



Mansour Fahimi, Ph.D.
EVP & Chief Data Scientist

Ashley Hyon
VP Survey Research Methods

ADDRESS BASED SAMPLING (ABS)

ABS FOR SCIENTIFIC SURVEY
SAMPLING APPLICATIONS

OUTLINE

What is ABS?

CDSF Data Structure

ABS Frame Compilation

ABS Frame Enhancements

Sampling Options

Survey Administration Options

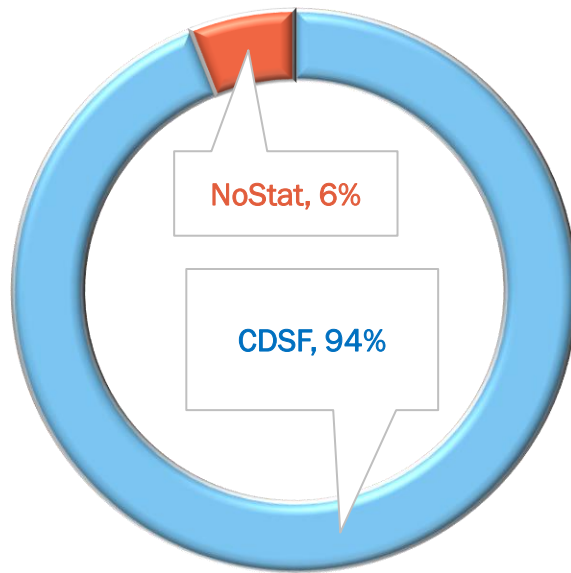
Weighting Considerations

Conclusion



WHAT IS ABS?

- ABS is address-based sampling using the Computerized Delivery Sequence File (CDSF) from the USPS:
 - Delivery codes for detection of business and residential addresses
 - Address validation and standardization
- CDSF is a database containing all delivery points in the US:
 - Removes undeliverable addresses and increases speed of delivery
 - Updated with daily feedback from thousands of letter carriers
- But the CDSF is a database for mail delivery:
 - Is ZIP-based and does not recognize Census geodemographics
 - Contains no population data for sample stratification



CDSF DATA STRUCTURE

- CDSF contains all addresses with standard delivery format:
 - Includes all LACS converted addresses
 - Meets all CASS certification requirements
 - Requires no additional address hygiene processing
- There is a diminishing number of simplified addresses:
 - Converting to city-style format for 911 needs
 - Further resolution utilizing commercial databases
- USPS NoStat file includes:
 - List of addresses currently not receiving mail
 - List of vacant delivery points on rural routes
 - Some information about units within drop points

Address Classification

- **Business:** Indicates the delivery point is a business address
- **Central:** The delivery point is serviced at a mail receptacle located within a centralized unit
- **Commercial Mail Receiving Agency:** Private business acting as a mail-receiving agent for specific clients
- **Curb:** The delivery point serviced via motorized vehicle at a mail receptacle located at the curb
- **Drop:** A delivery point or receptacle that services multiple residences such as a shared door slot or a boarding house in which mail is distributed internally by the site
- **Educational:** Identified as an educational facility such as colleges, universities, dormitories, sorority or fraternity houses, and apartment buildings occupied primarily by students
- **Neighborhood Delivery Collection Box Unit:** Services at a mail receptacle located within a cluster box
- **Seasonal:** Receives mail only during a specific season when addresses are occupied are identified
- **Throwback:** A street address but the delivery is made to the customer's P.O. Box address
- **Vacant:** Active in the past but is currently vacant (in most cases unoccupied over 90 days) and not receiving deliver

USPS IS ZIP CODE BASED

- ❖ **Zone Improvement Plan:** A 5-digit code that identifies a geographic delivery area or a single building or company with high mail volume
- ❖ **ZIP Sector:** The first two digits of the +4 addon representing a smaller geographic area within a ZIP Code such as several blocks or a group of streets
- ❖ **ZIP Segment:** The last two digits of the +4 addon representing a smaller geographic area within a ZIP Sector such as one floor of an office building or a side of a street

