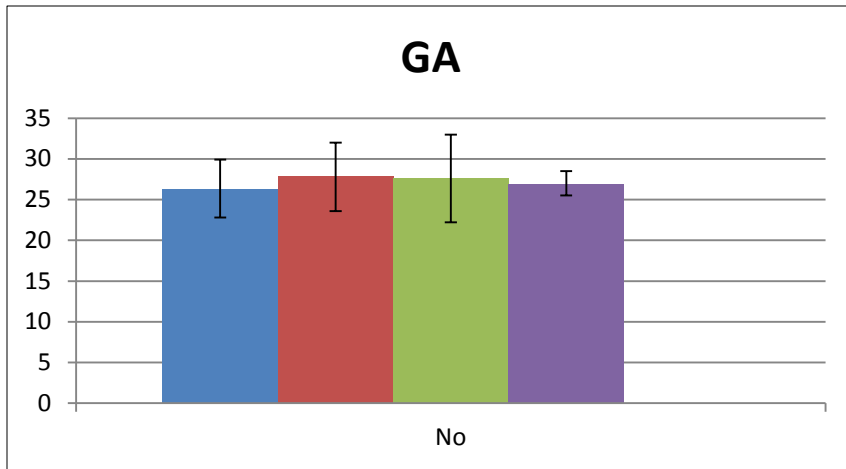
BRFSS 2012
DF-RDD CATI

NHIS 2012
HH CAPI

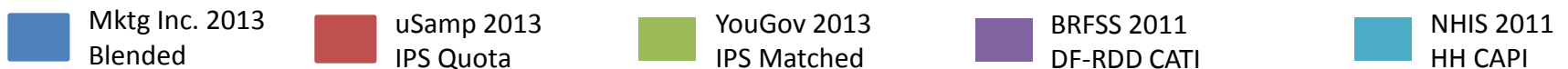
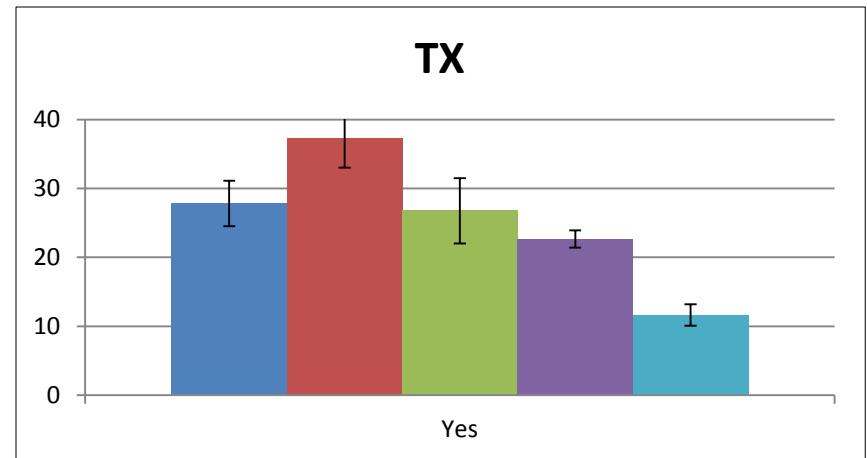
