Cell Phone Methodology

PAPOR Mini Conference
June 28, 2013

Paul Johnson
SSI

Science. People. Innovention

PAPOR Mini Conference
June 28, 2013

Papers/slides obtained with author permission.
Any misinterpretations are my own.

Science. People. Innovention

Recent Methodological Updates Adopted for the National Immunization Survey (NIS)

Vicki Pineau[1], Kirk Wolter[1], Robert Montgomery[1], Bess Welch[1], and Stacie Greby[2]

[1]NORC at the University of Chicago

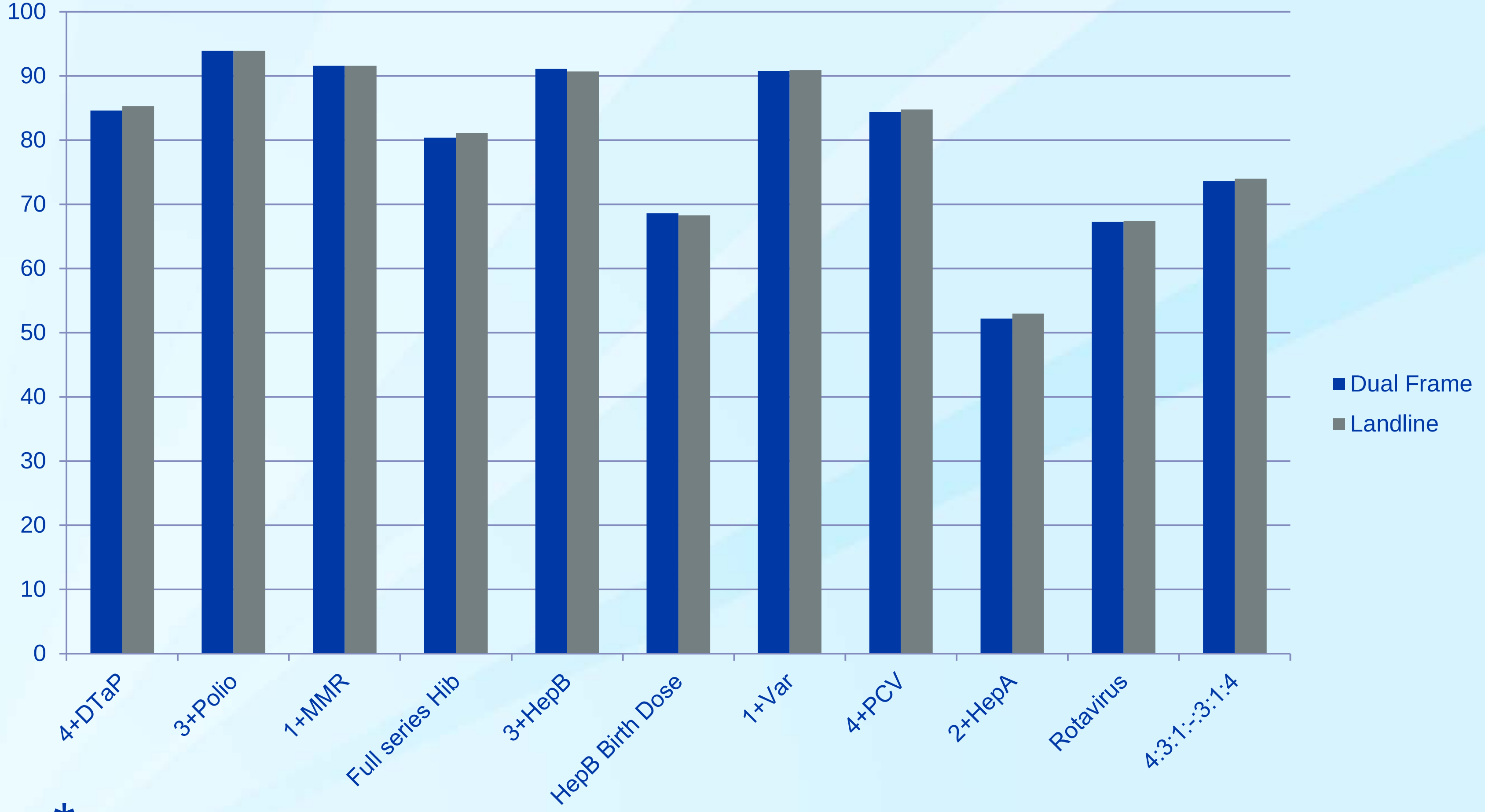
[2]Centers for Disease Control and Prevention

2013 AAPOR Annual Conference
May 17, 2013

Dual-Frame Results

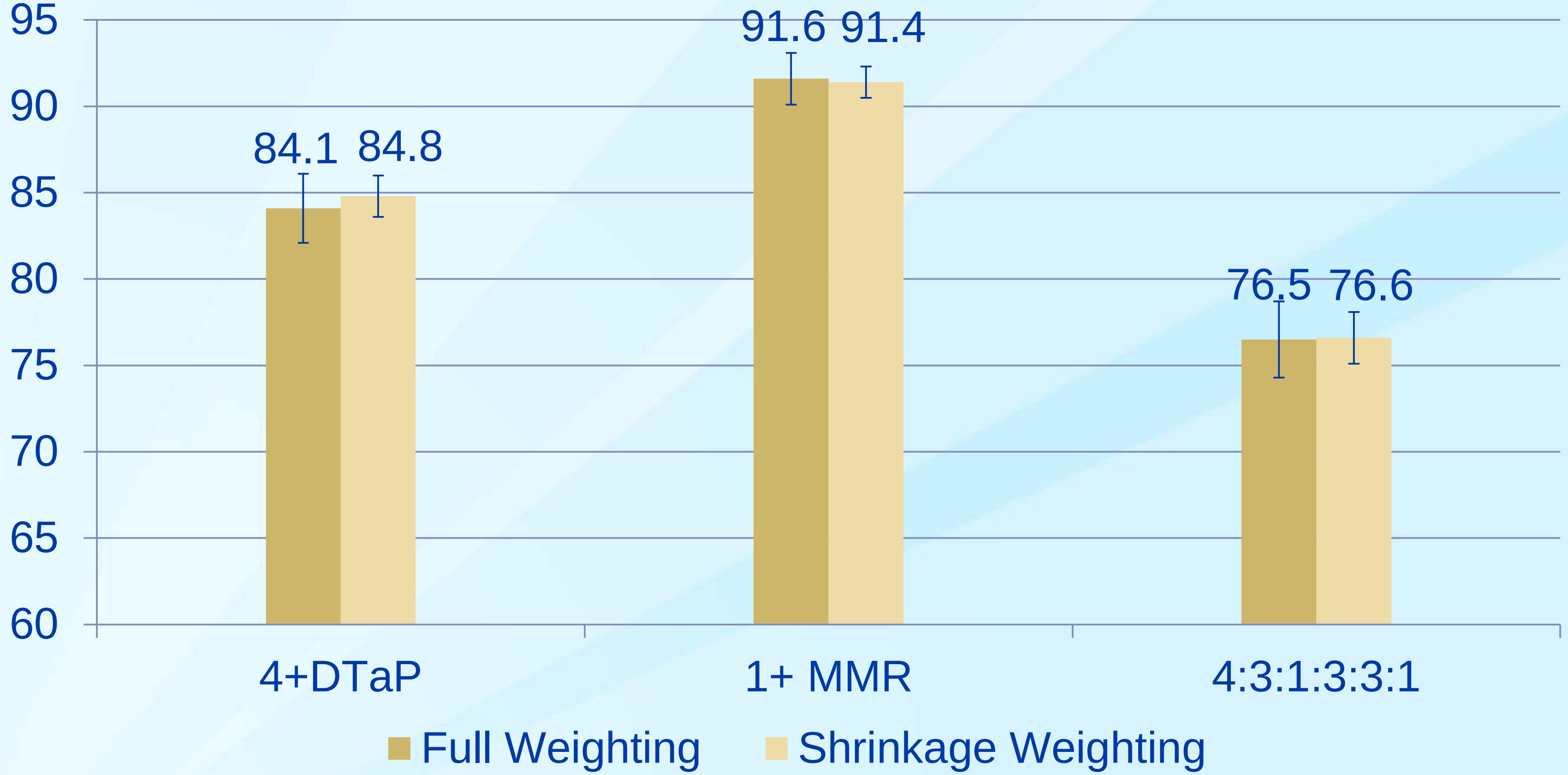
- **Minimized the perceived non-representative sample risk**
- **Maintained reasonable vaccination coverage estimates**
  - Comparable to LL only data and NHIS PRC
- **Response rates decreased**
- **Costs increased**
  - Expansion was supported by cost savings associated with the following methodological changes
    - Expansion of the NIS-Child age-eligibility criteria
    - Shortening the household telephone questionnaire
    - Incorporating efficient sample weighting methods via shrinkage weighting