Average ZIP Code Population

2000	6,772
2010	7,492
2020	7,931

ZIP Code Counts	Number	Percent
General	30,794	75.0%
P.O. Box	8,210	20.0%
Military	469.3	1.1%
Business	1,870	4.6%
Total	41,083	100.0%

CDSF DATA ELEMENTS



- ❖ Primary Address
- ❖ Secondary Address
- ❖ City
- ❖ State
- ❖ Zip
- ❖ ZIP4
- ❖ House Number
- ❖ Pre-Directional
- ❖ Street Name
- ❖ Street Suffix
- ❖ Post Direction
- ❖ Secondary Unit Descriptor
- ❖ Apt Number
- ❖ Carrier Route
- ❖ Delivery Point
- ❖ Delivery Point Check Digit
- ❖ Route Type
- ❖ Delivery Point Type Code
- ❖ Delivery Point Usage Code
- ❖ Drop Code
- ❖ Drop Count
- ❖ PO Box Throwback
- ❖ Seasonal Code
- ❖ Vacant Code
- ❖ OWGM
- ❖ Walk Sequence

MAJOR DELIVERY TYPE COUNTS

Delivery Type	Count	Distribution
City Style (street address)	123,896,531	85.73%
Rural Routes/Highway Contracts	63,824	0.04%
P.O. Box	14,235,673	9.85%
P.O. Box (Only Way to Get Mail/OWGM)	1,415,853	0.98%
Seasonal (vacation/second homes)	821,596	0.57%
Educational (off-campus housing)	98,031	0.07%
Vacant (long term 90-day vacancy)	2,960,792	2.05%
Throwback	201,174	0.14%
Drop Points	713,739	0.49%
Drop Units	2,013,366	1.39%
Drop Unit Augments	41,438	0.03%
City Style Augments	62,731	0.04%
P.O. Box Augments	14,569	0.01%
Simplified Addresses	102,769	0.07%
Total	144,512,494	100.00%

Why ABS?



Struggles with Traditional Methods

- Evolving coverage problems of telephone-based options
- Indecisive sampling and weighting procedures
- Growing rates of nonresponse to single mode methods