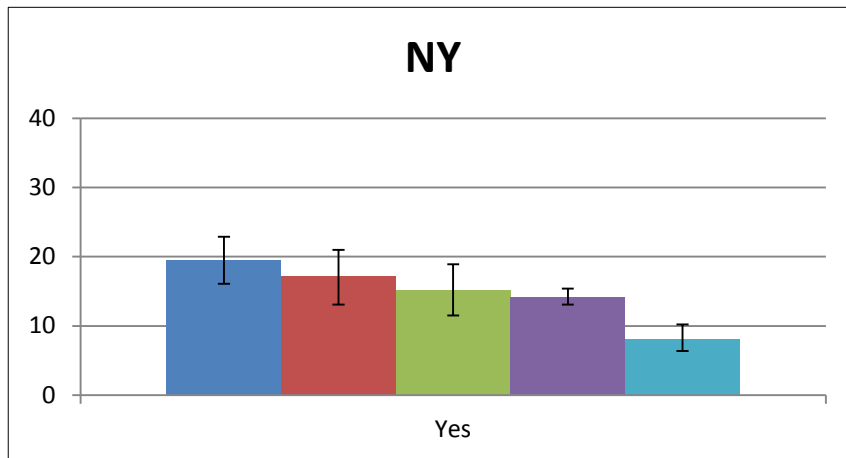
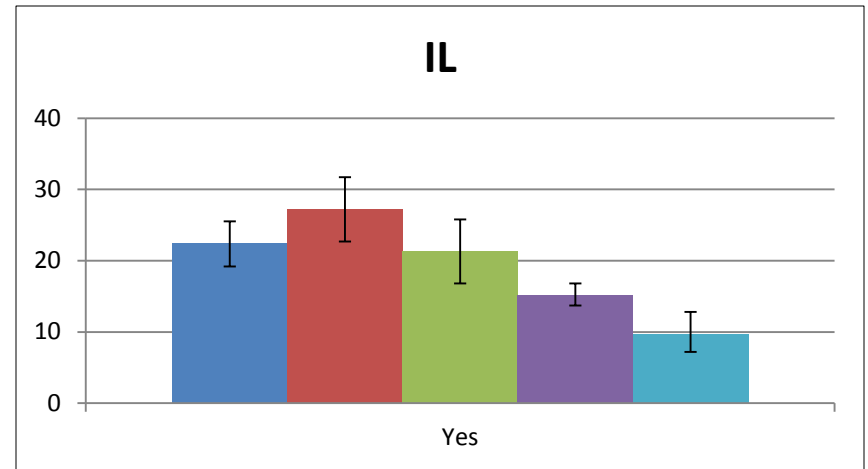
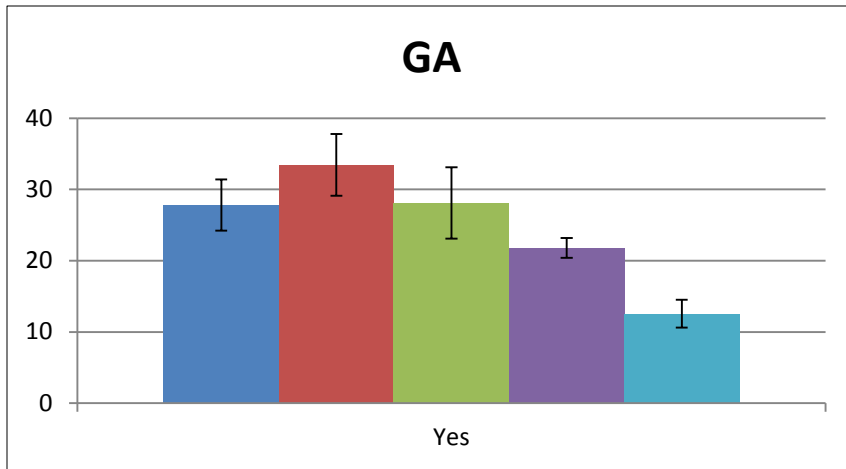
State: Health Insurance