Dual-Frame* vs. Landline Only Vaccination Coverage Rate Estimates, NIS, United States, 2011



*Dual-frame vaccination coverage rate estimates computed using shrinkage weighting to minimize MSE

Comparison of Bias and Variance of Shrinkage Weighting to Full Weighting, NIS, United States, Q2-Q4 2011



No statistically-significant differences at the α=0.05 level.

Reduction in width of 95% confidence intervals.

Birth Cohorts Using Any Day of Quarter Design, NIS, United States, 2011

	Month of Interview											
	Q1 2011			Q2 2011			Q3 2011			Q4 2011		
Month of Birth	Jan-11	Feb-11	Mar-11	Apr-11	May-11	Jun-11	Jul-11	Aug-11	Sep-11	Oct-11	Nov-11	Dec-11
Dec-07	37	38	39	40	41	42	43	44	45	46	47	48
Jan-08	36	37	38	39	40	41	42	43	44	45	46	47
Feb-08	35	36	37	38	39	40	41	42	43	44	45	46
Mar-08	34	35	36	37	38	39	40	41	42	43	44	45
Apr-08	33	34	35	36	37	38	39	40	41	42	43	44
May-08	32	33	34	35	36	37	38	39	40	41	42	43
Jun-08	31	32	33	34	35	36	37	38	39	40	41	42
Jul-08	30	31	32	33	34	35	36	37	38	39	40	41
Aug-08	29	30	31	32	33	34	35	36	37	38	39	40
Sep-08	28	29	30	31	32	33	34	35	36	37	38	39
Oct-08	27	28	29	30	31	32	33	34	35	36	37	38
Nov-08	26	27	28	29	30	31	32	33	34	35	36	37
Dec-08	25	26	27	28	29	30	31	32	33	34	35	36
Jan-09	24	25	26	27	28	29	30	31	32	33	34	35
Feb-09	23	24	25	26	27	28	29	30	31	32	33	34
Mar-09	22	23	24	25	26	27	28	29	30	31	32	33
Apr-09	21	22	23	24	25	26	27	28	29	30	31	32
May-09	20	21	22	23	24	25	26	27	28	29	30	31
Jun-09	19	20	21	22	23	24	25	26	27	28	29	30
Jul-09	18	19	20	21	22	23	24	25	26	27	28	29
Aug-09	17	18	19	20	21	22	23	24	25	26	27	28
Sep-09	16	17	18	19	20	21	22	23	24	25	26	27
Oct-09	15	16	17	18	19	20	21	22	23	24	25	26
Nov-09	14	15	16	17	18	19	20	21	22	23	24	25
Dec-09	13	14	15	16	17	18	19	20	21	22	23	24
Jan-10	12	13	14	15	16	17	18	19	20	21	22	23
Feb-10	11	12	13	14	15	16	17	18	19	20	21	22
Mar-10	10	11	12	13	14	15	16	17	18	19	20	21
Apr-10	9	10	11	12	13	14	15	16	17	18	19	20
May-10	8	9	10	11	12	13	14	15	16	17	18	19

Color Key

- Age Eligible Day of Screening
- Age Eligible Any Day of Quarter (Old)
- Age Eligible Any Day of Quarter (Young)

Health Care Research

Multicultural Research

Sports & Leisure Research

Religion & Jewish Community Research

Teen & Young Population Research

Public Health Research

Older American Research

Sampling Cell Phones by Rate Center: Efficacy, Coverage and Incidence

David Dutwin, SSRS

David Malarek, MSG

SSRS social science research solutions

Example Map of Rate Centers

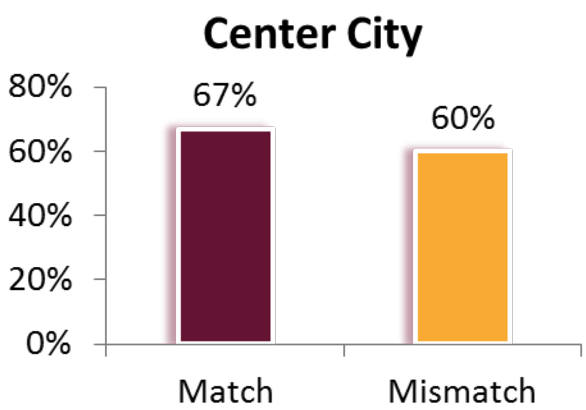
DC Metro

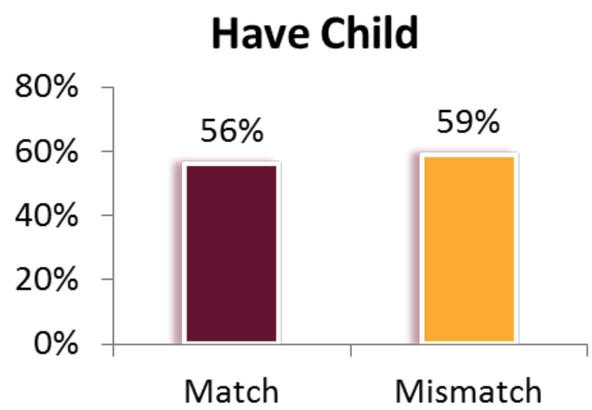
Data

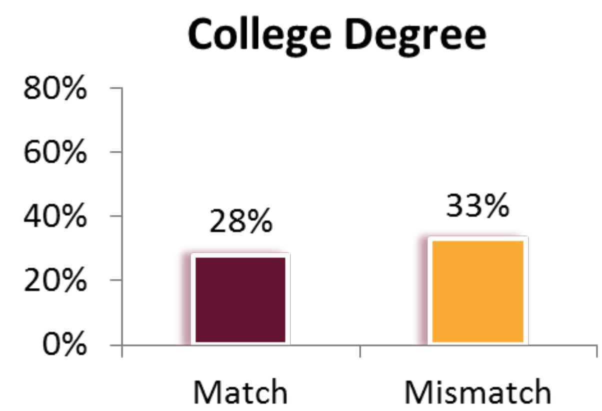
- MSG’s Rate Center Dataset:
  - Rate center geographic data
  - Appended Census data for demographics
  - Square Miles
  - Population
  - Households
  - 18+ Population
  - Ethnicity
  - Vacancy
  - Tenure
  - CPO Households
- SSRS EXCEL omnibus survey:
  - 12,229 cell phone interviews
  - Full demographic battery
- Unified EXCEL/Rate Center Dataset

