

PAPOR DECEMBER 6, 2018

SHORT COURSE I: OPTIMIZING SAMPLING FRAMES

ASHLEY H. HYON

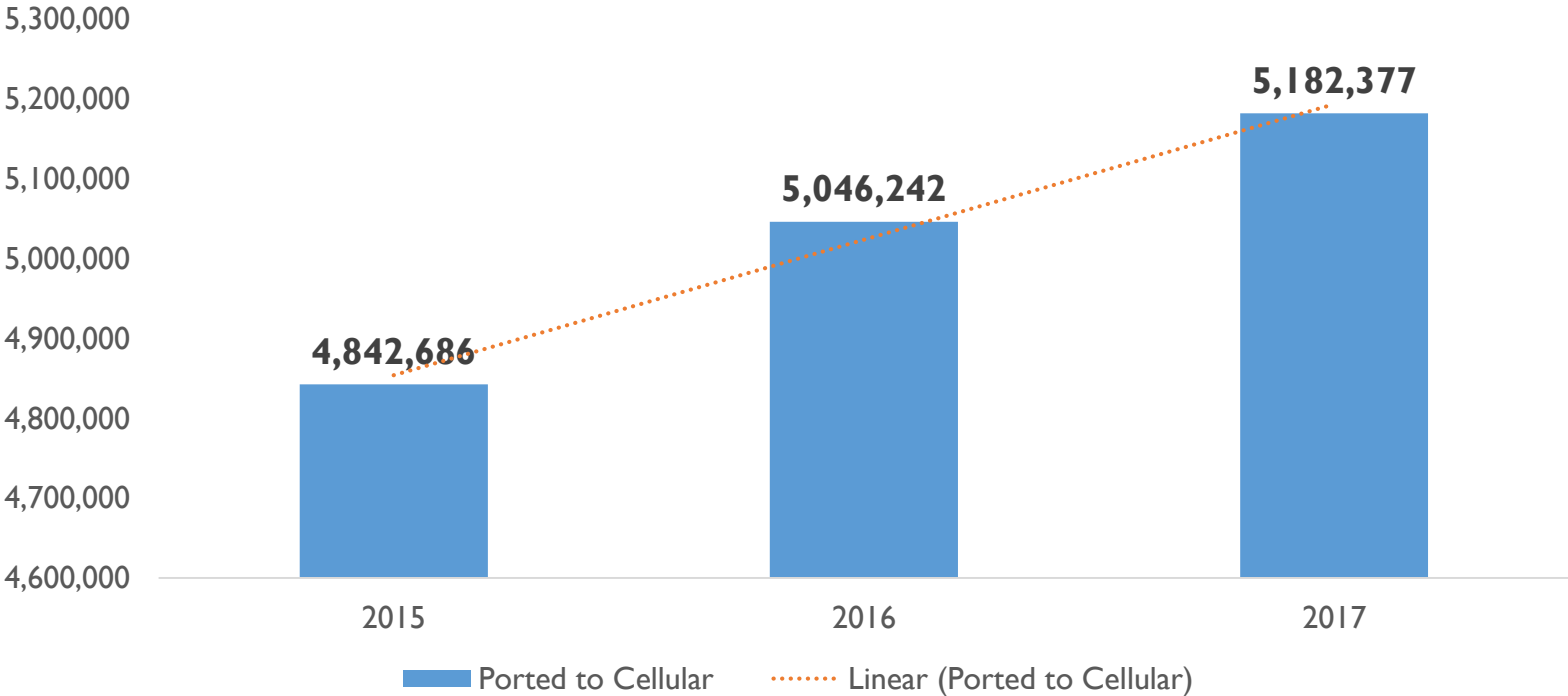
Outline

- Sampling Frames
 - Cellular RDD
 - ABS
 - Convenience Consumer Frames
 - On-Line Frames
- Hybrid Designs – Mix Bag Approaches
 - Increasing Feasibility
 - Expanding the Universe (multi-frame)
 - Increasing Efficiency
 - Demographic Density
 - Incorporating Auxiliary Variables
 - Expanded Capabilities
 - Increasing Response Rates
 - Additional Modes of Contact
 - Tailored Approach