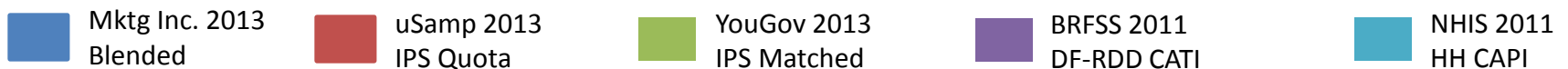
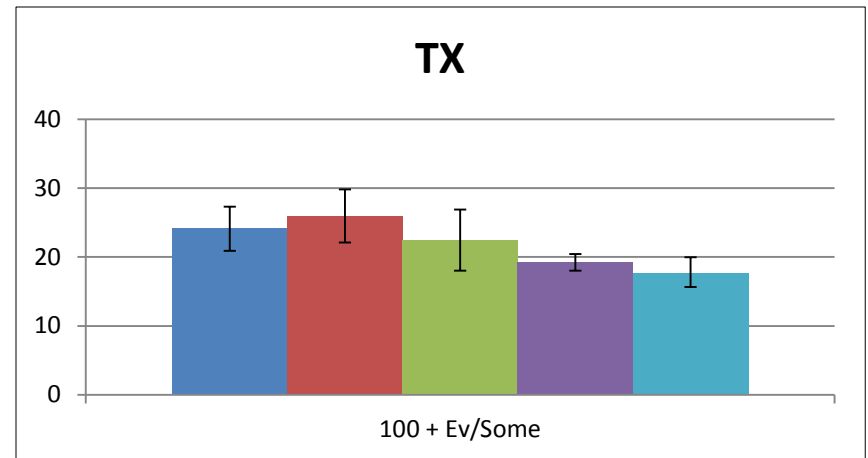
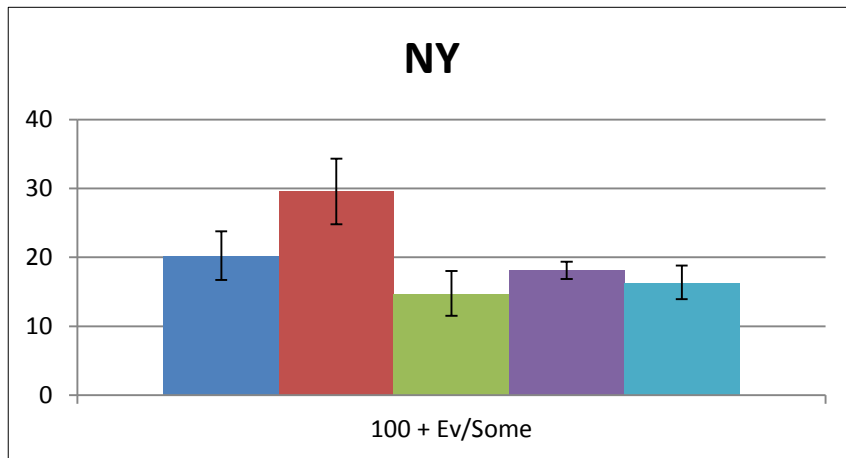
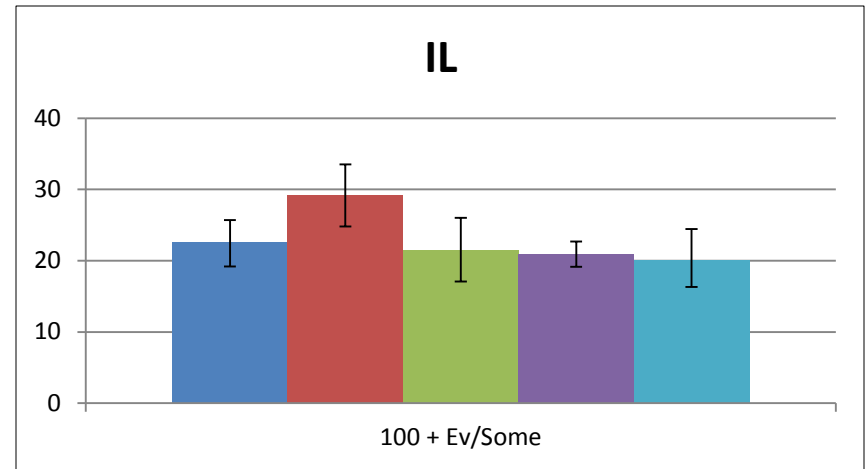
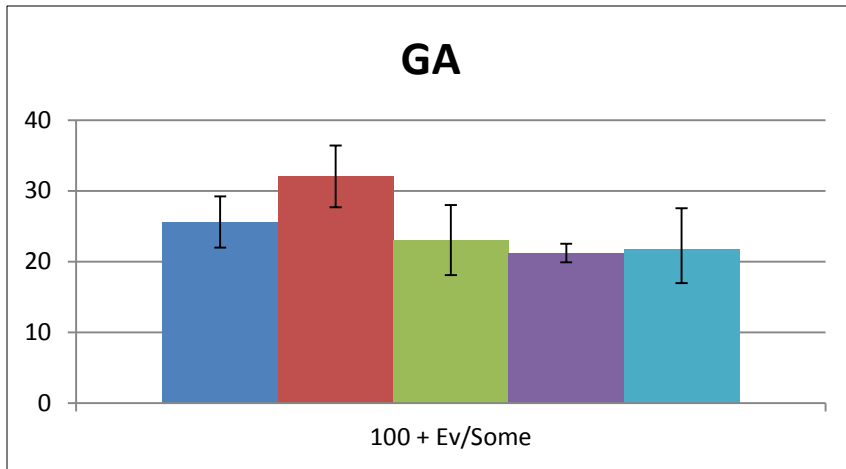
State: Primary Care Provider