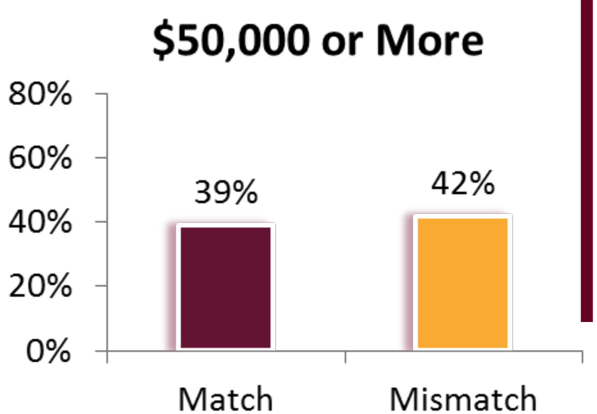
Significant Differences by Respondent Characteristics
Rate Center Mismatches

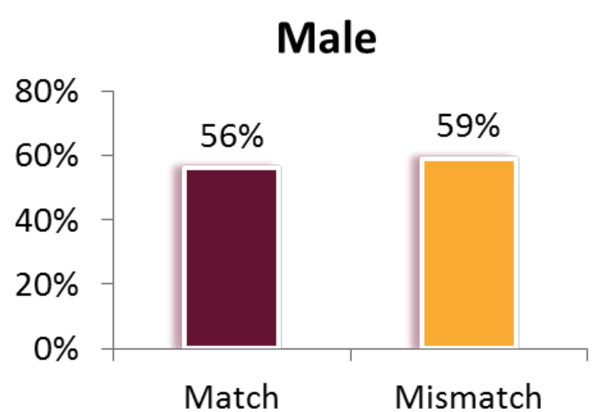
50 percent of cell owners live in their rate center