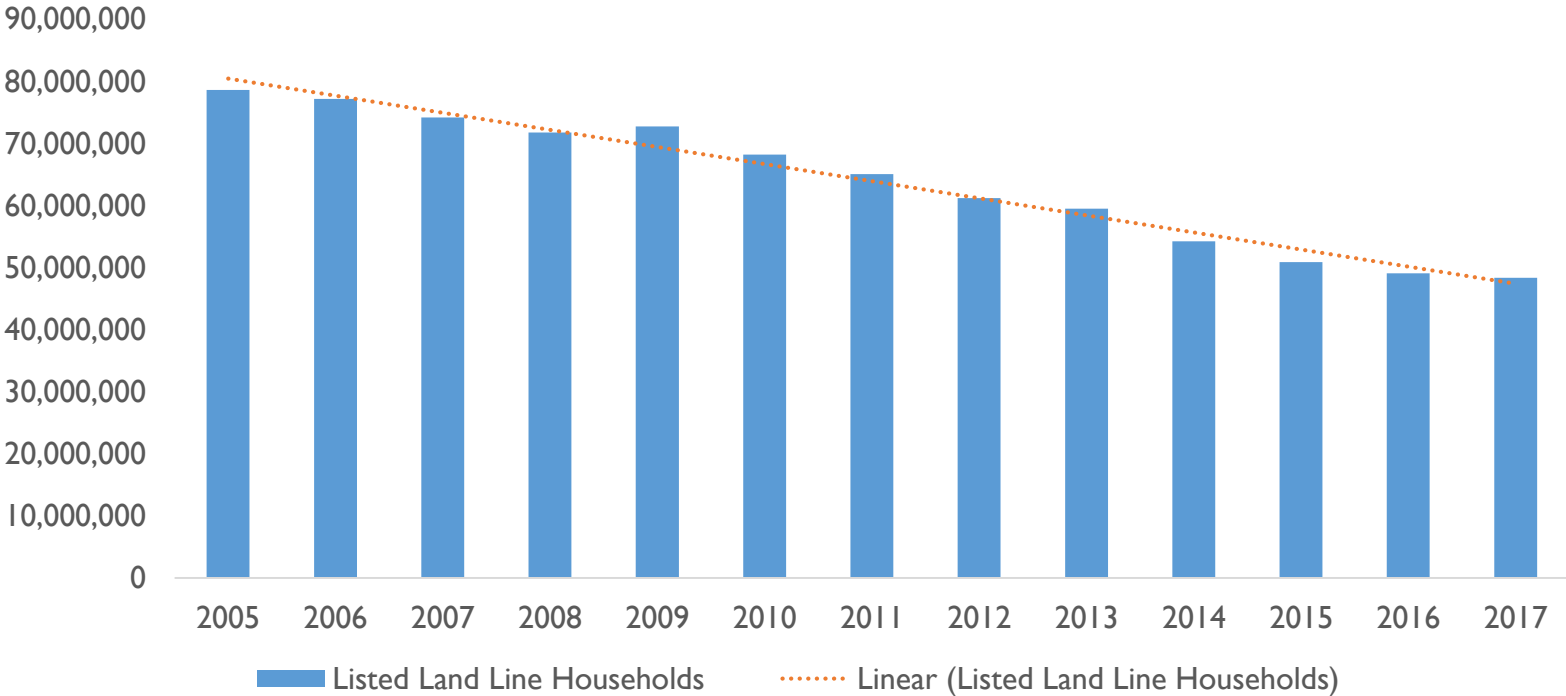
Sampling Frames – Cellular RDD

Land Line Ports To Wireless

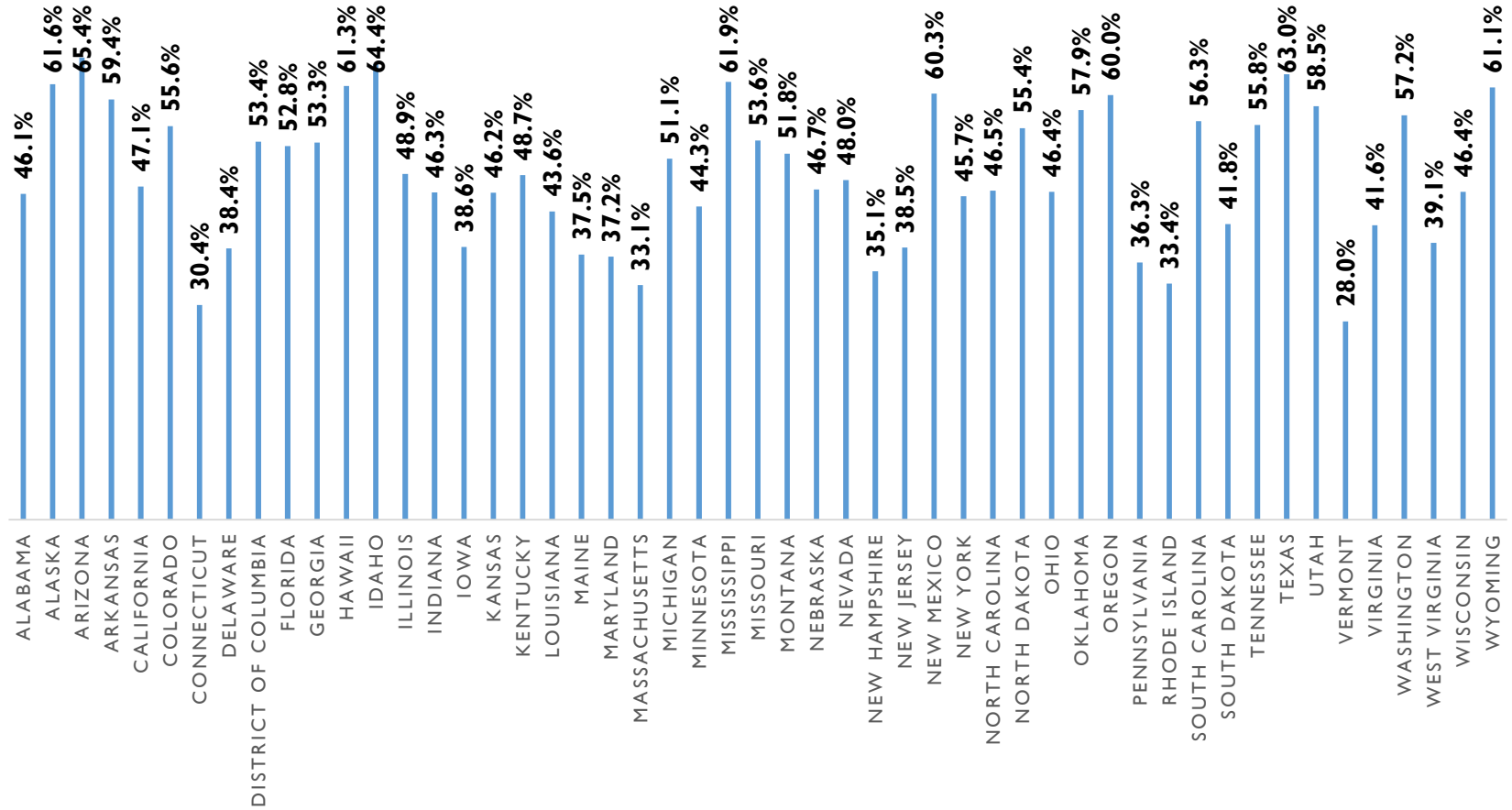


Sampling Frames – Cellular RDD

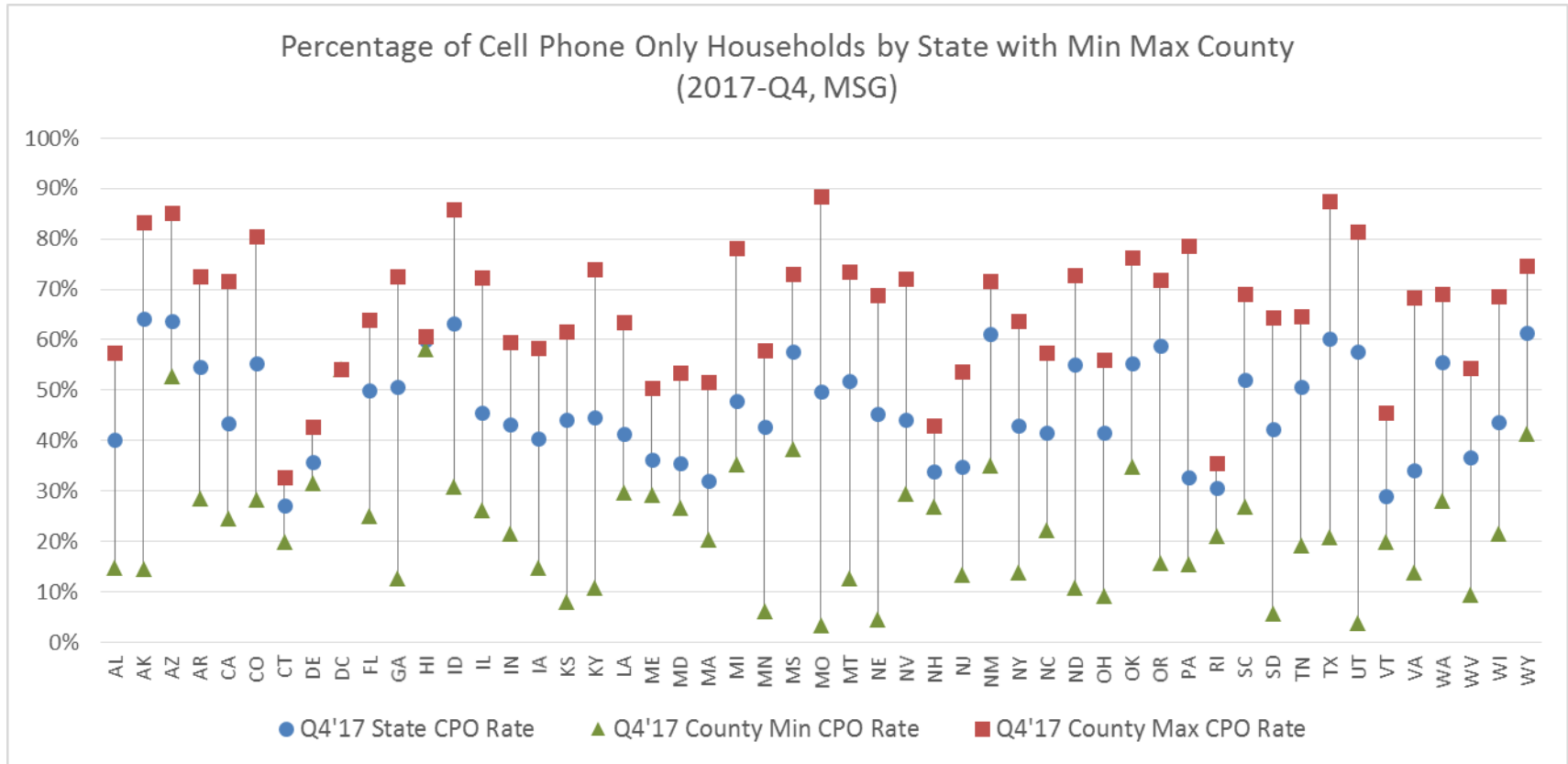
Listed Land Line Households



Sampling Frames – Cellular RDD

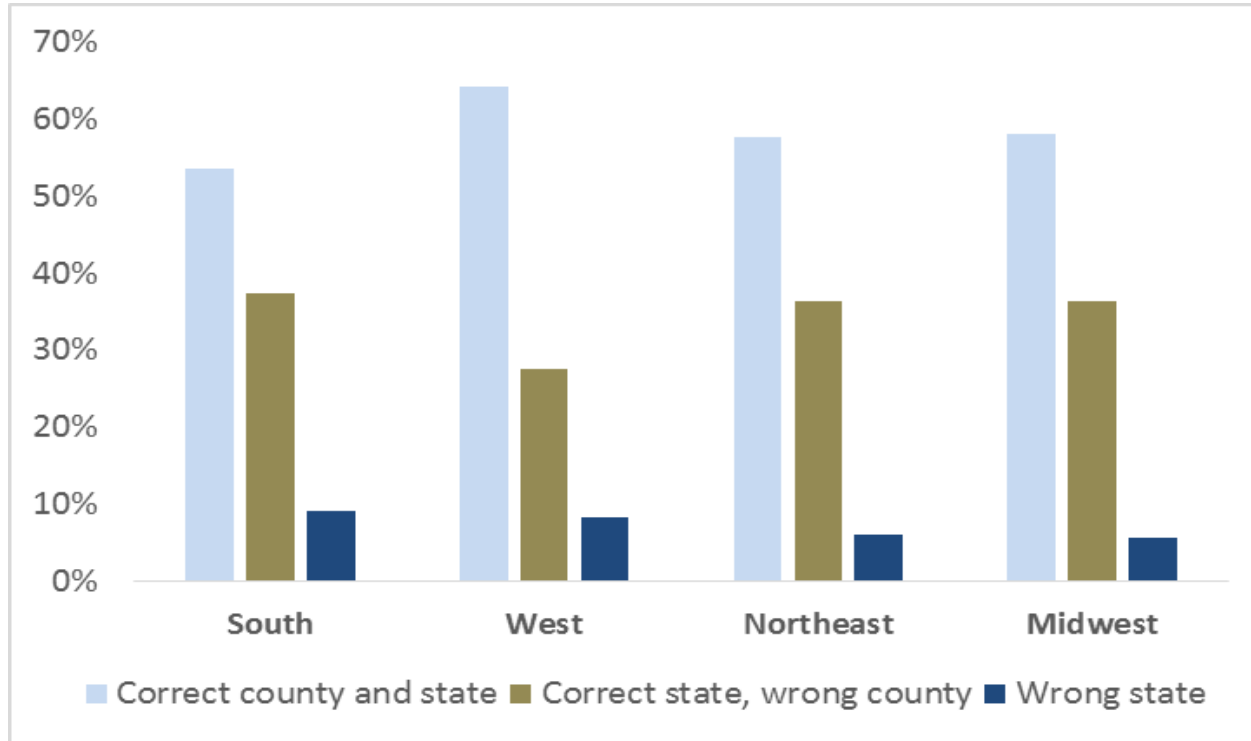


Sampling Frames – Cellular RDD



Sampling Frames – Cellular RDD

ACCURACY GROUPS BY CENSUS REGIONS



Citation: Data from AAPOR 2018 Presentation: “Sample and Respondent Provided County Comparisons Among Cellular Respondents. Authored by Carol Pierannunzi, Senior Survey Methodologist CDC et al.

Sampling Frames – Cellular RDD

- Rate Center is the basic unit of Geography for Land Line & Cellular Thousand Blocks
- Rate Center Boundaries Updated Quarterly
 - Defined by 2010 Census Blocks
 - Each US Census Block assigned uniquely to a single Rate Center
 - Boundaries provide more flexibility when stratifying by geography smaller than state
- Rate Center only identifies where the Cellular IK Block is “homed”
- Cellular RDD Universe = 511,792,000



Sampling Frames – Cellular RDD

DEMOGRAPHIC TARGETING BY DENSITY

Rate Center Geographic Boundaries

are linked to census geography →

Census-based demographic sources

are then used to generate Rate

Center specific demographic profiles

that can be used for stratification or

oversampling.

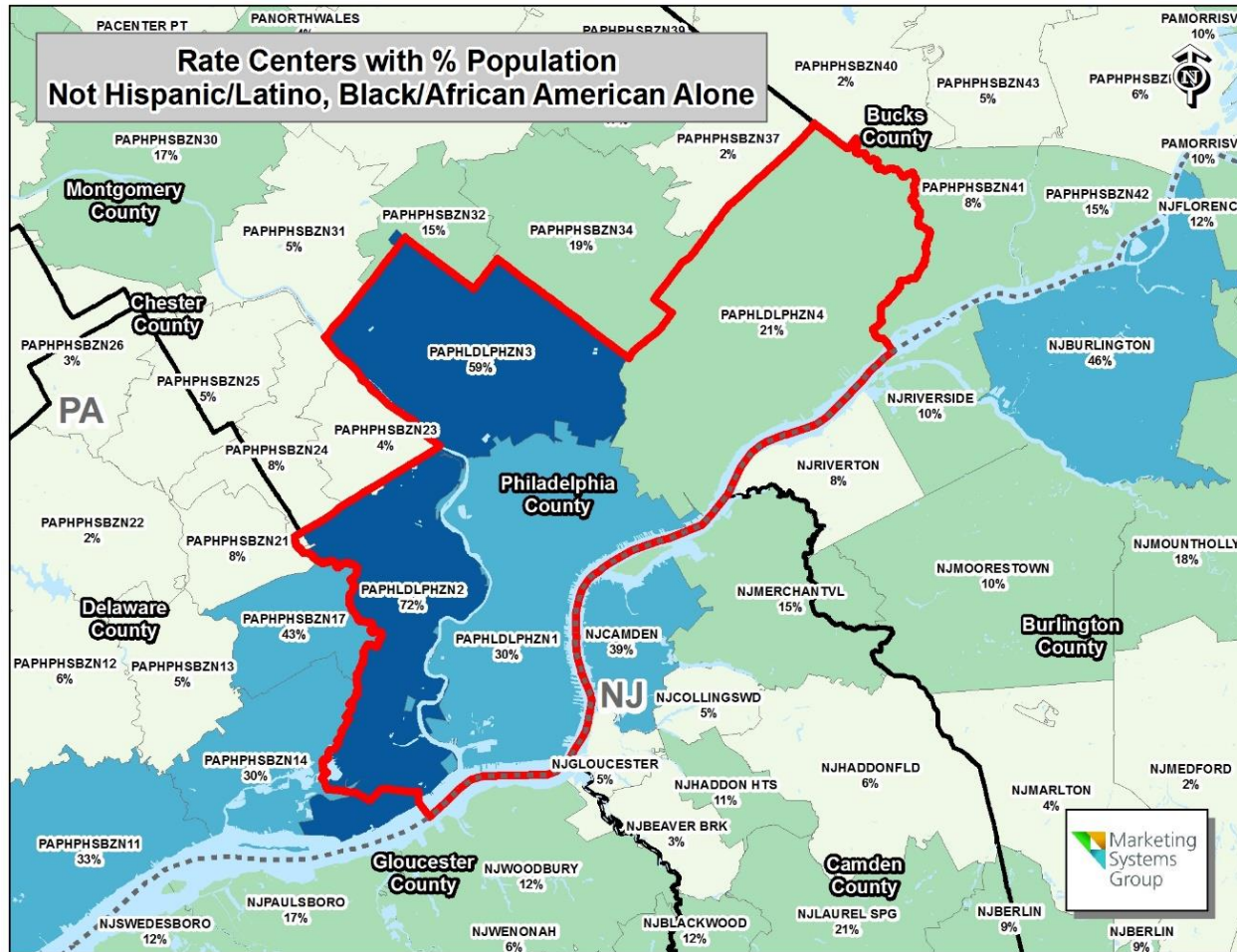
- Current Year Households
- Current Year Population - Gender
- Current Year Population - Race
- Current Year Population - Age
- Current Year Household Income