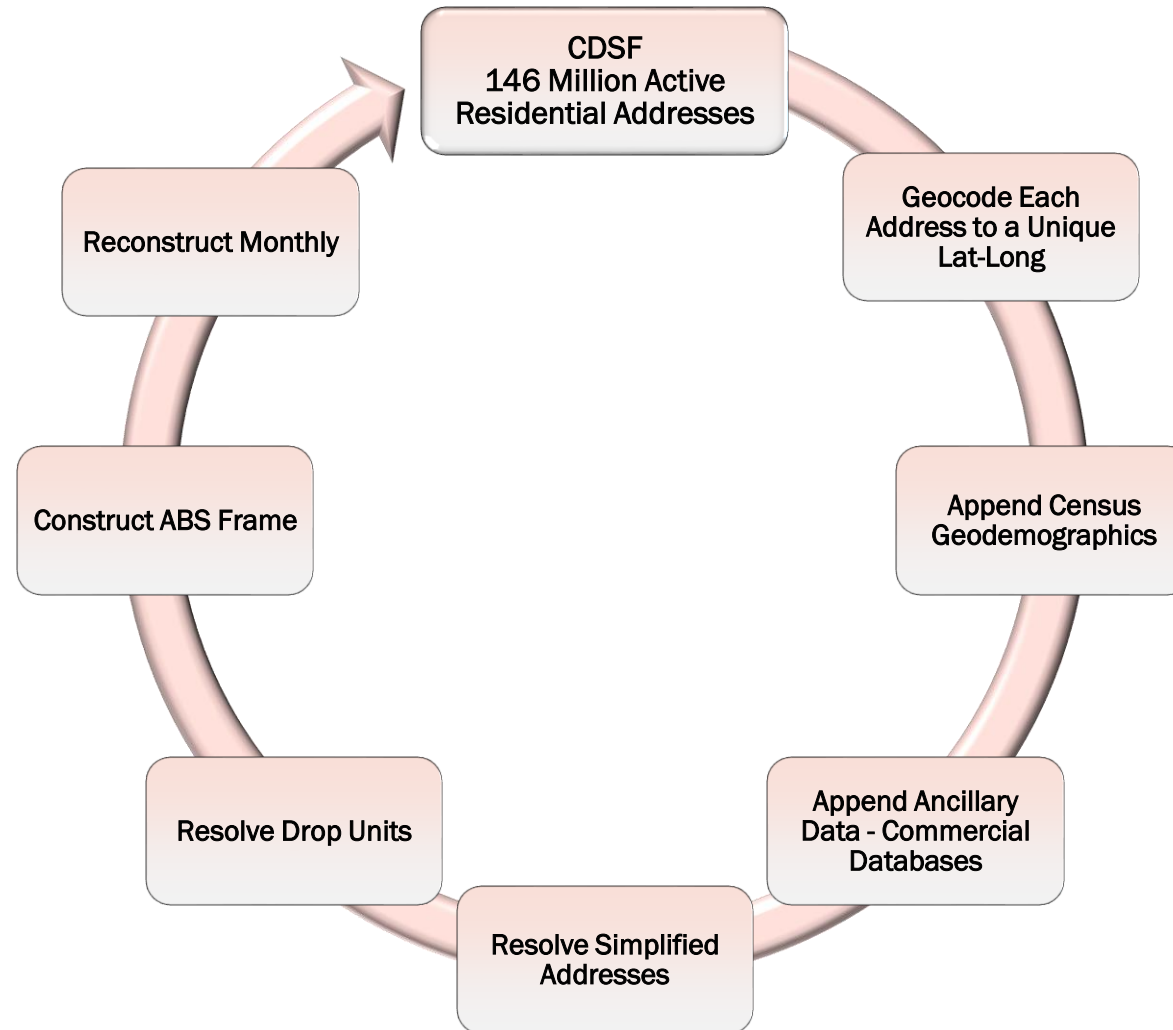
Multi-mode Approach

- Improved coverage
- Boost response rates
- Reduce cost

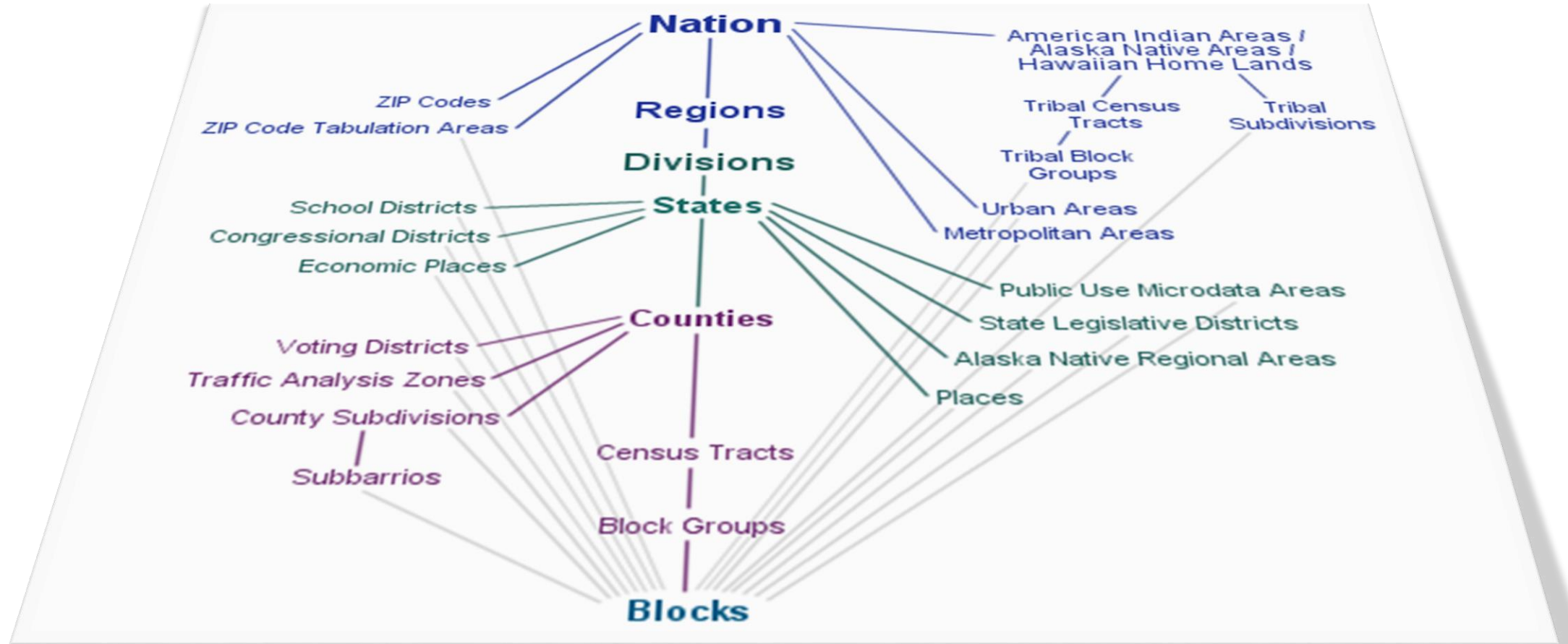
Effective Survey Sampling Protocols via ABS