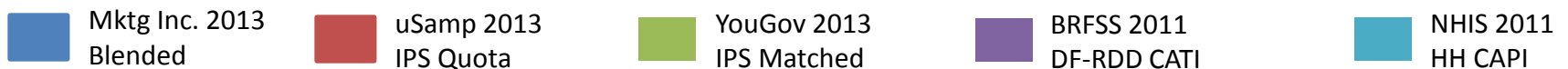
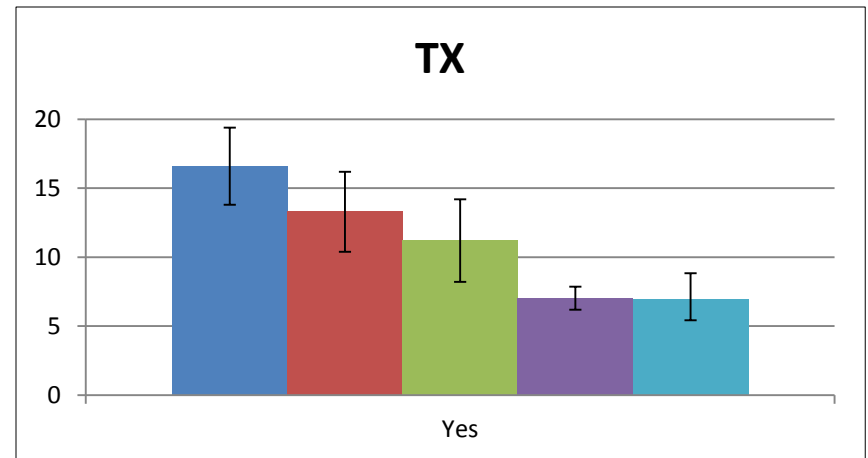
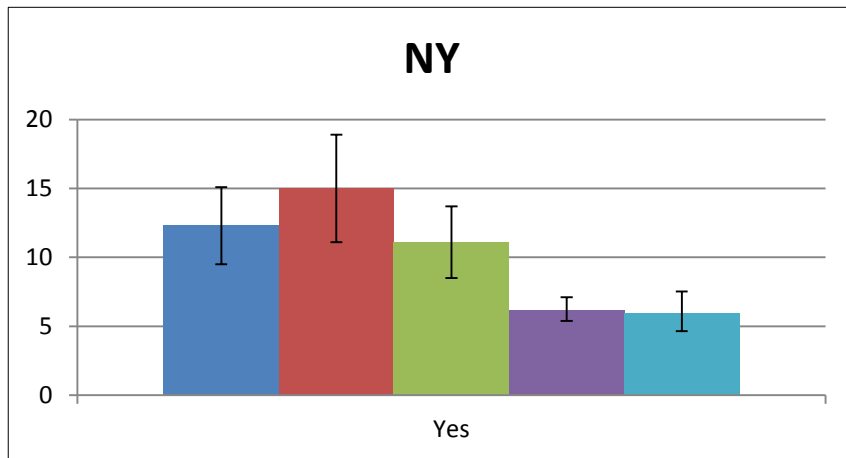
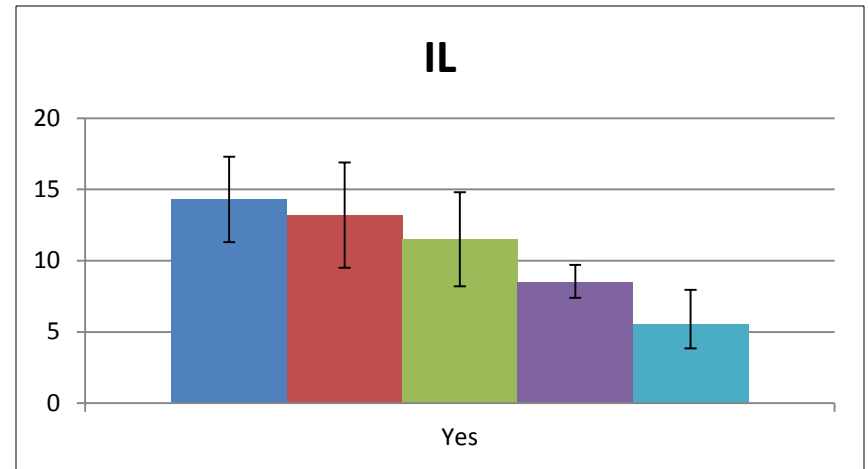
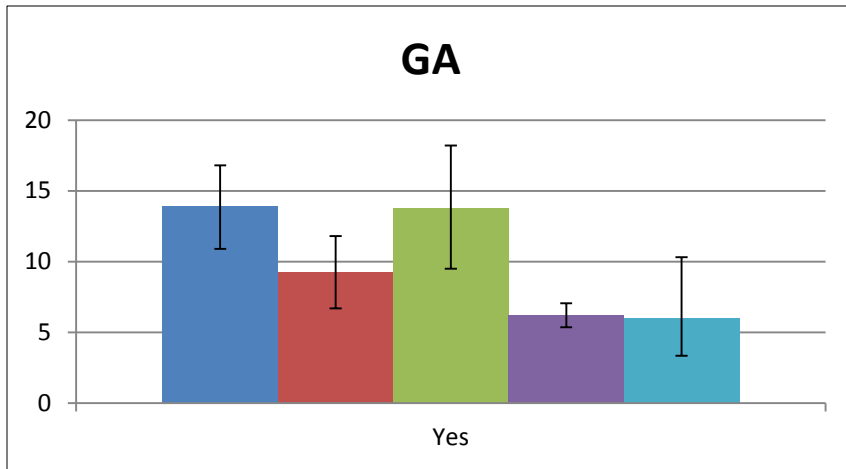
State: Cost Barrier