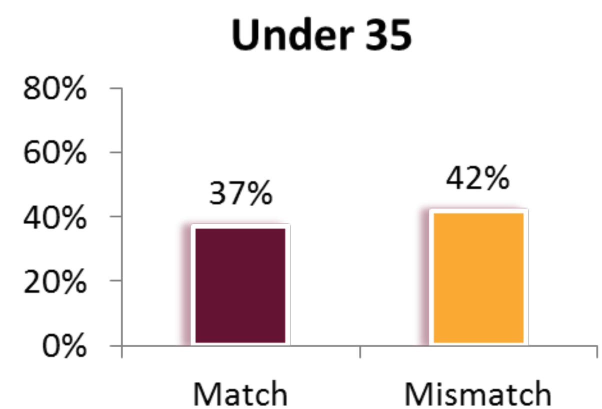




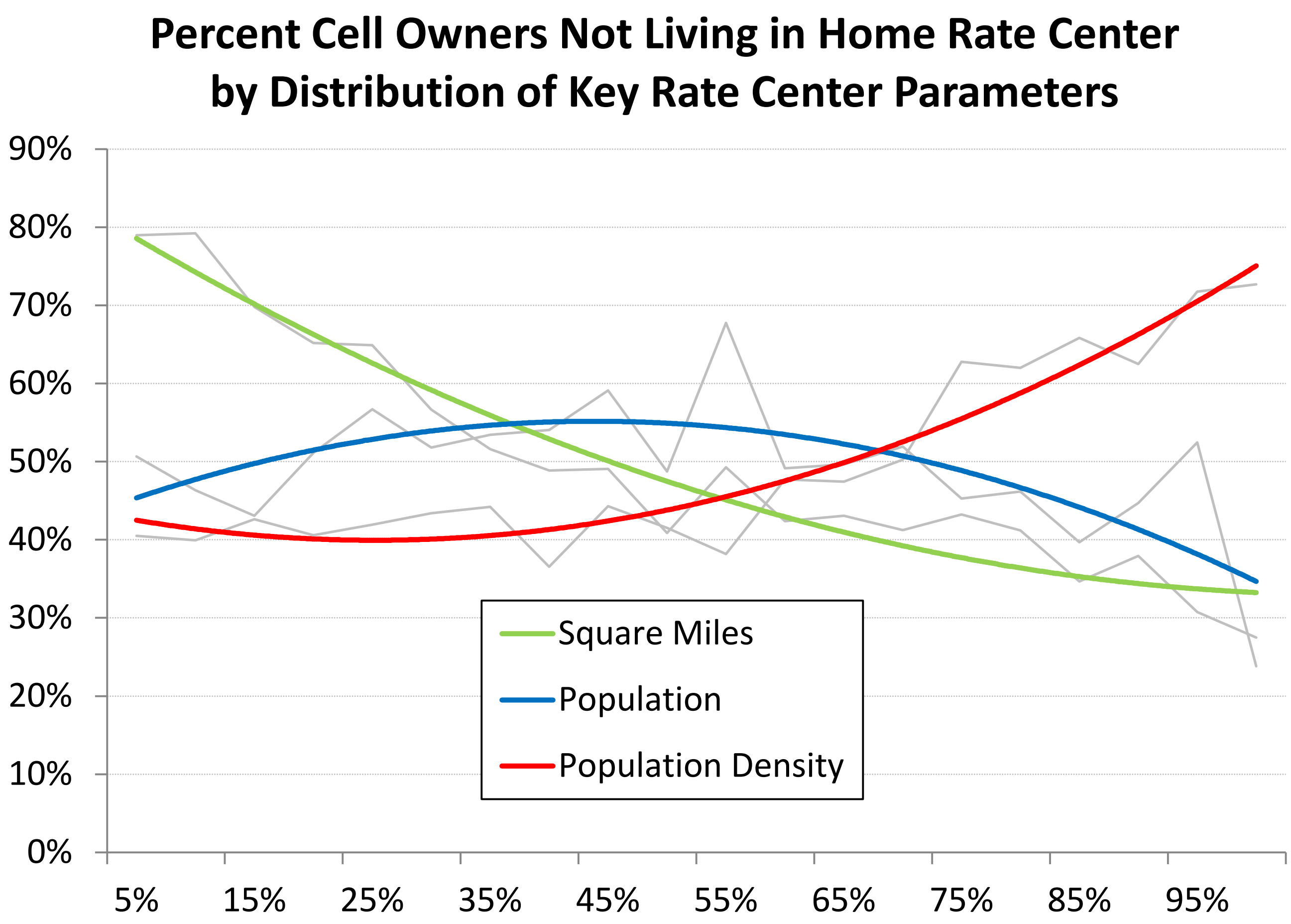




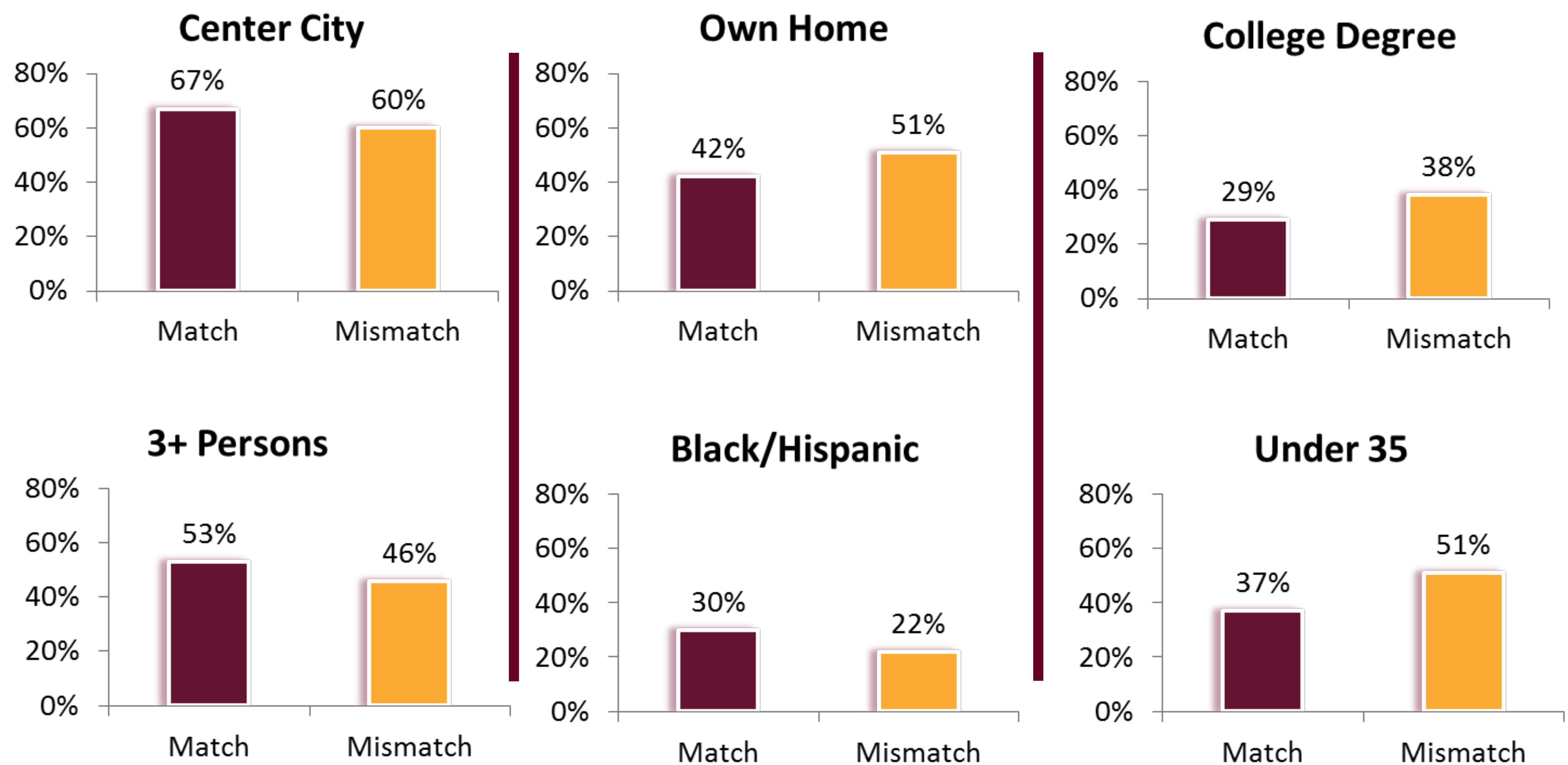




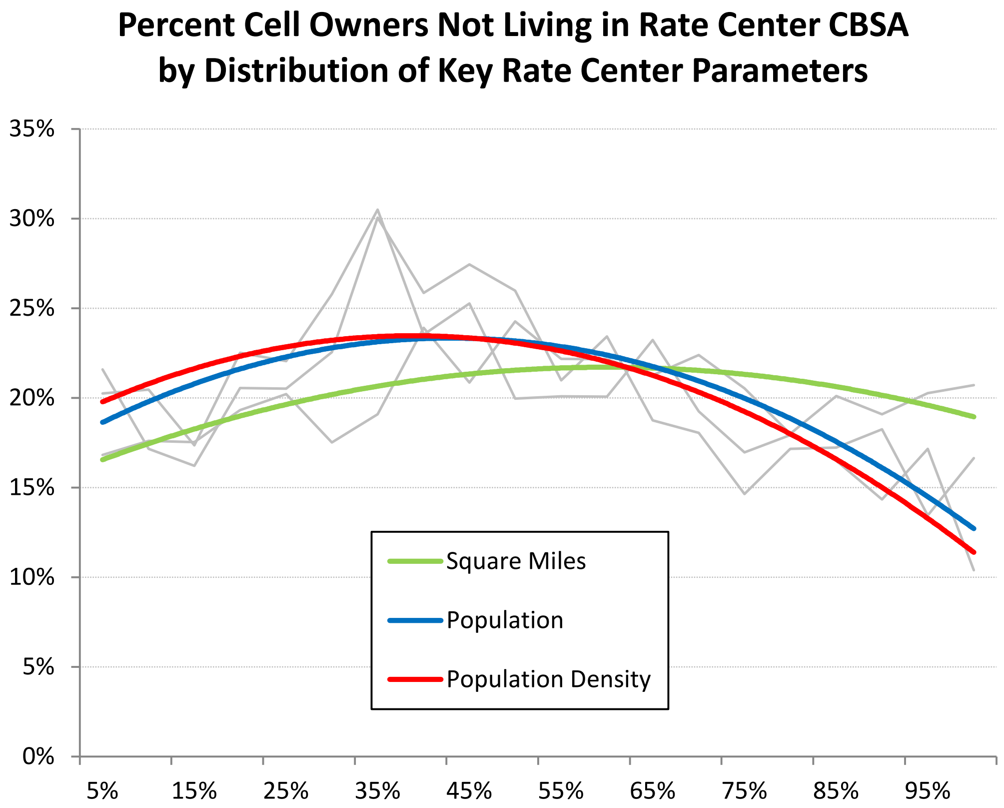
Differences by Rate Center Characteristics



Significant Differences by Respondent Characteristics
CBSA Mismatches



Differences by Rate Center Characteristics, CBSA Level



Rate Centers: Conclusions

1. Half of cell phone owners live in their rate center
2. Over three quarters live in the CBSA of their rate center
3. Mismatches more likely in mid-atlantic and northeast; in small rate centers; in rate centers high in population density
4. Mismatches more likely to be non-metro, educated, and young; CBSA mismatches more likely to be home owners, educated, white, young, and in single family households
5. CBSA on average attain 73 percent incidence; 75 percent coverage; logically, geographies smaller than a CBSA will be worse on both metrics

Alternative Sample Selection and Data Collection Strategies for Balancing Cell Phone Response Distribution Across County/Region Level Geographies in a Dual Frame Telephone Survey

AAPOR 2013 Methodological Briefs: Cell Phones

Howard Speizer, Marcus Berzofsky, Tom Duffy, Jamie Ridenhour (RTI)
Tim Sahr (Ohio State University)

May 17, 2013

Rate Center

- Rate Center holds some promise as a stratification variable because it is assigned to every cell number
- Classification error however, is an issue:
  - Incorrect state=7.7%
  - Classification error statewide=33%
  - In Hamilton County the false-positive rate was 48%

OMAS Results with Adjusted Rate Center

MedicaidRegion	Estimate of cell completes based on population totals	Estimate of cell completes based on %COP	Estimate of cell completes based on rate center	Estimate of cell completes based on adjusted rate center	Cell achieved - expected based on modified rate center	Non-response ratio
Central	1039	1162	1056	1094	39	1.04
East Central	695	675	745	701	-21	0.97
Northeast	1048	891	721	678	-77	0.89
Northeast Central	257	222	322	309	-37	0.88
Northwest	573	569	626	633	8	1.01
Southeast	302	249	253	254	-31	0.88
Southwest	772	887	940	978	42	1.04
West Central	507	536	527	544	77	1.14

Sample Design for Next OMAS

- Using 2012 survey data we can estimate rate center error and non-response by county to improve the cell phone sample stratification

- Allocate desired number of cell phone completes to counties based on corrected rate center distribution to get desired nominal sample size by county

Steps to Determine Sample Size by County

- Adjust nominal sample size by following adjustments to get starting sample size of cell phone numbers
  1. Frame error adjustment: ratio of initial rate center distribution and corrected rate center distribution
  2. Response rate adjustment: expected aggregate response rate times the ratio of prior survey number of completes over prior survey expected number of completes
  3. Out of state adjustment: ratio of number of contacted persons in rate center county indicating they lived out of state over number of contacted persons in rate center county
  4. Ineligibility adjustment: aggregate ineligibility adjustment specific to study design other than out of state adjustment

Understanding Bias in Appended Wireless Billing ZIP Code Data

Tara Merry, Andy Weiss, Mikelyn Meyers, Paul Schroeder (Abt SRBI)

Kristie Johnson (NHTSA)