Sampling Frames – Cellular RDD

RATE CENTER DEMOGRAPHIC DENSITY REPORT

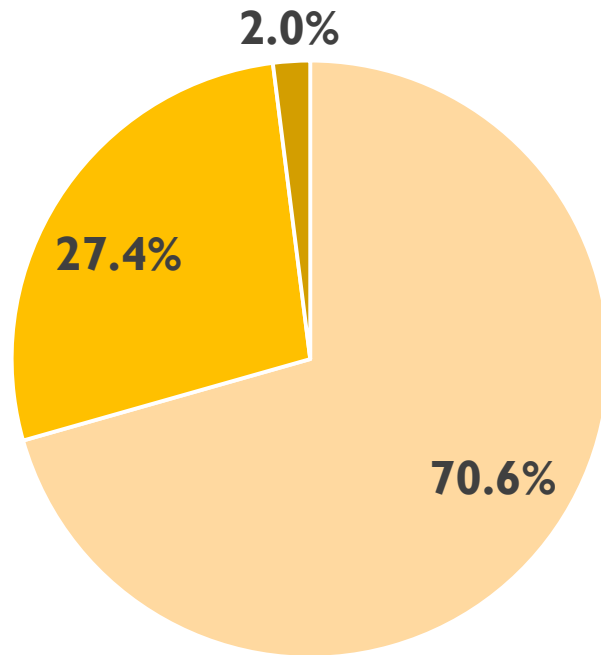
Rate Center ID	Primary FIPS Code	Total Housing Units	Number of Cellular 1k Blocks	Total Population (2010)	% Inclusion (Population)	Pop Total	Pop NHS Black	% NHS Black
PAPHLDLPHZN1	42101	237,091	1,876	495,900	100%	523,812	159,658	30.48%
PAPHLDLPHZN2	42101	120,636	931	272,015	100%	279,191	199,799	71.56%
PAPHLDLPHZN3	42101	135,601	167	307,888	100%	312,281	184,688	59.14%
PAPHLDLPHZN4	42101	176,850	171	449,428	100%	461,217	95,965	20.81%
PAPPHSBZN23	42091	10,266	139	25,998	3.05%	25,732	1,107	4.30%

Sampling Frames – Cellular RDD



Sampling Frames – Cellular RDD

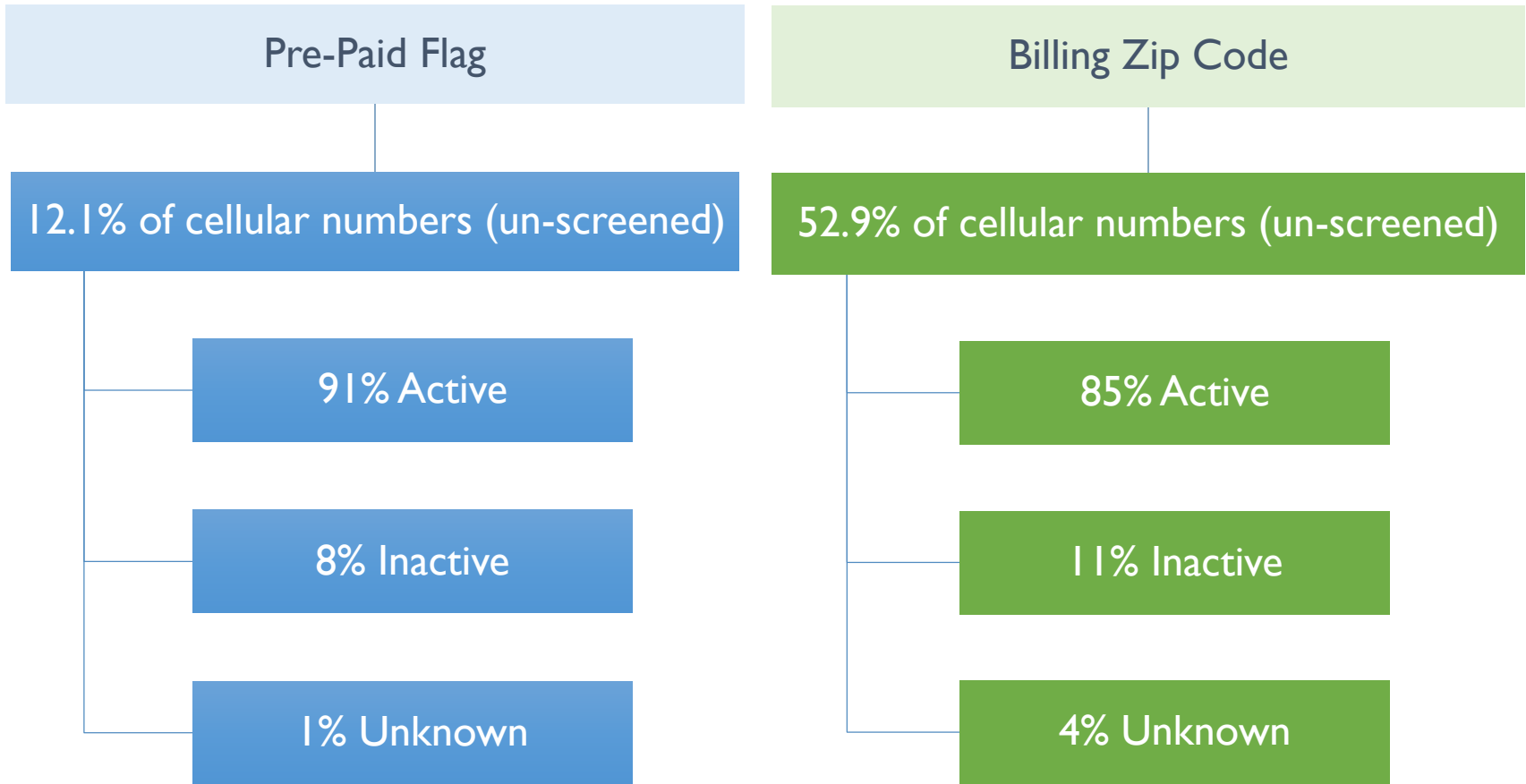
Screening for Activity Cell-WINS



■ Active ■ In-active ■ Unknown

- Determines whether an individual telephone number is Active, Inactive or Unknown at that point in time.
- Real time database queries to various US Telephony databases.
- FCC compliant – the individual telephone numbers do not ring through to the device

Sampling Frames – Cellular RDD



Sampling Frames – ABS

- Address based sampling is a viable sampling alternative when complete coverage of the household population is desired
- Strengths of using this method for sample surveys:
 - Single Frame Approach to cover all types of HHs - Weighting calculations are simplified
 - Auxiliary information is available for every address on the frame (Census data, Claritas, others)
 - Very useful for sampling design/planning/allocation
 - Greater control of geographic level of interest (down to Zip+4)
 - Can support mixed mode survey designs from the start

Sampling Frames – ABS

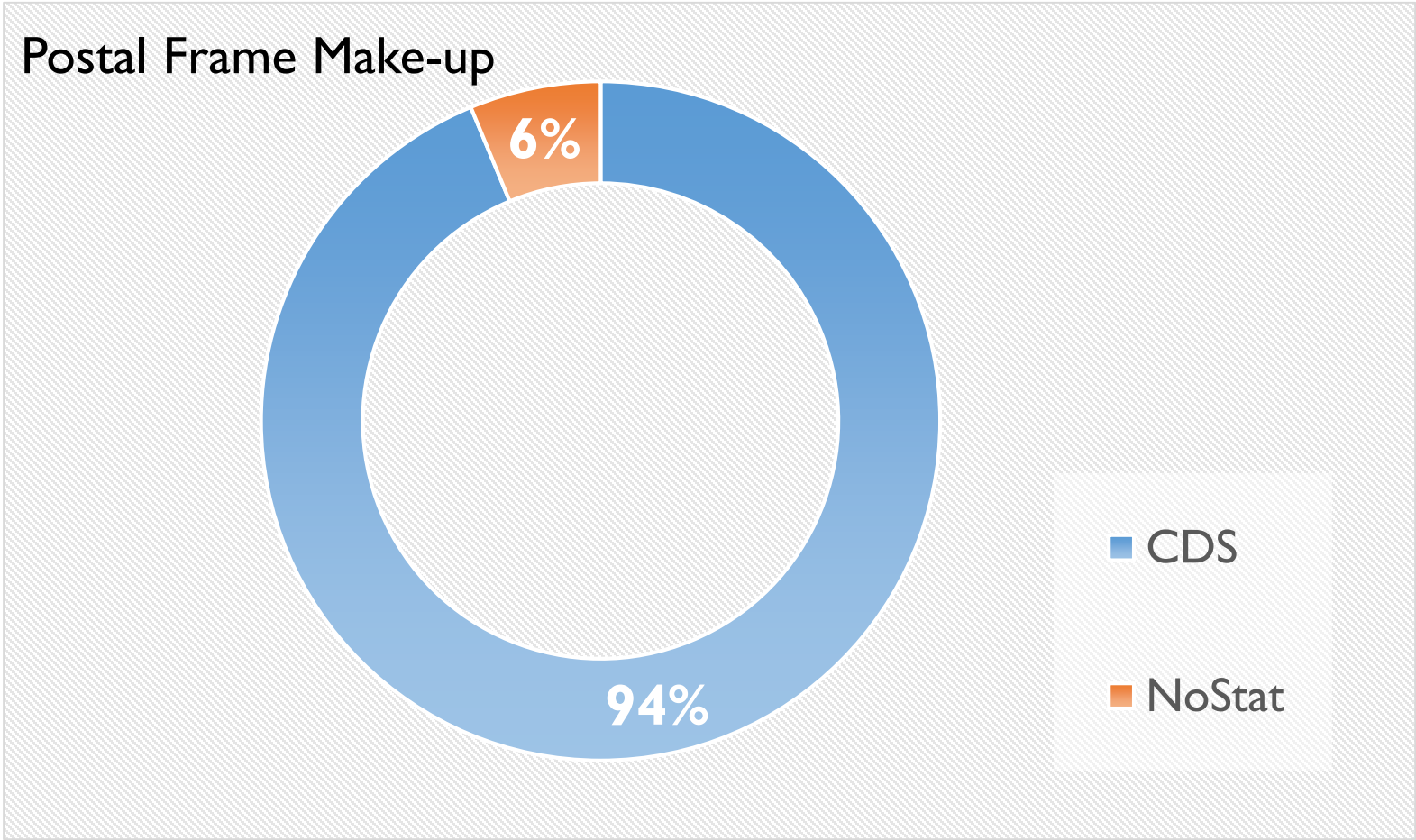
- Based on USPS CDS File of 146 million active residential addresses
- Complimented with USPS NoStat File (inactive)
- Address-frame covers over 99% of all households
- Reconstructed monthly
- Augment simplified addresses – where possible
- Augment drop units – where possible
- Every address geo-coded by MSG (Lat/Long)
- Append 2000 and 2010 Census Geography

Sampling Frames – ABS

- USPS Computerized Delivery Sequence File (CDS)
 - Snapshot of the official master USPS database of active mailing addresses
 - All addresses are Delivery Point Verified (DPV)
 - Includes all LACS converted addresses
 - Meets all CASS certification requirements
 - Requires no additional address hygiene processing

- USPS NoStat File
 - USPS database of addresses currently not receiving mail
 - Includes new/planned housing developments
 - Includes vacant delivery points on rural routes
 - Internal drop addresses

Sampling Frames – ABS



Sampling Frames – ABS

ABS frame count

Address Type	Count
CityStyle (<i>street address</i>)	123,896,531
Rural Routes/Highway Contracts	63,824
POBox	14,235,673
POBox (<i>Only Way to Get Mail/OWGM</i>)	1,415,853
Seasonal (<i>vacation/second homes</i>)	821,596
Educational (<i>off-campus housing</i>)	98,031
Vacant (<i>long term 90 day vacancy</i>)	2,960,792
Throwback (<i>mail being forwarded to PO Box</i>)	201,174
DropPoints (<i>multi-dwellings with no unit information – one central mail drop off</i>)	713,739
DropUnits	2,013,366
DropUnitAugments	41,438
CityStyleAugments	62,731
POBOXAugments	14,569
Simplified Addresses	102,769