- Detailed geodemographic data
- Geodemographic stratification
- Multiple-mode capabilities

FROM CDSF TO ABS SAMPLING FRAME



TOPOLOGY OF THE CENSUS GEOGRAPHY



TOPOLOGY OF THE CENSUS GEOGRAPHY

Counties

- ❖ Local level of government below state
- ❖ There are 3,142 counties or county-equivalent units in the US
- ❖ Average county population is about 100,000:
 - Los Angeles County has a population of about 10 million
 - Loving County in Texas has a population of only 169
- ❖ Louisiana is divided into parishes
- ❖ Alaska is divided into boroughs

Tracts

- ❖ Census tracts are geographic entities within counties with boundaries that follow visible features
- ❖ Census tracts cover between 2,500 to 8,000 residents
- ❖ Tracts are homogeneous with respect to population characteristics, economic status, and living conditions
- ❖ Census tracts are subdivided into Census block groups and blocks

TOPOLOGY OF THE CENSUS GEOGRAPHY

Census Block Groups (CBG)

- ❖ The smallest geographical unit for which the bureau publishes data
- ❖ CBGs are identified by a single number within a tract that identifies all the blocks that compose the CBG
- ❖ There are over 211,000 CBGs in the US and Puerto Rico, each containing an average of 39 blocks
- ❖ CBGs generally contain between 600 and 3,000 people with an optimum size of 1,500 people

Census Blocks

- ❖ Typically bounded by streets, roads, creeks, or other features
- ❖ Blocks typically have a four-digit number where the first digit indicates which CBG it belongs to
- ❖ In the US and Puerto Rico there are about 8,200,000 census blocks
- ❖ Block population can range from zero to blocks with hundreds of inhabitants
- ❖ There are 2,700,000 blocks with zero population in the US

DATA FRAME ENHANCEMENTS by MSG

- ❖ Decennial Census Summary Files (SF1):
 - ✓ The only source providing data down to the census block level
 - ✓ Contains 286 tables on age, gender, households, families, and housing units
- ❖ American Community Survey (ACS) Provides:
 - ✓ Annual estimates based on survey of 3.5 million addresses
 - ✓ 5-year ACS provides population estimates down to the CBG level
- ❖ Claritas Pop-Facts Provides:
 - ✓ Current and 5-year projections for common geodemographics
 - ✓ The only source for ZIP and DMA based population estimates
- ❖ Specialty Databases:
 - ✓ Aliments, behaviors, preference, etc.
 - ✓ Registered voters

MODES OF CONTACT

- Mailing address for all delivery points:
 - Name
 - Household characteristics
- E-mail addresses:
 - Unique email for about 40% of addresses
 - Multiple emails per person
- Phones Numbers:
 - Overall append rate of about 55% from white pages
 - About 40% are cellular from consumer files

ABS FRAME ENHANCEMENTS (Behavioral & Lifestyle Segments)