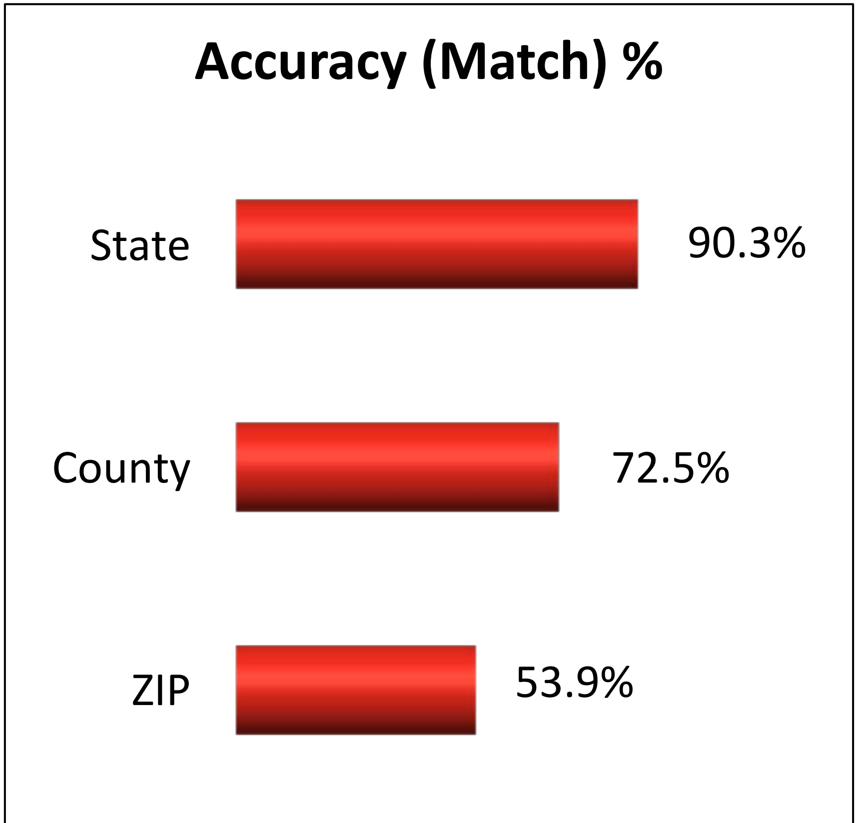
AAPOR 2013 Conference May 16-19, 2013 Boston, MA

Appending Billing ZIP Process

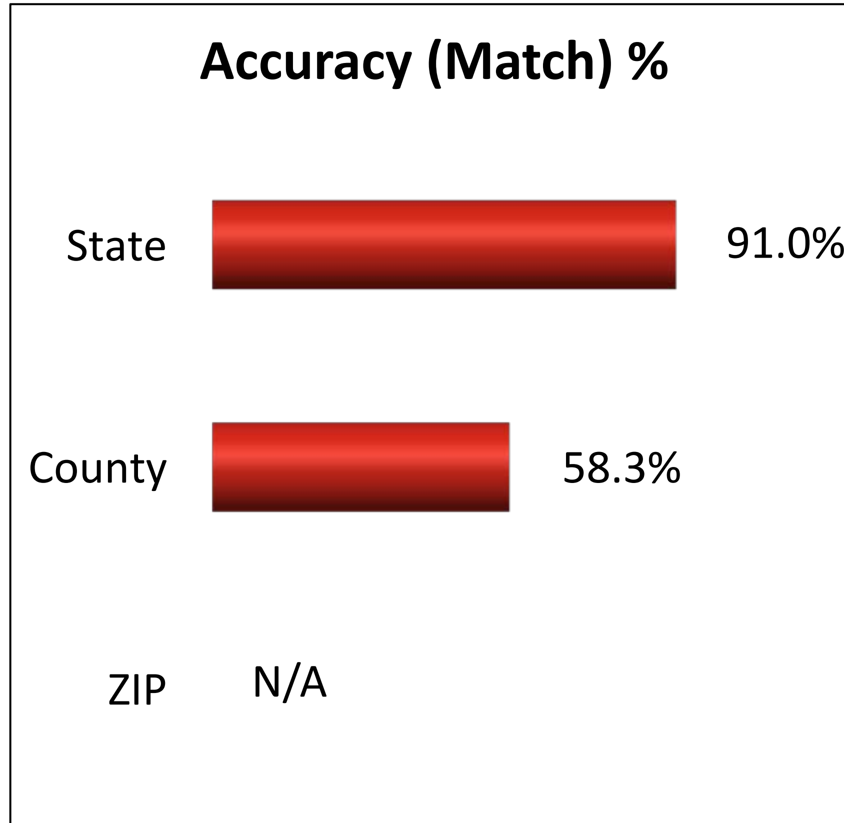
- SSI's *Wireless Geo-ID* Process
  - Post-production process (run after sample is pulled)
  - Appends billing ZIP code of the cell phone number using information in proprietary data sources
- Three geographic data sources for comparison:
  - Rate center data (all sample records): based on the FIPS of the rate center in the cell sample
  - Billing data (matched cases only): based on data appended from Wireless Geo-ID process
  - Respondent-reported data (screened cases only): geographic residence information reported by respondent in survey

National Data: Accuracy of Billing ZIP Compared to Rate Center

Billing vs. respondent-reported data



Rate center vs. respondent-reported data



How do matched and unmatched cases differ: *National vs. NYC Demographics*

Unmatched are more likely to be:	National Sample (Unmatched vs. Matched)	NYC Sample (Unmatched vs. Matched)
Less educated	HS or less (54.1% vs. 37.5%) [2]	HS or less (60.9% vs. 28.6%) [2]
Lower income	Less than $50K (64.4% vs. 48.3%) [2]	Less than povertyX2 (68.1% vs. 35.0%)[2]
Hispanic, non-White	Hispanic (26.0% vs. 14.2%) [2] White (47.8% vs. 66.3%) [2]	Hispanic (42.4% vs. 24.0%)[2] White (14.0% vs. 35.9%)[2]
Renters	Rent (48.4% vs. 39.9%) [2]	*Not asked*
Younger	16 to 34 (48.3% vs. 44.0%)[1]	*Not significant*
Cell-phone only	*Not significant*	CPO (72.5% vs. 62.6%)[1]

[1]p<.05, [2]p<.001

Conclusion & Recommendations

- Appended billing ZIP can be used to improve efficiency
  - Low match rate can limit use/benefit
  - Unmatched cases *are different*, should include at least some
- Lots of regional variability in accuracy of rate center & appended ZIP data (completeness/accuracy)
  - Use all available info when defining frame, but beware variation due to change over time or small sample sizes
- Develop initial sampling design/stratification but plan to evaluate after implementation and adjust if needed
  - Pilot test can be useful, but needs to be large enough
  - Test with limited replicates, but need time to evaluate/refine

Cell-Phone Sampling Frames: Effectiveness and Dependability of Recent-Usage Data

American Association for Public Opinion Research
Boston, Massachusetts
May 19, 2013

Presented by: Robert DeHaan and Robin Gentry
Acknowledgements: SSI, Neustar

ARBITRON

Why Cell Phone RDD and Why Now?