State: Current Smoker



State: Heavy Drinker



Quantifying Uncertainty

- The use of Frequentist confidence intervals with data from a non-probability sample is theoretically inappropriate
- Bayesian credible intervals are a more appropriate way to quantify uncertainty when analyzing data from a non-probability sample
- In our pilot studies, however, both methods yielded highly similar, if not identical, results

Uncertainty Comparison

Variable	Confidence Interval		Credible Interval	
Obesity	29.22	32.61	29.12	32.56
Diabetes	9.88	11.95	9.85	11.94
High BP	26.64	29.62	26.56	29.58

Major Benefits

- **Time** (samples constructed to be representative):
 - < 15 days for a national survey ~ 4,000 interviews
 - ~ 30 days for most states ~3,000 interviews
 - ~ 30 days for large (5+ million) MSAs ~2,000 interviews
- **Cost:**
 - Internet opt-in panels: \$5-\$15 per completed interview
 - Costs include editing and weighting
 - Dual-frame RDD State direct costs average ~\$70/CI
 - Considerable additional costs for editing and weighting

Preliminary Results

- **Great deal of similarity**
 - Results of sample matching comparable with BRFSS and NHIS
 - Variation among surveys consistent across states
 - Internet opt-in panels fairly accurate at lower levels of geography
 - Quota sampling not as accurate
- **Differences can be attributed to:**
 - Coverage effects (sample selection* outcome interaction)
 - Use of different control totals and weighting methods
 - Mode effects (face-to-face, telephone, Internet)
 - Question differences and order effects
 - Temporal changes (2013 vs. 2011)
 - Sample size differences
 - Cross-sectional differences

Acknowledgements

- Stephen Ansolabehere, Harvard University & CCES
- Steven Gittelman, Mktg.Inc.
- Douglas Rivers, Stanford University & YouGov
- Meena Khare, NCHS
- Georgia, Illinois, New York, and Texas Departments of Health, BRFSS
 - Rana Bayakly, Madhavi Vajani, Francis Annor, GA
 - Bruce Steiner, IL
 - Bethany Hawke, NY
 - Rebecca Wood, TX
- Haci Akcin, CDC
- Derek Ford, CDC
- Guixiang Zhao, CDC
- Soyoun Park, CDC