Sampling Frames – Convenience Consumer

➤ Sources:

- Land line sampling frames are developed using information known from HHs that have an identified and listed landline number.
- Consumer cellular frames are based on credit/purchasing information where a secondary source is used to append cellular numbers
- Voter file frames made up land line and cellular numbers

➤ Geographic capability – down the Census Block

➤ Demographic Targeting at the household and individual levels

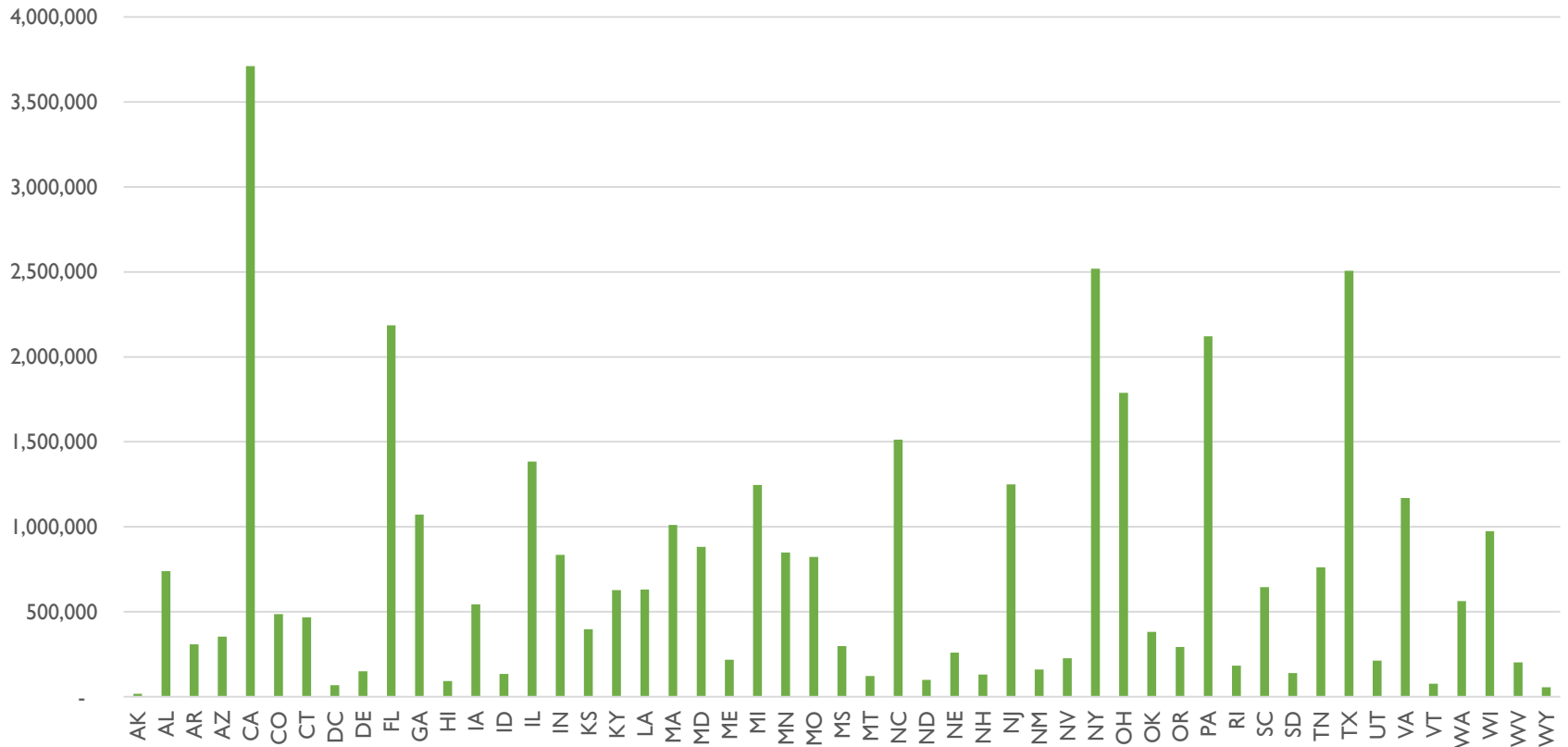
➤ Updated monthly/Bi-monthly/quarterly

Sampling Frames – Convenience Consumer

- How can these frames be utilized?
 - Sample into areas down to Census Block
 - Individual/household level demographic/psychographic targeting
 - Disproportionate Stratified Sampling Methodologies
 - Over sampling
 - Appending additional modes of contact
 - Capture some of the in-migration respondents that are missed in a Cellular RDD Frame for a given area