Demographics

Age, income, education, occupation, nationality, family size



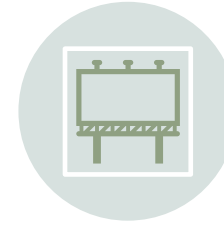
Psychographics

Consumer beliefs, opinions, habits and lifestyles



Geography

Location and descriptors, such as cost of living and urbanicity



Attitudes

Reasons why consumer use products and services



Needs

Perceived drivers and trade-offs in purchase behavior



Purchase Behaviors

Specific products, services, and brand names consumers purchase



Buying Preferences

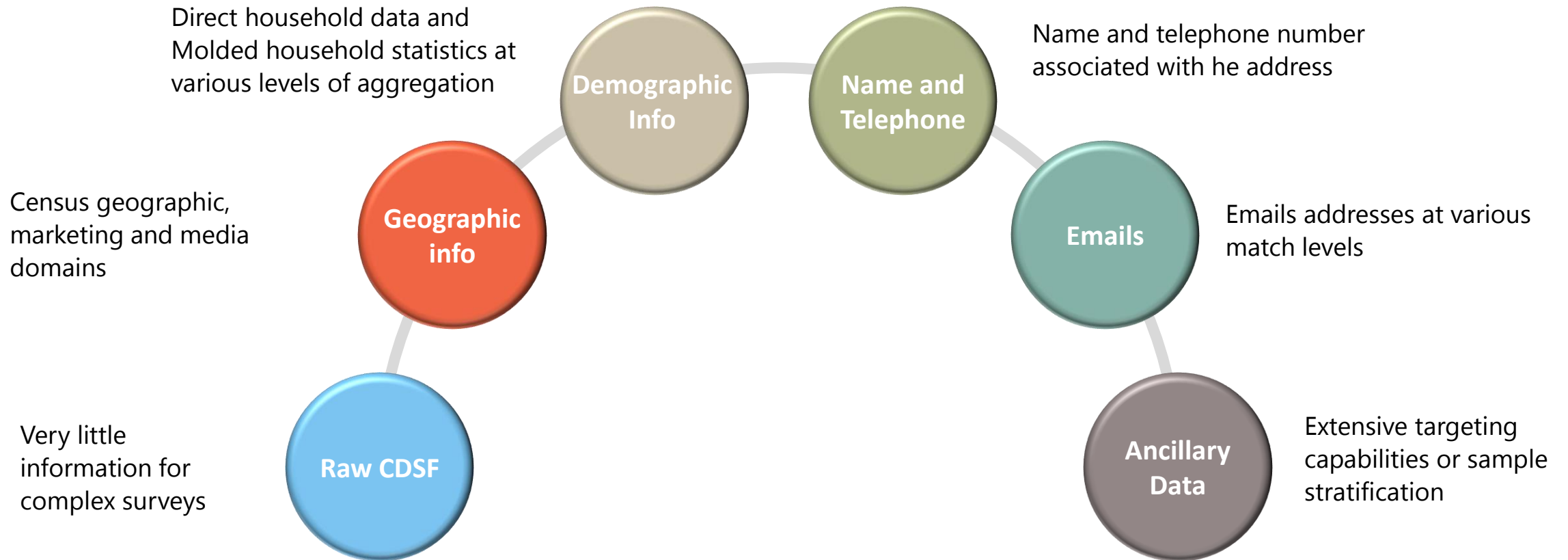
Where and why consumers buy products & services



Media Preferences

Preferred channels and media providers

FROM CDSF TO ABS SAMPLING FRAME



SAMPLING OPTIONS

- Simple Random Sampling
- Systematic Random Sampling using ZIP+4 for sorting
- Stratified Sampling:
 - Geography
 - Demography
 - Geodemography
 - Ancillary Indicators:
 - Behavioral – smokers
 - Political – registered voters
 - Purchase habits
 - Preferences