» Dual Frame (Address Based v Cell Phone RDD)

» Increasing Incidence of Cell Phone Only/Mostly Households

» New Innovations
  - Usage Based Activity Indicators
  - Billing Zip Append
  - Pre-placement mailings

ARBITRON

Study Overview (Activity Indicator Split)

» Grouped according to usage over last 10 months

1	2
High Usage	Moderate Usage

+

3
No Usage

- 16,050 “active” numbers hand-dialed
- 2 weeks across 6 diverse markets
- Pre-placement postcard (50/50)
- Placement phone calls made
- Cell Phone vs. Landline usage questions asked

Contactability Study

ARBITRON

Contact/Consent Rates

$$\text{Consent Rate} = \frac{\text{Agree}}{(\text{Agree + Non-Agree})}$$

Disposition	High Usage	Moderate Usage
Operational Consent Rate	**17.11%**	**16.39%**
Agree	12.80%	9.75%
No Answer	19.66%	16.24%
Won't Talk	38.36%	28.68%
Answering Machine	4.00%	4.27%
Disconnect	19.18%	35.86%
Other	6.00%	5.20%

Usable: Agree (Agree)

Usable: Non-Agree (No Answer, Won't Talk, Answering Machine)

Unusable (Disconnect, Other)

Conclusions

» Activity Indicators are useful when screening out unused numbers (Indicator 3)

» Pre-placement postcard was not particularly effective, but better results could be expected with a letter and a dollar

» Improved proportionality amongst 12-24 year olds, offset by reduced representation amongst 25-34 year olds

» Cell Phone Only density in frame is nearly double what was found in a 2006 Arbitron study

Attempting to Boost RDD Cell Sample Productivity by Identifying Non-working Numbers Prior to Dialing

Princeton Survey Research Associates International

Science. People. Innovention

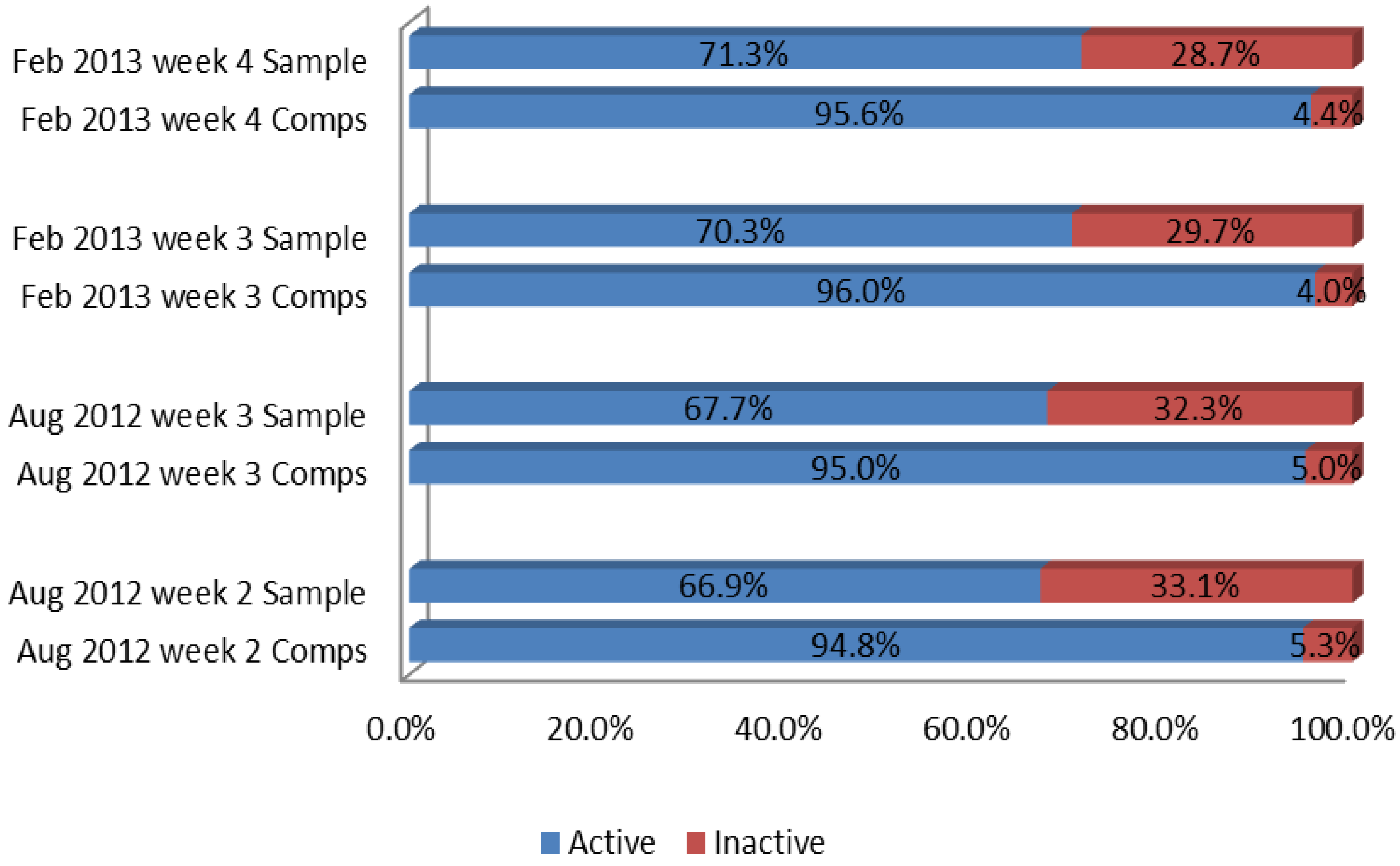
Michelle Mosher – SSI

Jonathan Best – Princeton Survey Research Associates International

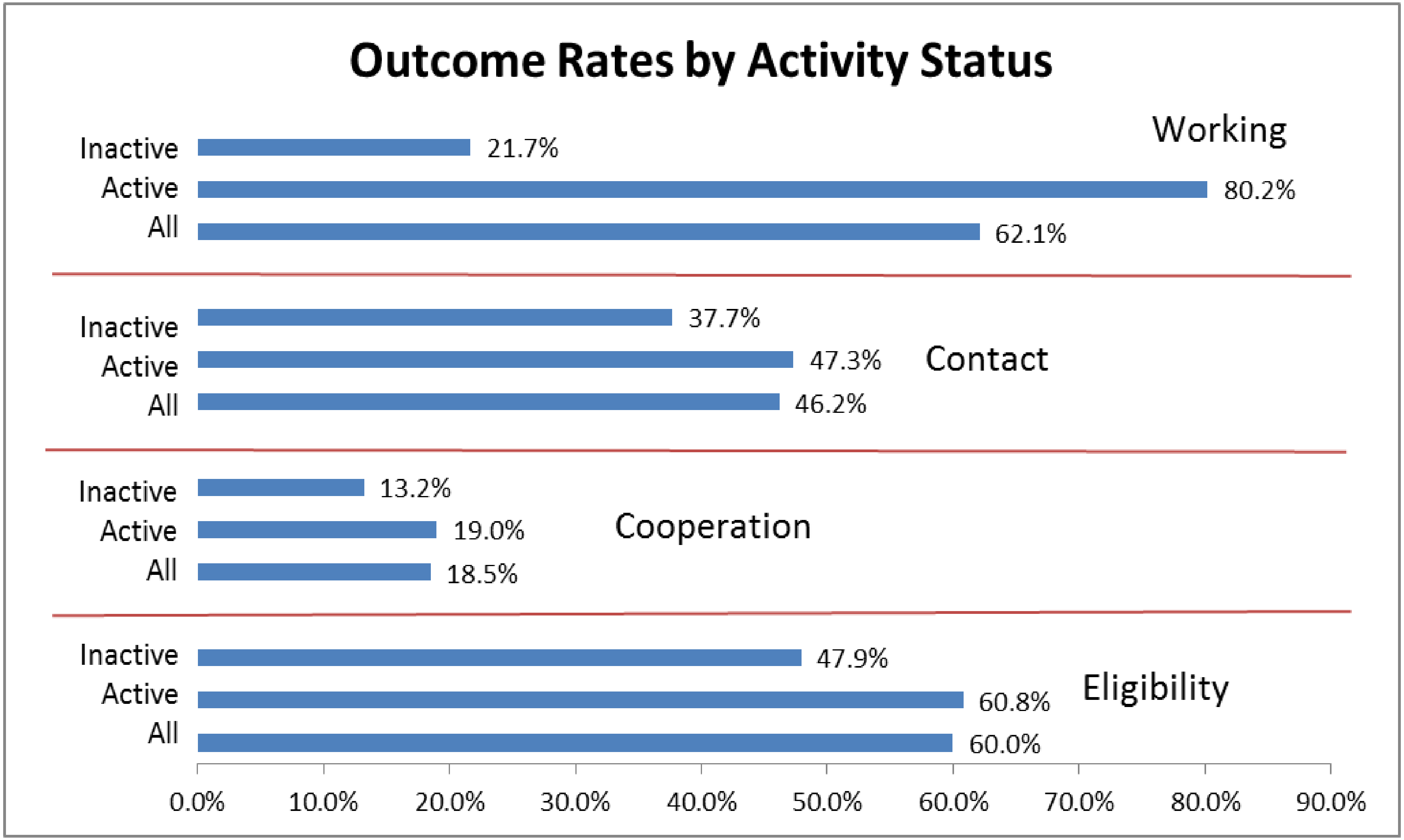
What is the 'Phone Activity Flag'?