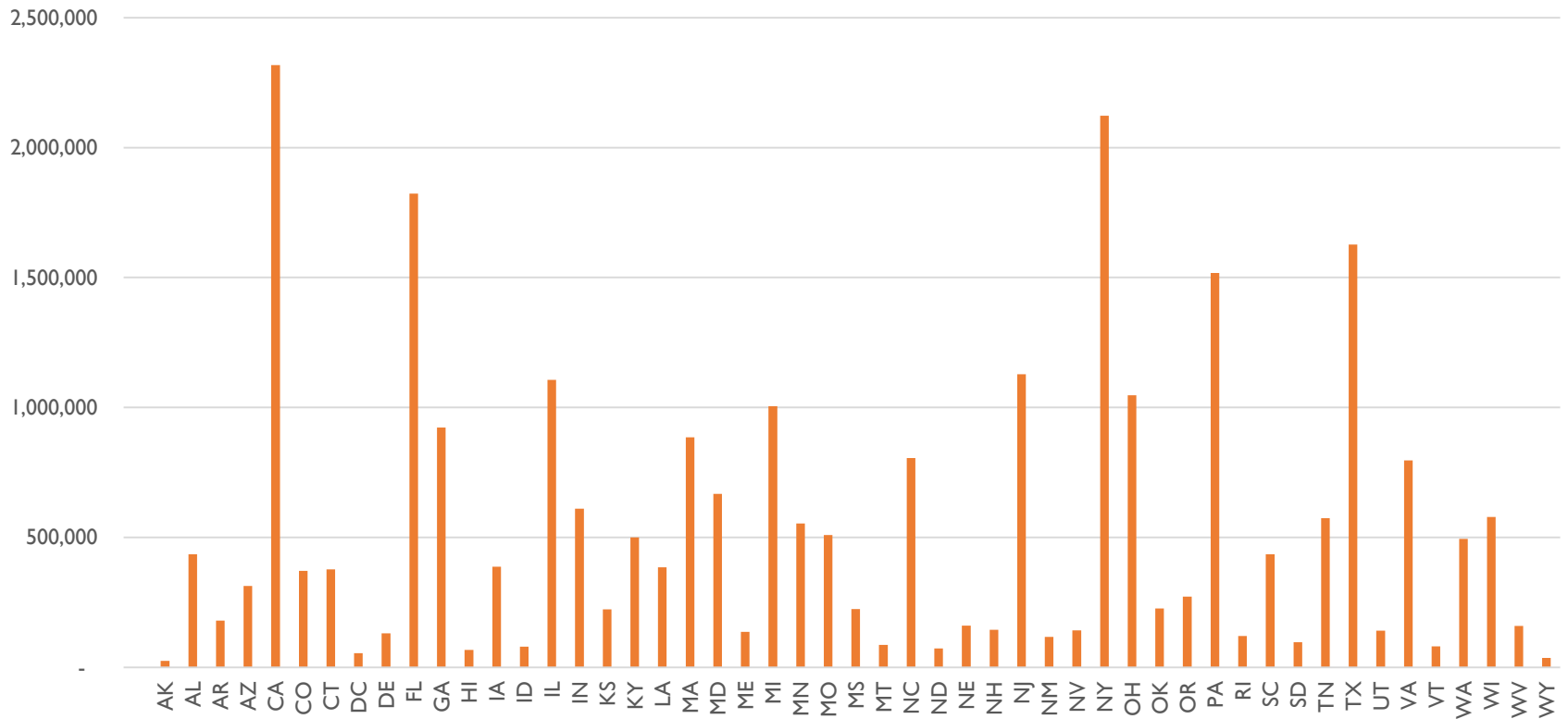
Sampling Frames – Convenience Consumer

LISTED LAND LINE COUNTS UNIQUE PHONE NUMBER PER HOUSEHOLD = 37,900,150



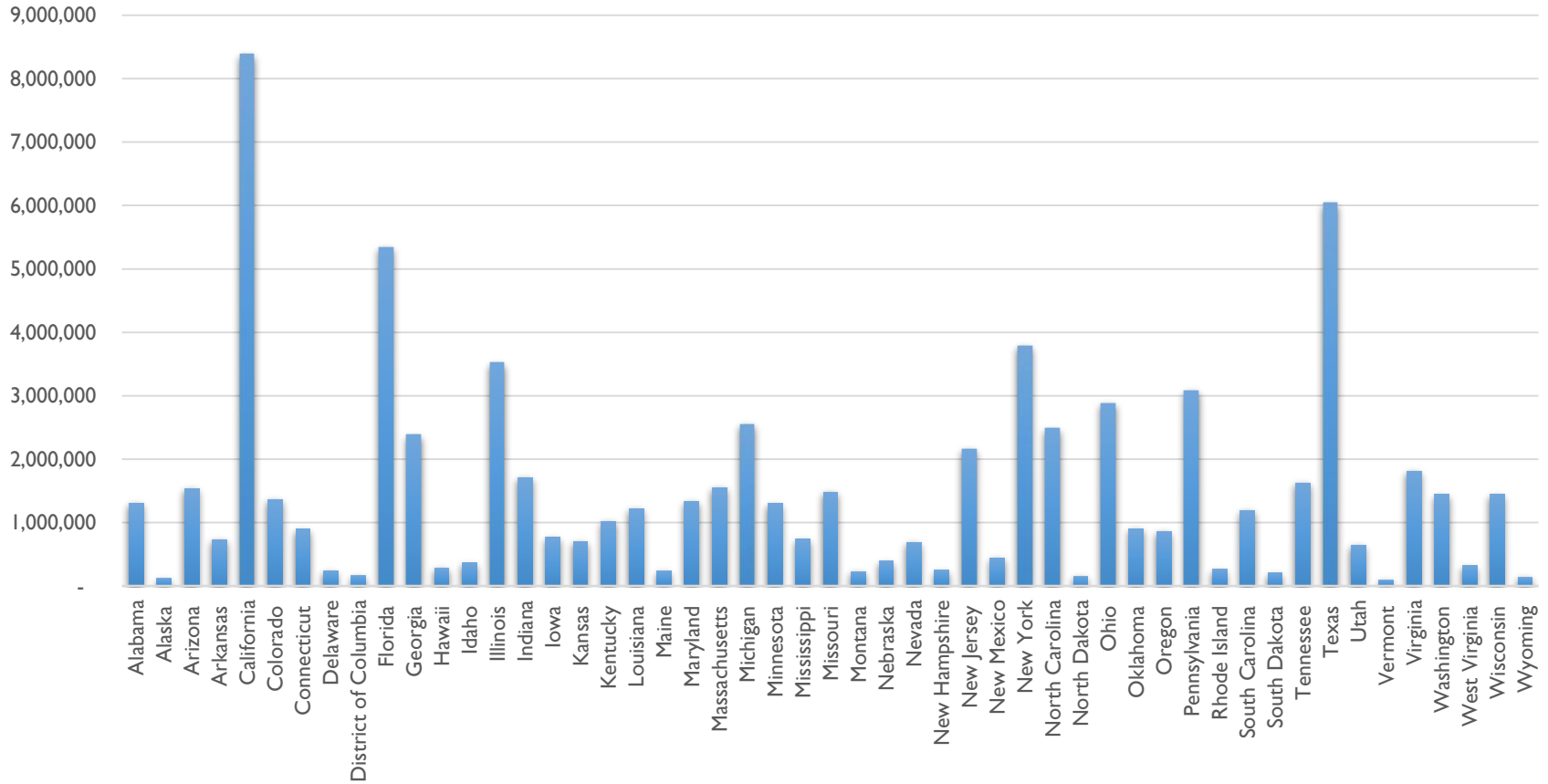
Sampling Frames – Convenience Consumer

VOTER FILE COUNT - UNIQUE LAND LINE NUMBER PER HOUSEHOLD = 27,266,926



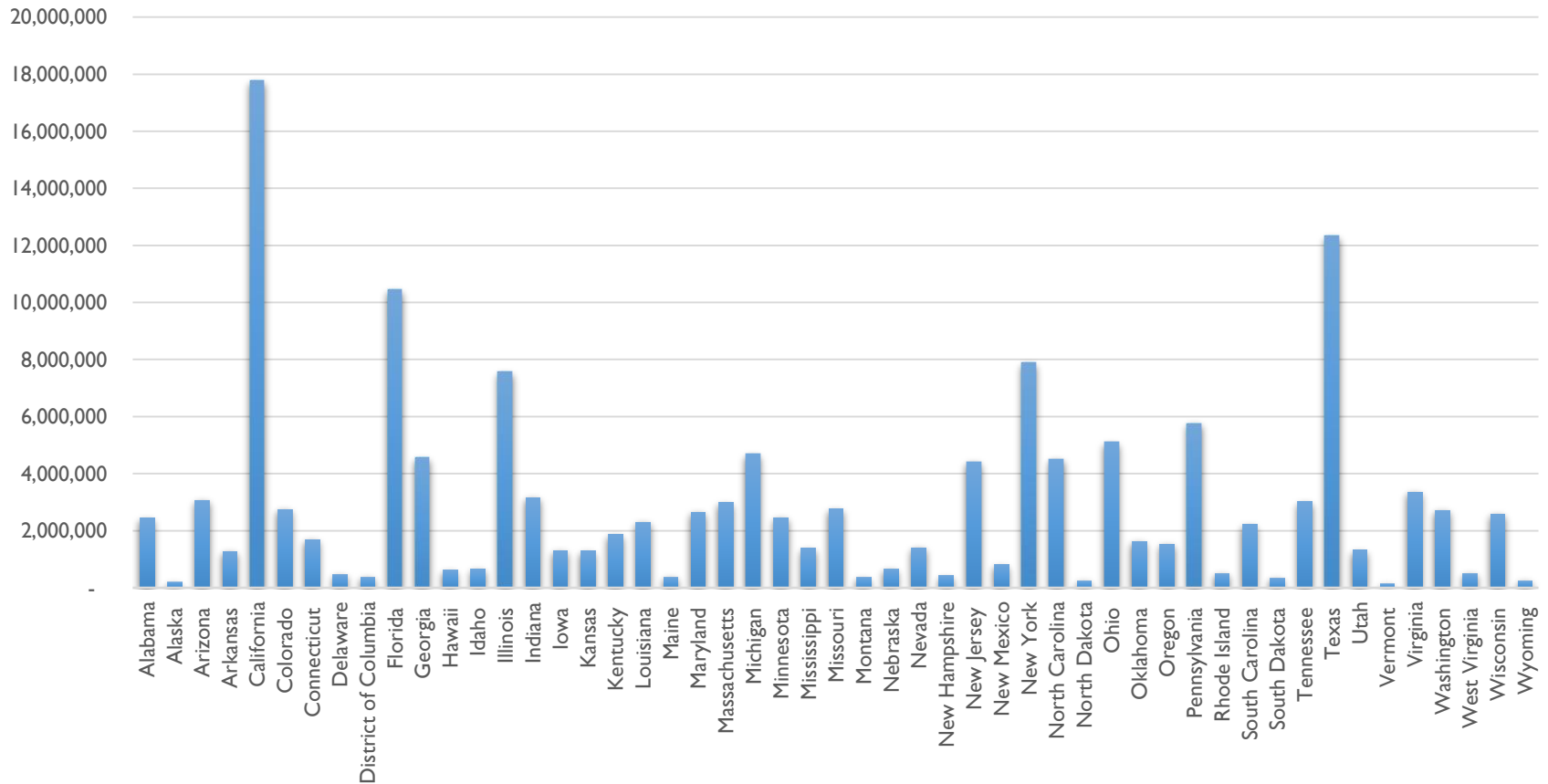
Sampling Frames – Convenience Consumer

CONSUMER CELLULAR - ONE PER HOUSEHOLD = 74,820,165



Sampling Frames – Convenience Consumer

CONSUMER CELLULAR - ALL PER HOUSEHOLD = 145,192,009



Sampling Frames – On-line

- Non-Probability Recruited On-line Panels: Unequal and unknown chance of selection
 - Cheaper and faster
 - Numerous panel providers
 - Specialty Panels
 - Business (B2B)
 - Ethnic
 - Ailments
 - Medical Professionals
 - Millennials
 - Teens
 - Tradesmen
 - Gamers

Sampling Frames – On-line

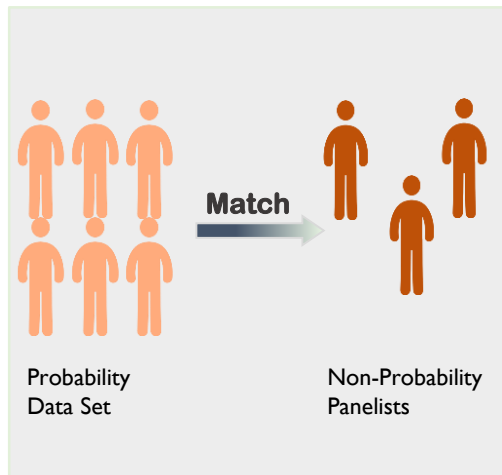
- Probability Recruited On-line Panels: Equal and known chance of selection
 - Sample quality and faster than phone or mail
 - Not many panel providers
 - Knowledge Panel
 - Amerispeak
 - Others you need to run the whole study through them or for internal use only (i.e. Gallup Panel or American Trends Panel)
 - Very expensive to recruit
 - Limited specialty panels

Increasing Feasibility

- Sampling from multiple sample frames to increase feasibility
 - Supplementing on-line surveys with off-line line frames:
 - Small areas where there is insufficient sample
 - Hard to reach populations
 - Expanded universe
 - Supplementing off-line frames with on-line frames:
 - Increase the number of younger respondents or sub populations
 - Reduce data collection cost
 - Shorten field time

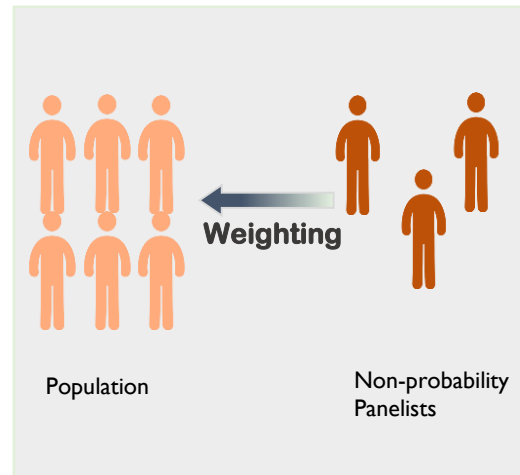
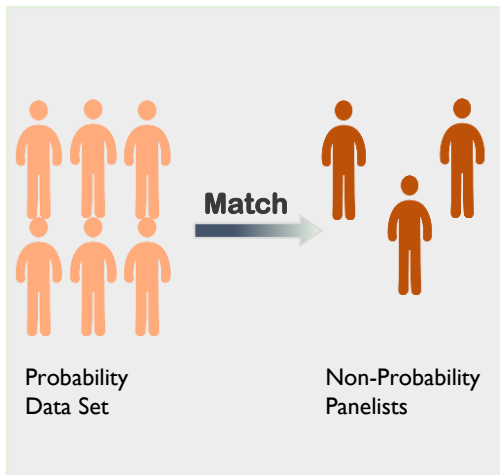
Increasing Feasibility

- Developing a technique to get “representative like sample” from a non-probability panels utilizing a high quality probability dataset as a matching reference and matching based on demographic, psychographic and behavioral variables. The completed surveys will then be weighted producing a pseudo probability on-line sample



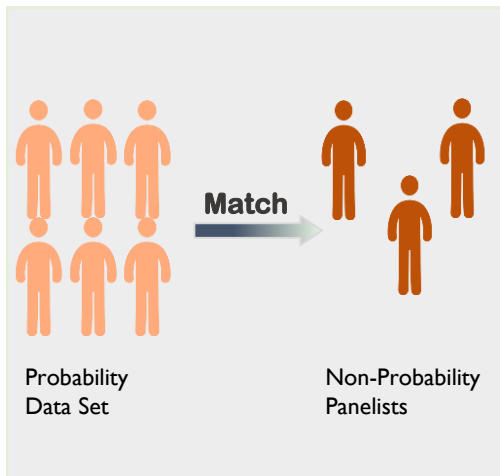
Increasing Feasibility

- Developing a technique to get “representative like sample” from a non-probability panels utilizing a high quality probability dataset as a matching reference and matching based on demographic, psychographic and behavioral variables. The completed surveys will then be weighted producing a pseudo probability on-line sample



Increasing Feasibility

- Developing a technique to get “representative like sample” from a non-probability panels utilizing a high quality probability dataset as a matching reference and matching based on demographic, psychographic and behavioral variables. The completed surveys will then be weighted producing a pseudo probability on-line sample



Increasing Efficiency – Demographic Density

➤ Public Data Sources

- CENSUS 2010 SFI (Summary File)
 - Geography is a perfect fit at the Block level
 - Estimates are now dated
 - Not very many variable options other than Total Pop, HHs and HUs.
- Claritas Current Year (2018)
 - Updated Yearly
 - More variables are available
- CENSUS ACS 5 Year (2012 - 2016 released soon)
 - Updated Yearly
 - Has the most # of variables of all sources
 - Can provide more variables for an age 18+ Universe (including educational attainment)

Increasing Efficiency – Demographic Density

TITLE: Population 5+: Speak Spanish at Home

Source: Claritas 2018.1

Geography: Census Block Groups

Percent:	Geography:	Population Age 5+		Population 5+: Speak Spanish at Home		Incidence		Coverage	Total Households	
		Individual	Cumulative	Individual	Cumulative	Individual	Cumulative		Individual	Cumulative
100_100	9	369	369	369	369	100.0	100.0	0.0	126	126
095_100	291	588,897	589,266	568,463	568,832	96.5	96.5	1.4	178,864	178,990
090_095	842	1,441,434	2,030,700	1,330,567	1,899,399	92.3	93.5	4.7	427,377	606,367
085_090	904	1,484,588	3,515,288	1,299,852	3,199,251	87.6	91.0	7.9	439,550	1,045,917
080_085	1,039	1,754,177	5,269,465	1,444,940	4,644,191	82.4	88.1	11.4	520,500	1,566,417
075_080	1,188	1,873,907	7,143,372	1,452,537	6,096,728	77.5	85.3	15.0	567,685	2,134,102
070_075	1,343	2,091,970	9,235,342	1,514,842	7,611,570	72.4	82.4	18.7	638,974	2,773,076
065_070	1,733	2,690,335	11,925,677	1,816,905	9,428,475	67.5	79.1	23.2	850,554	3,623,630
060_065	1,966	3,047,219	14,972,896	1,905,169	11,333,644	62.5	75.7	27.8	982,492	4,606,122
055_060	2,255	3,408,932	18,381,828	1,959,351	13,292,995	57.5	72.3	32.6	1,141,609	5,747,731
050_055	2,413	3,680,200	22,062,028	1,931,899	15,224,894	52.5	69.0	37.4	1,263,550	7,011,281
045_050	2,731	4,153,979	26,216,007	1,969,643	17,194,537	47.4	65.6	42.2	1,438,280	8,449,561
040_045	2,982	4,596,154	30,812,161	1,952,255	19,146,792	42.5	62.1	47.0	1,643,061	10,092,622
035_040	3,601	5,684,110	36,496,271	2,124,929	21,271,721	37.4	58.3	52.2	2,097,246	12,189,868
030_035	4,244	6,475,885	42,972,156	2,095,356	23,367,077	32.4	54.4	57.4	2,410,652	14,600,520
025_030	5,437	8,339,153	51,311,309	2,286,885	25,653,962	27.4	50.0	63.0	3,203,961	17,804,481
020_025	7,508	11,454,527	62,765,836	2,565,806	28,219,768	22.4	45.0	69.3	4,459,694	22,264,175
015_020	10,624	16,253,066	79,018,902	2,810,259	31,030,027	17.3	39.3	76.2	6,497,935	28,762,110
010_015	17,116	25,871,249	104,890,151	3,176,643	34,206,670	12.3	32.6	84.0	10,554,490	39,316,600
005_010	34,057	50,146,566	155,036,717	3,576,412	37,783,082	7.1	24.4	92.8	20,832,177	60,148,777
000_005	108,146	144,737,201	299,773,918	2,937,074	40,720,156	2.0	13.6	100.0	60,882,835	121,031,612
000_000	6,753	6,746,123	306,520,041	0	40,720,156	0.0	13.3	100.0	2,911,348	123,942,960