SAMPLE ALLOCATION OPTIONS

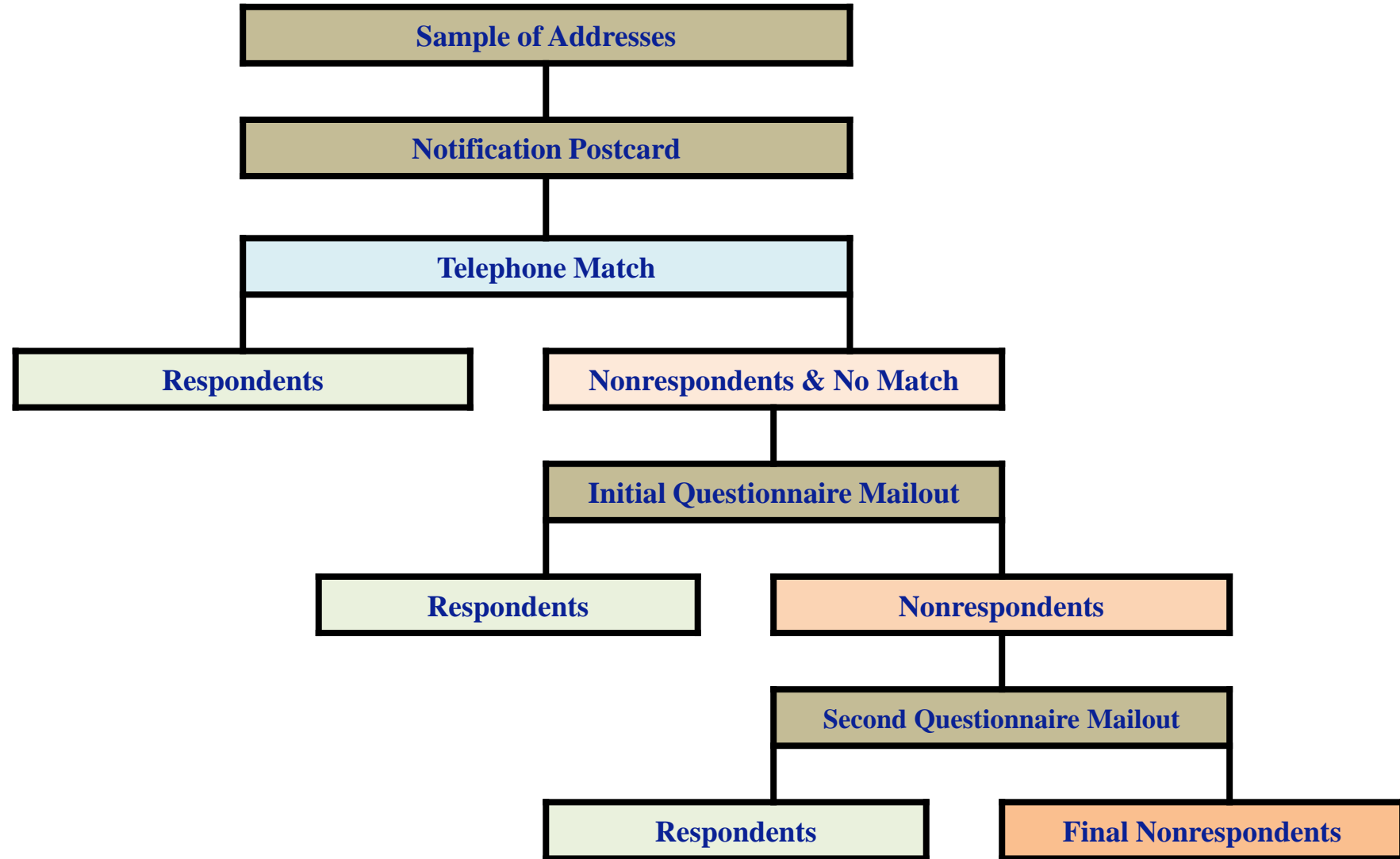
$$n_i = n \times \frac{\frac{N_i S_i}{\sqrt{C_i}}}{\sum_i \frac{N_i S_i}{\sqrt{C_i}}}$$