- Phone activity = the frequency of calls generated from a specific wireless telephone number
  - Sourced from caller ID network information
  - Information from over 175 carriers
    - Majority of regional carriers
    - Majority of cable companies
    - Inbound calls from 'out of network' also captured
    - 95% of outgoing calls captured (remaining 5% are mostly activity from text only numbers with no voice capability)

Omnibus Sample – Screening Results



Disposition Comparison



Effects on Productivity

Interviewer Hours by Activity Status				
	Active	Inactive	Total	
Interviews (Ints)	480	20	500	
Interviewer hours (Int hr)	123.8	31.7	155.5	
Productivity (Ints/Int hr)	3.9	0.6	3.2	

$$\frac{(3.9 - 3.2)}{3.2} = 21\%$$

Conclusions

- Cell phone screening does a good job at identifying unproductive numbers. Over 60% of all non-working numbers are flagged as inactive.
- Inactive numbers lead to few interviews – only about 4 to 5 percent of total completes
- The screening increases the working rate of a Continental US cell sample from 62% to 80%
- No evidence that excluding the inactive sample leads to bias in basic sample demographics
- We estimate that excluding the inactive numbers would lead to increases in cell sample productivity of 17-21%.

Experian[SM]

Marketing Services

The Mechanics of GPS Geo-Location for Mobile Devices:

Their Potential for Measurement Error and Some Illustrative Data

Dr. Max Kilger and TraShawna Boals

OVERVIEW

- Uses of geo-location in survey and market research
- Technologies for generating Geo-location data
- How GPS works
- Measurement Errors
- Illustrative Data from passive meter technology

Uses of geo-location in survey and market research

- What is geo-location research?
- Examples:
  - Active: When a participant is located near a point of interest to the research, a survey is triggered.
  - Passive: When a participant has a mobile app installed on device and it periodically captures GPS coordinates.

Technologies for generating Geo-location data in mobile devices

- Global Positioning Systems (GPS)
- Cell Tower Triangulation
- Wifi geo-location

Technologies for generating Geo-location data in mobile devices

- Global Positioning Systems (GPS)
  - U.S. system - 31 satellites in medium earth orbit
  - At least 4 satellites visible at any point on Earth
  - More satellites visible = more accurate geo-location fix

How GPS Geo-Location Works

Accuracy: +/- 10 meters

- GPS signal has three basic components
  - **Date/Time/Satellite Status** – ultra-accurate time signal used in calculating satellite-GPS receiver trip time
  - **Ephemeris** – very accurate estimate of the satellite position at the time of signal
  - **Almanac** – reference data with satellite id's, approximate orbits, satellite clock statuses

- GPS-equipped device position determined by combining time and ephemeris data (position) from multiple satellites

How aGPS Geo-Location Works

- aGPS – assisted GPS is a way to speed up GPS-based geo-location and is found on many smartphones
  - Obtains a quicker TTFF **–** Time to first fix – time it takes to determine geo-location by acquiring and storing information about the locaiton of satellites via cellular network.
  - aGPS utilizes GPS information from cell towers with known geo-locations to establish rough location and shorten iterative geo-location process for the GPS signal calculations

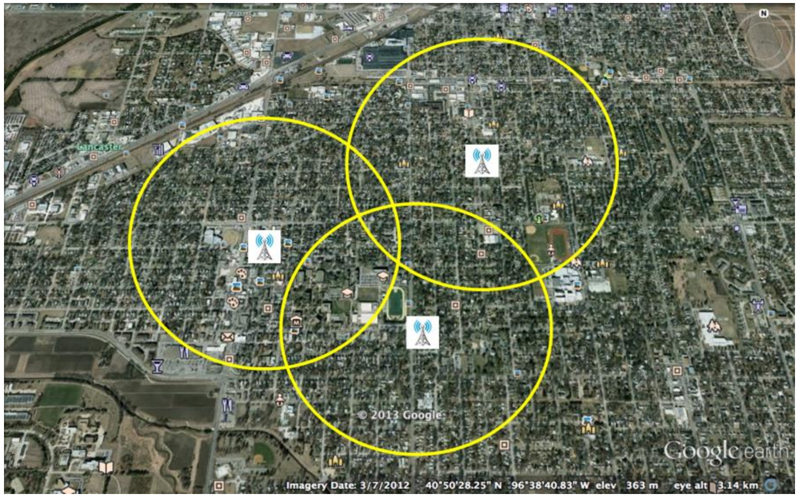
Technologies for generating Geo-location data in mobile devices

- Cell Tower Triangulation
  - Multiple cell towers with known locations
  - Cell tower signals overlap by purpose
  - More towers = better geo-location fix

How Cell Tower Triangulation Works

Accuracy: +/- 1000 meters down to 50 meters

- Geo-location using multiple cell tower signals
- Multiple pieces of information utilized
    - Signal strength from each tower
    - Known geo-location of each tower
    - When available, directional information based on which antenna segment is active



Technologies for generating Geo-location data in mobile devices

- Wifi geo-location
  - Typical 30 meter signal
  - Wifi signal detected by mobile device
  - Large databases of wifi points of presence used to ascertain fix