Increasing Efficiency – Auxiliary Variables

➤ Basic Demographics

- Head of Household name
- Telephone number
- Age/Gender
- Surname
- Income
- Ethnicity
- Education
- Own/Rent
- Presence of Children
- Marital Status
- Own/Rent



Increasing Efficiency – Auxiliary Variables

➤ Income

Income	Codes
	A=\$1,000-\$14,999
	B=\$15,000-\$24,999
	C=\$25,000-\$34,999
	D=\$35,000-\$49,999
	E=\$50,000-\$74,999
	F=\$75,000-\$99,999
	G=\$100,000-\$124,999
	H=125,000-\$149,999
	I=\$150,000-\$174,999
	J=175,000-\$199,999
	K=\$200,000-\$249,999
	L=\$250,000+
	U=Unknown

Increasing Efficiency – Auxiliary Variables

➤ Education

Education	1 = High School Diploma
	2 = Some College
	3 = Bachelor Degree
	4 = Graduate Degree
	5 = Less Than High School Diploma

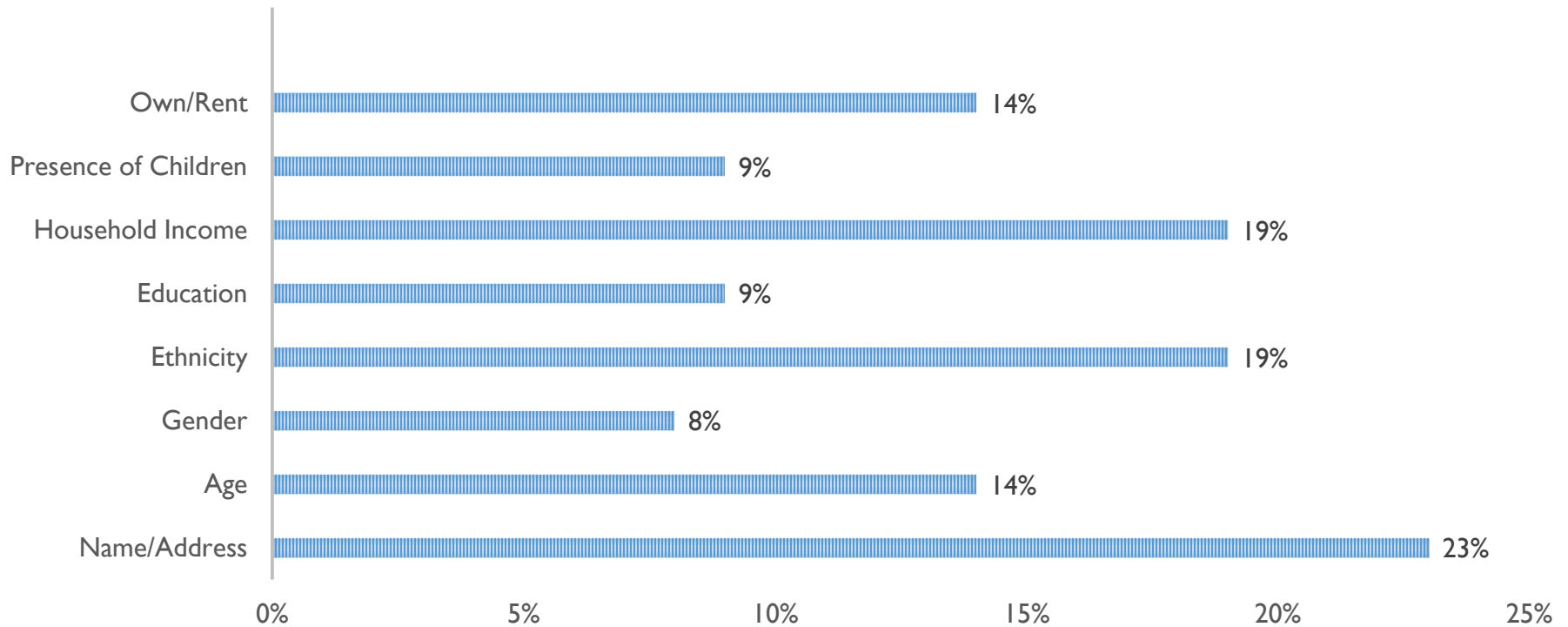
Increasing Efficiency – Auxiliary Variables

➤ Ethnicity

Ethnicity	1 = African	18 = Jewish
	2 = African American	19 = Korean
	3 = Asian Other	20 = Middle Eastern
	4 = Chinese	21 = Miscellaneous Other
	5 = Czech	22 = Native American
	6 = Dutch	23 = Polish
	7 = Eastern European	24 = Polynesian
	8 = English	25 = Portuguese
	9 = French	26 = Russian
	10 = German	27 = Scandinavian
	11 = Greek	28 = Scotch
	12 = Hawaiian	29 = Swiss
	13 = Hispanic	30 = Ukrainian
	14 = Indonesian	32 = Vietnamese
	15 = Irish	33 = Western European
	16 = Italian	31 = Unknown
	17 = Japanese	

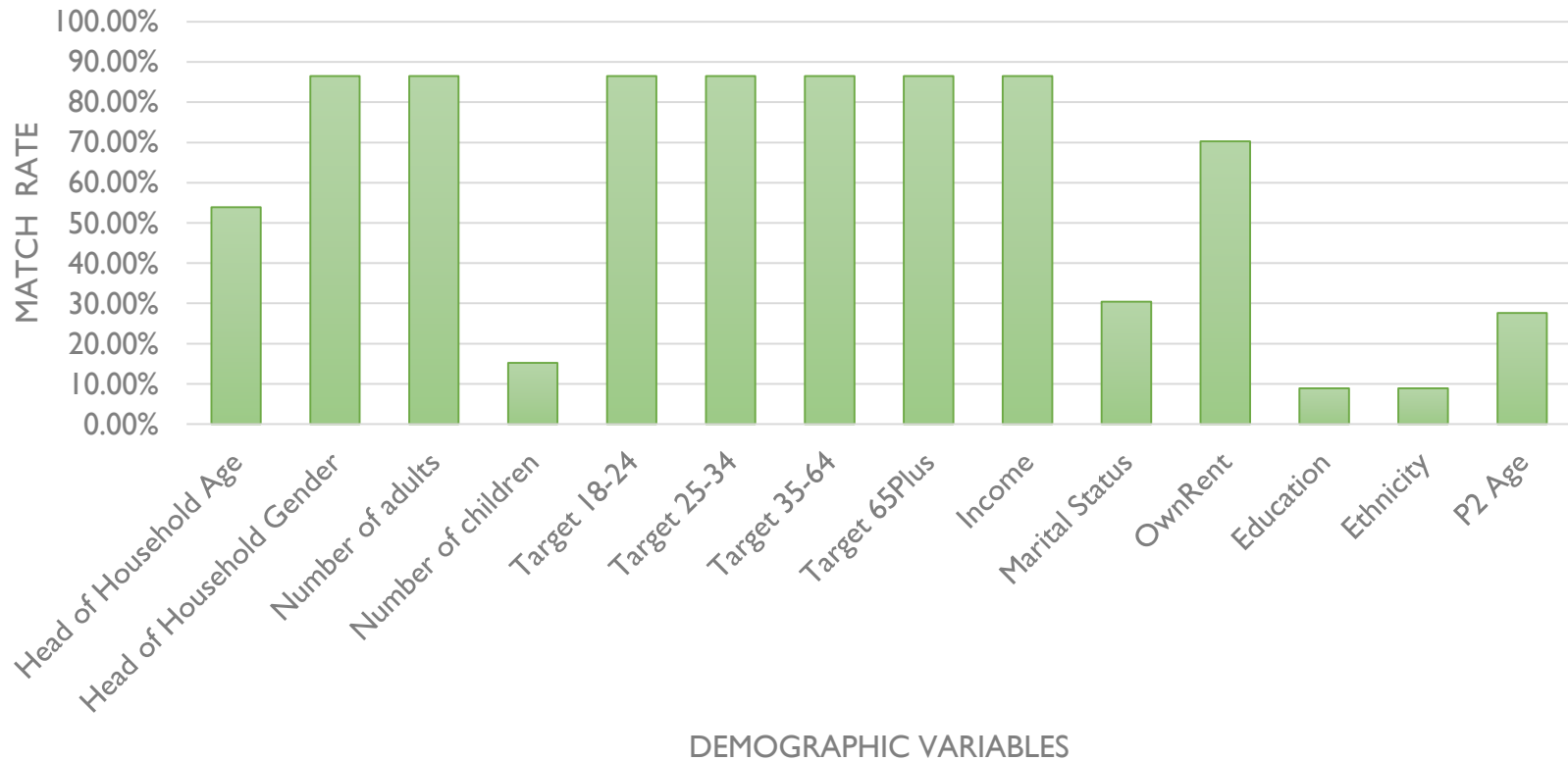
Increasing Efficiency – Auxiliary Variables

Cellular RDD to Consumer Cellular Match Rates



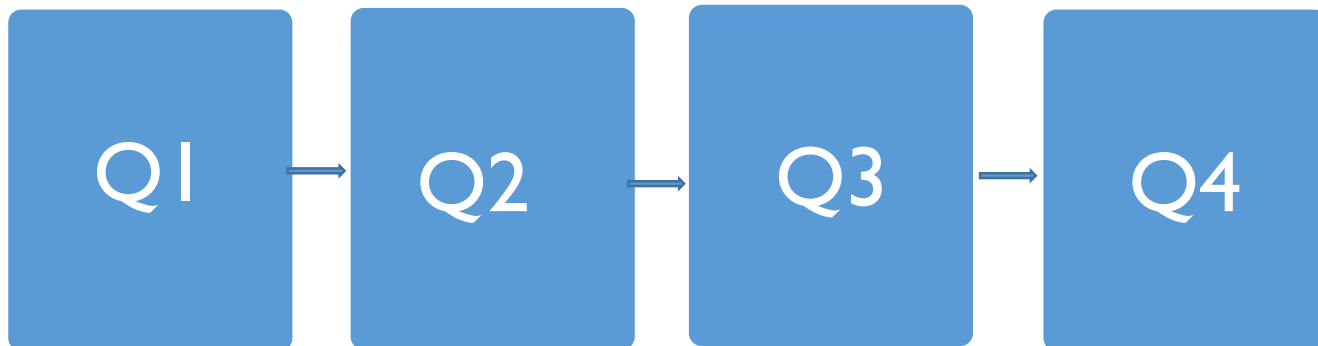
Increasing Efficiency – Auxiliary Variables