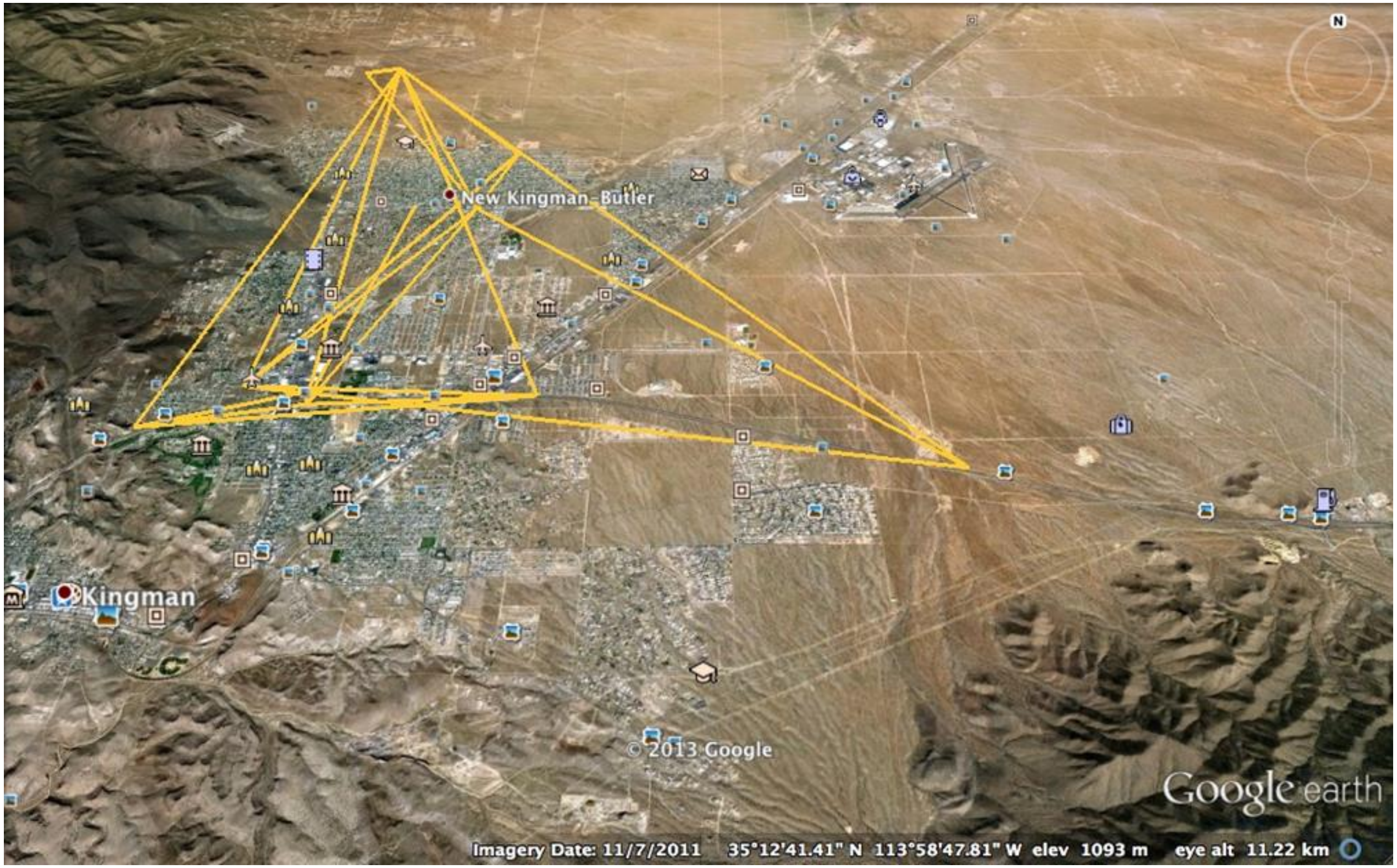
How Wifi Geo-location Works

Accuracy: +/- 100 meters down to 5 meters

- Wifi geo-location uses signals from one or more wifi sources
  - Typical sources include privately/publically owned wifi routers and access points
  - These devices use TCP/IP protocols – this means that each wifi device has a unique MAC address – no other device has the same address
  - Entities like Google used initiatives like StreetView to collect wifi MAC addresses and exact geo-location data
  - Large databases containing MAC address and geo-location allows mobile devices that encounter wifi devices to acquire fix

Illustrative Data from Passive Measurement Technology

- Sample data from mobile measurement panel utilizing passive application
- App collects GPS data from mobile device – 2 day path shown below



GPS Measurement Errors

- GPS satellite signal is line of sight – thus things like buildings are a problem
- Esoteric error sources
  - Ionosphere disturbances
  - Solar flares
  - Relativistic effects due to speed of satellite

- Common error sources
  - Multipath errors due to buildings
  - Faraday effects due to things like steel vehicle bodies
  - Clock drift on GPS-enabled devices

Example of GPS Error on Smartphone

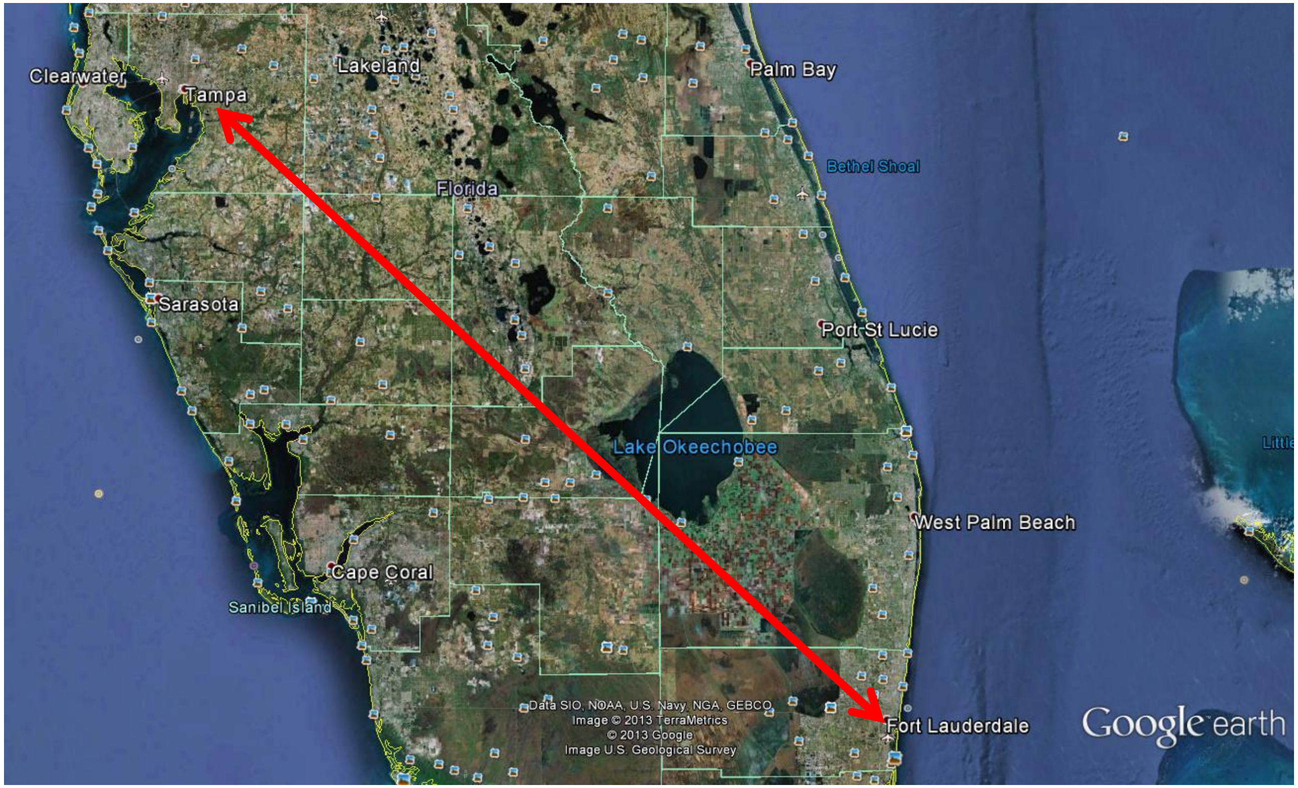
- The GPS calculated path is ~150 meters off the roadway
- Likely to be Faraday cage effect due to steel body of vehicle

Cell Tower Triangulation Error

- Scene from “Breaking Bad” burying a body? No, cell tower triangulation error
- Person likely entered building 300 meters south and lost satellite signal

Wifi Geo-location Error

- Star Trek "Transporter Effect"
- Ft Lauderdale→ Tampa → Ft Lauderdale in 90 seconds



Privacy Issues

- Significant privacy issues with geo-location data

  - Only 4 date/time/geo-location data points needed to identify single smarphone from 1.5 million smartphones[1]

  - 34.8% of U.S. adults say that they are willing to provide some personal information to a company in order to get something they want[2]

1. De Montjoye, Y., Hidalgo, C., Verleysen, M. and Blondel, V., 2013.
2. Experian Simmons Fall 12 Month National Consumer Study

Summary

- Three current geo-location technologies – often used in concert
- Each of these technologies has inherent measurement error sources
- Geo-location technologies and data present non-trivial privacy issues

Contact

- Max Kilger – max.kilger@experian.com
- TraShawna Boals – trashawna.boals@experian.com

Cell Phone Revolution

Communication Revolution

New Challenges

New Opportunities