ABS to Listed Land Line Match Rates



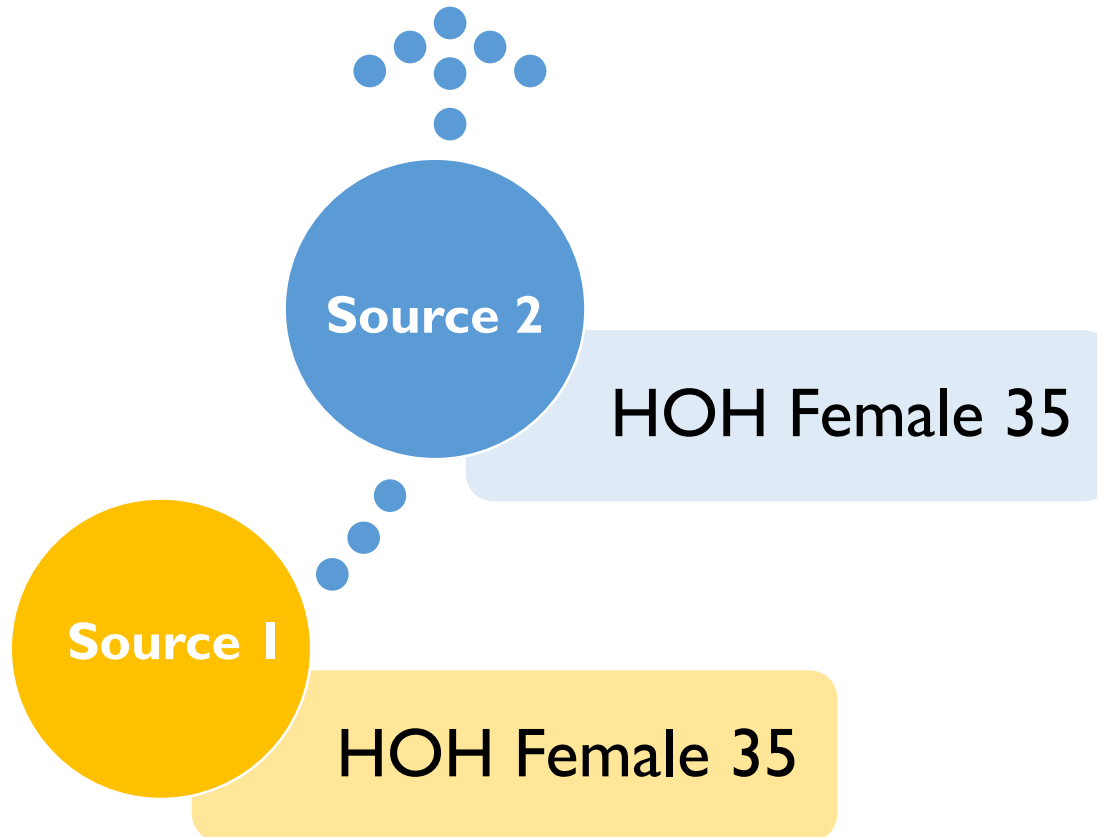
Increasing Efficiency – Auxiliary Variables

- Multiple points in time a household or individual variable is present the higher likelihood of being correct



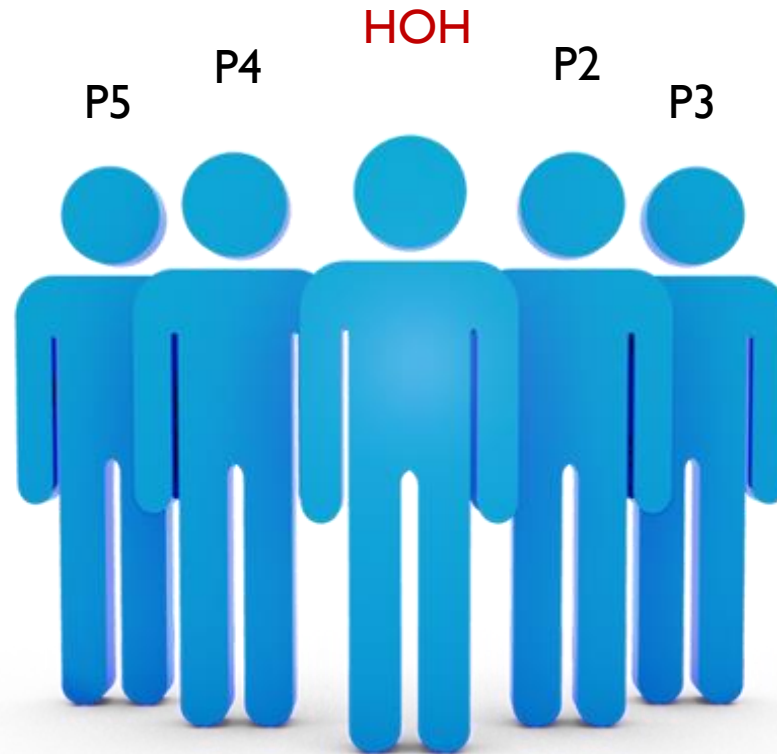
Increasing Efficiency – Auxiliary Variables

Higher likelihood HOH is a female aged 35



Increasing Efficiency – Auxiliary Variables

- Flagging multiple people in the household that meet your target



Increasing Efficiency – Auxiliary Variables

ADDITIONAL VARIABLES

- HH is Phone Objector
- HH length of residence
- Number of adults in HH
- HH contributes to charities
- HH has premium credit card
- HH has interest in reading
- HH has computer and peripherals
- Marital status of Member 1
- Gender of Member 1 known
- Religion of Member 1 Protestant
- Educational attainment of Member 1 known
- Occupational group of Member 1
- European Ethnicity of Member 1
- Member 1 owns business
- Multi-buyer HH
- Book buyer HH
- Gifts & Gadgets buyer HH
- General Merchandise buyer HH
- Presence of Child in HH
- Mosaic group of HH
- HH on Federal DNC list
- Home owner Probability Model score of HH
- HH has Cat Enthusiast

Citation: Data from BigSurv 2018 Presentation, Barcelona Spain: “Investigating the Value of Appending New Types of Big Data to Address-based Survey Frames and Samples”. Authored by Paul J. Lavrakas et al.

Increasing Efficiency – Auxiliary Variables

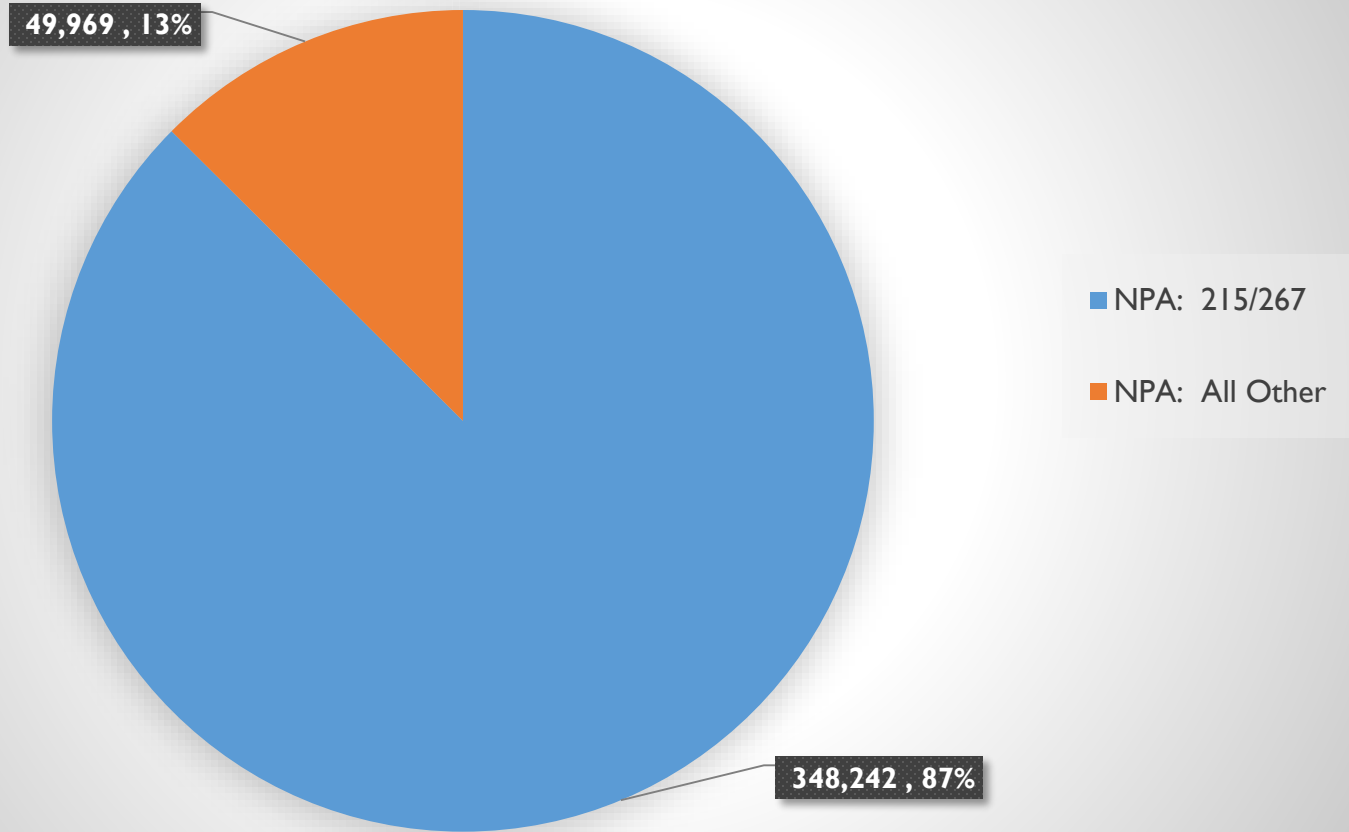
ADDITIONAL VARIABLES

- Bivariate examples of how these variables relate to Response/ Nonresponse in the 2018 NCS, and its overall AAPOR RRI for completion/consent of 27%
 - 47% of Premium credit card HHs completed/consented
 - 45% of Cat Enthusiast HHs completed/consented
 - 44% of Contributor to charities HHs completed/consented
 - 43% of HHs where Member 1 owns a business completed/consented
 - 39% of Higher status occupation or Retired completed/consented
 - 39% of Gifts/Gadgets buyer completed/consented
 - 38% of HHs where Educational of Member 1 is known completed/consented
 - 37% of HHs with General Merchandise buyer completed/consented
 - 34% of HH with Interest in Reading completed/consented
 - 33% of Phone Objector HHs completed/consented

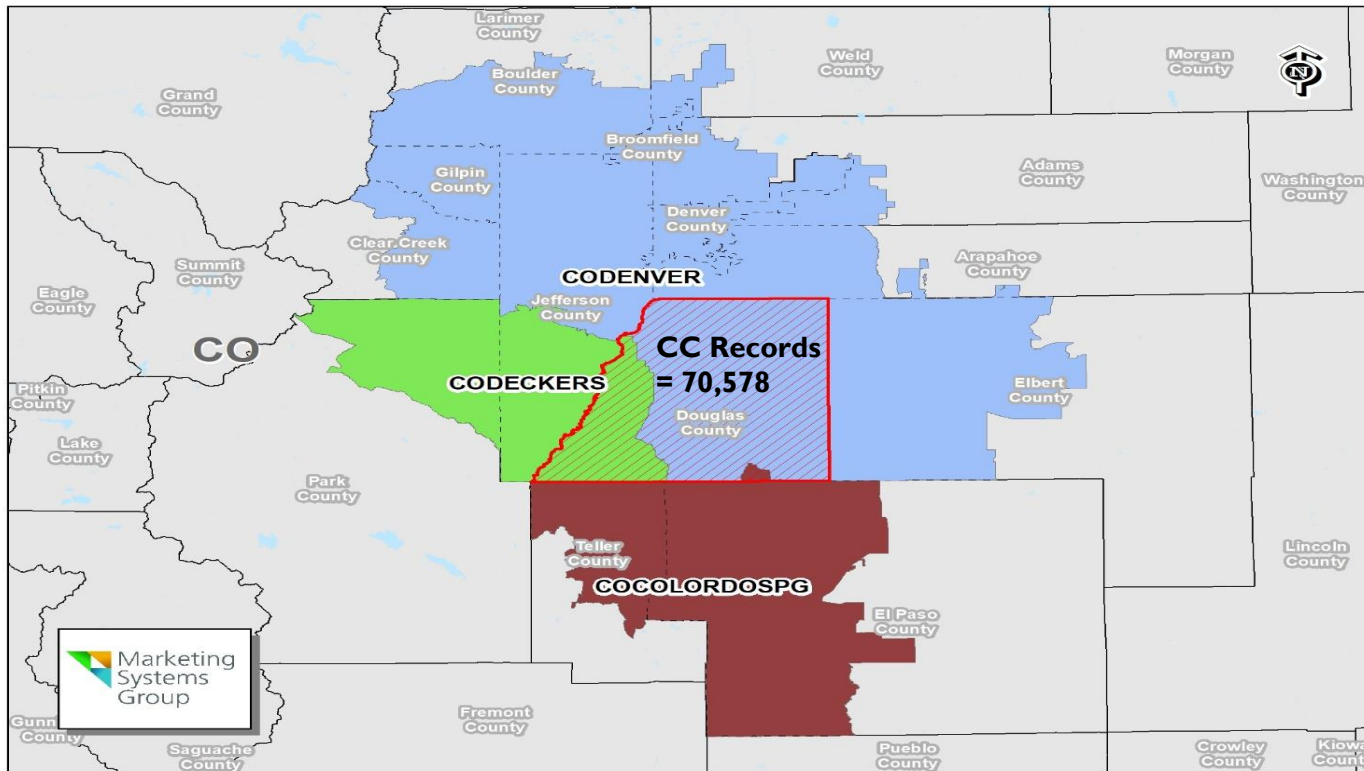
Citation: Data from BigSurv 2018 Presentation, Barcelona Spain: “Investigating the Value of Appending New Types of Big Data to Address-based Survey Frames and Samples” . Authored by Paul J. Lavrakas et al.

Increasing Efficiency – Auxiliary Variables

Philadelphia – Capturing Some of the In Migration



Increasing Efficiency – Auxiliary Variables



Rate Center ID	Primary FIPS Code	Total Housing Units	Cellular Telephone Universe	Total Population (2010)	% Inclusion (Population)	County	Name
CODENVER	08031	1,131,766	4,226	2,677,879	10.62%	08035	Douglas County
CODECKERS	08093	10,371	0	19,485	3.75%	08035	Douglas County
COCOLORDOSP G	08041	253,156	893	619,949	0.07%	08035	Douglas County

Increasing Efficiency – Expanded Capabilities

- Ability to implement a disproportionate stratified sampling design:
 - A cost-effective allocation plan that reduces screening costs
 - Keeping the design effect in an acceptable range (1-2)
 - Ensuring the coverage of hard-to-reach population within each stratum.

Hard to Reach Target Populations



vs.



Increasing Efficiency – Expanded Capabilities

STRATIFIED CELL RDD METHODOLOGY (S-RDD)

- Stratify the RDD frame into 4 distinct strata
- Identify some of the in-migration
- Identify cellular numbers that are on the commercial databases
- Identify the cellular numbers not in the area

Miami-Dade/Broward County		
Stratum 1	1,681,700	In RDD Frame; in Consumer Cell; in Target Geography
Stratum 2	5,275,283	In RDD Frame; Not in Consumer Cell (unlisted)
Stratum 3	312,146	Not in RDD Frame; In Consumer Cell; In Target Geography
Stratum 4	253,017	In RDD Frame; in Consumer Cell; outside Target Geography

Increasing Efficiency – Expanded Capabilities

TARGETED ABS METHODOLOGY (T-ABS)

- T-ABS is a disproportionate stratified sampling design that employs a cost effective optimal allocation plan that reduces screening costs
- Calculated design effects can be used for quality control by keeping them within the acceptable range of 1-2
- It's a single frame approach that assigns the entire population into one of the mutually exclusive strata
- Ensures the coverage of hard-to-reach population within each stratum.

Increasing Efficiency – Expanded Capabilities

DSS based on Auxiliary Variables

Over Generate a file of addresses from the ABS Frame (dependent on the estimated match rate of the demographic variable)

Match the addresses against the commercial databases to flag households that meet the demographic criteria of interest

Stratify into mutually exclusive strata:
Targeted vs. Remainder

Disproportionately sample from each by over sampling the targeted stratum and under sampling the remainder stratum

Increasing Efficiency – Expanded Capabilities

➤ Example I: Utilizing Auxiliary Variables for ABS

- Aiming to achieve 1000 homeowner Completes in San Francisco
- Excluded Vacant, Seasonal, Educational and Traditional PO Boxes
- Pulled all available addresses and ran them against all and the commercial sources to flag for homeowners:
 - Total households = 370,894
 - Total households flagged as homeowners = 91,950

Increasing Efficiency – Expanded Capabilities

Table 1

Simple Random Sample - Number of Contacts (35% Owners Overall)	
Stratum	Contacts
San Francisco County	2,824
Total sample	2,824
Design effect	1.00

Table 2

Targeted ABS - Number of Contacts

Stratum	Optimal Allocation 1	Optimal Allocation 2
Owner	861	900
Remainder	982	707
Total sample	1,843	1,607
Design effect	1.19	1.40

Table 3

Targeted ABS - Expected Number of Homeowners

Stratum	Optimal Allocation 1	Optimal Allocation 2
Owner	861	900
Remainder	139	100
Total target sample	1,000	1,000
Design effect	1.19	1.40

Increasing Efficiency – Expanded Capabilities

Table 1: SRS

Without Stratification: Owners						
Record Type	Completes	Response Rate	Eligibility Rate	Bad Sample	Yield Rate	Needed Sample (completes/yield rate)
Has phone	400	40%	35%	10%	13%	3,175
No phone	600	10%	35%	10%	3%	19,048
Total	1000	22%	35%	10%	7%	22,222
Total Sample = 22,222						

Table 2: T-ABS

Stratum 1: Owners						
Record Type	Completes	Response Rate	Eligibility Rate	Bad Sample	Yield Rate	Needed Sample (completes/yield rate)
Has phone	360	40%	70%	10%	25%	1,429
No phone	540	10%	70%	10%	6%	8,571
Total	900	22%	70%	10%	14%	10,000
Stratum 2: Remainder						
Record Type	Completes	Response Rate	Eligibility Rate	Bad Sample	Yield Rate	Needed Sample (completes/yield rate)
Has phone	40	40%	14%	10%	5%	794
No phone	60	10%	14%	10%	1%	4,762
Total	100	22%	14%	10%	3%	5,556
Total Sample = 15,556						

Increasing Efficiency – Expanded Capabilities

DSS based on Demographic Density

Run a Density Report based on census geography for the desired demographic

Impute the cut-of points for the incidence and coverage for the desired demographic

Stratify into mutually exclusive strata:
Targeted vs. Remainder

Disproportionately sample from each by over sampling the targeted stratum and under sampling the remainder stratum

Increasing Efficiency – Expanded Capabilities

- Example 2: Utilizing Demographic Density for Cellular RDD
 - Aiming to achieve 1000 Hispanic Completes in San Francisco
 - Ran a density report for percent Hispanic based on Census Block Groups and stratified into three strata based on percent Hispanic:

Total Pop	Hispanic	Incidence	Coverage	Density Level
236,461	76,974	33%	56%	High
224,170	31,513	14%	23%	Medium
424,317	28,078	7%	21%	Low
884,948	136,565		100%	

Increasing Efficiency – Expanded Capabilities

Table 1

Simple Random Sampling - Number of Contacts (16% Hispanic Overall)	
Stratum	Contacts
San Francisco County	6,250
Total sample	6,250
Design effect	1.00

Table 2

Targeted ABS - Number of Contacts			
Stratum	Optimal Allocation 1	Optimal Allocation 2	Optimal Allocation 3
Hispanic - High (33%)	2,148	2,305	2,305
Hispanic - Medium (14%)	1,335	1,067	1,425
Hispanic - Low (7%)	1,734	1,511	756
Total sample	5,217	4,883	4,486
Design effect	1.12	1.19	1.71

Table 3

Targeted ABS - Expected Number of Hispanics			
Stratum	Optimal Allocation 1	Optimal Allocation 2	Optimal Allocation 3
Hispanic - High (33%)	699	750	750
Hispanic - Medium (14%)	187	150	200
Hispanic - Low (7%)	114	100	50
Total target sample	1,000	1,000	1,000
Design effect	1.12	1.19	1.71

Increasing Efficiency – Expanded Capabilities

Table 1: SRS

Without Stratification: Hispanics						
Record Type	Completes	Response Rate	Eligibility Rate	Bad Sample	Yield Rate	Needed Sample (completes/yield rate)
Has phone	400	40%	16%	10%	6%	6,944
No phone	600	10%	16%	10%	1%	41,667
Total	1000	22%	16%	10%	3%	48,611
Total Sample = 48,611						

Table 2: T-ABS

Stratum 1: High Hispanic						
Record Type	Completes	Response Rate	Eligibility Rate	Bad Sample	Yield Rate	Needed Sample (completes/yield rate)
Has phone	300	40%	33%	10%	12%	2,560
No phone	450	10%	33%	10%	3%	15,361
Total	750	22%	33%	10%	6%	17,921
Stratum 2: Medium Hispanic						
Record Type	Completes	Response Rate	Eligibility Rate	Bad Sample	Yield Rate	Needed Sample (completes/yield rate)
Has phone	80	40%	14%	10%	5%	1,581
No phone	120	10%	14%	10%	1%	9,524
Total	200	22%	14%	10%	3%	11,104
Stratum 3: Low Hispanic						
Record Type	Completes	Response Rate	Eligibility Rate	Bad Sample	Yield Rate	Needed Sample (completes/yield rate)
Has phone	20	40%	7%	10%	2%	839
No phone	30	10%	7%	10%	1%	4,762
Total	50	22%	7%	10%	1%	5,601
Total Sample = 34,627						

RR – Additional Modes of Contact

- Append Phones Numbers (land line/cellular) – 50%-60% match rate
- Append Mailing Address
 - Land line - 30%-40% match rate
 - Cellular – 15%-25% match rate
- Append E-mail Addresses – 10%-20% match rate

RR – Additional Modes of Contact

Matching Levels (Exact or Inexact)

1. Individual Level
 - a. First Name
 - b. Last Name
 - c. Address
2. Household Level
 - a. Last Name
 - b. Address
3. Address Level
 - a. Address



RR – Additional Modes of Contact

Matching Levels (Exact or Inexact)

1. Individual Level
 - a. First Name
 - b. Last Name
 - c. Address
2. Household Level
 - a. Last Name
 - b. Address
3. Address Level
 - a. Address



Matching Specifications

- Exact Matching
- Fuzzy Matching
- One unique email per
- Multiple emails per
- Pinging for bounce backs
- Permission Pass

RR – Tailored Approach

- Sequential data collection (least to most expensive)
- Language specific
- Varying rates of incentive based on response propensity
- Mode preference based on demographic/psychographic characteristics
- Pre-notification letters

THANK YOU!

Ashley H. Hyon

Ahyon@m-s-g.com

610-994-8321 (O)

215-870-3338 (C)