$$n \times \frac{N_i}{N}, \text{ if } C_i \equiv C \text{ \& } S_i \equiv S$$

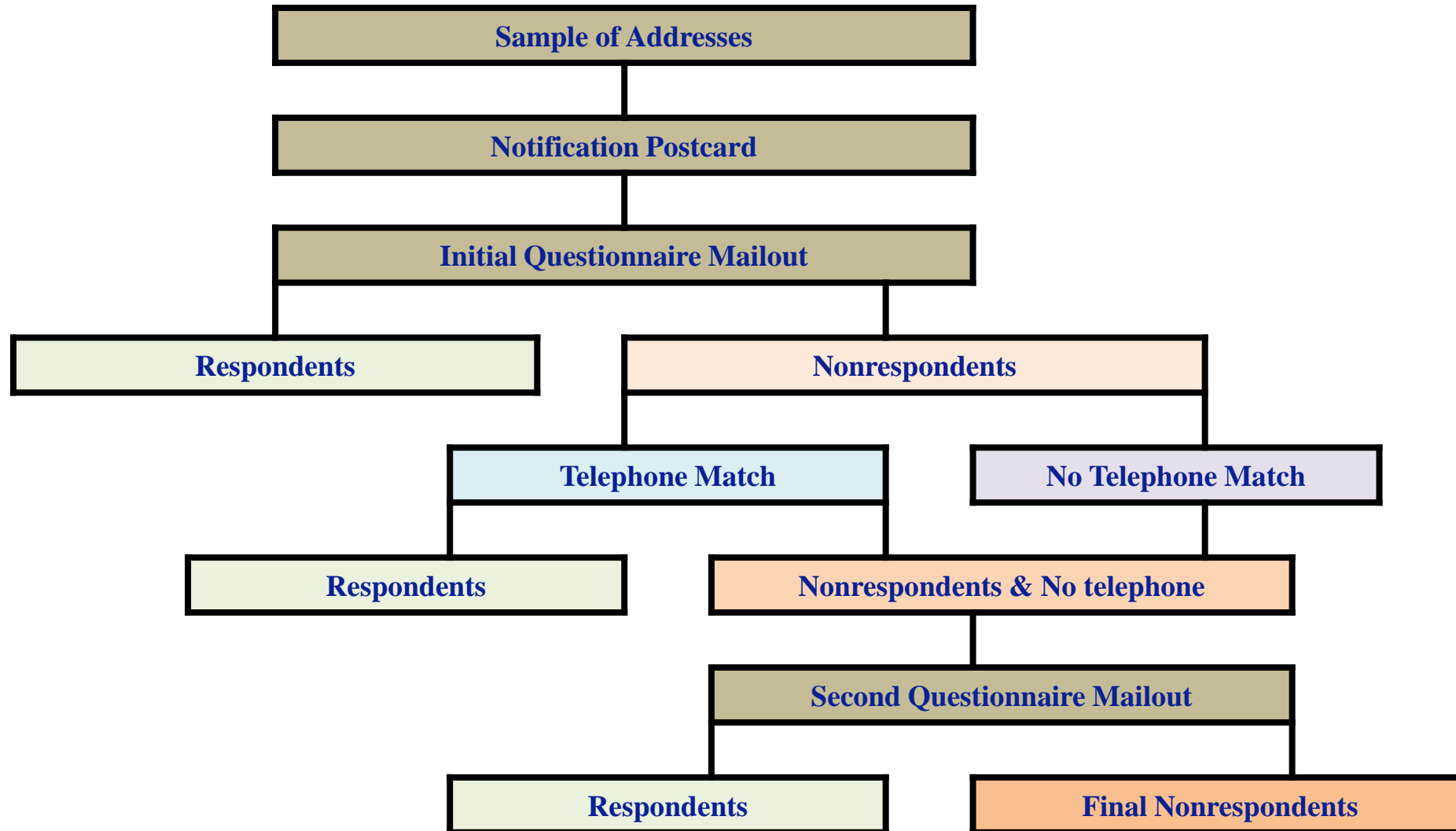
$$n \times \frac{N_i S_i}{\sum_i N_i S_i}, \text{ if } C_i \equiv C$$

$$n \times \frac{\frac{N_i}{\sqrt{C_i}}}{\sum_i \frac{N_i}{\sqrt{C_i}}}, \text{ if } S_i \equiv S$$

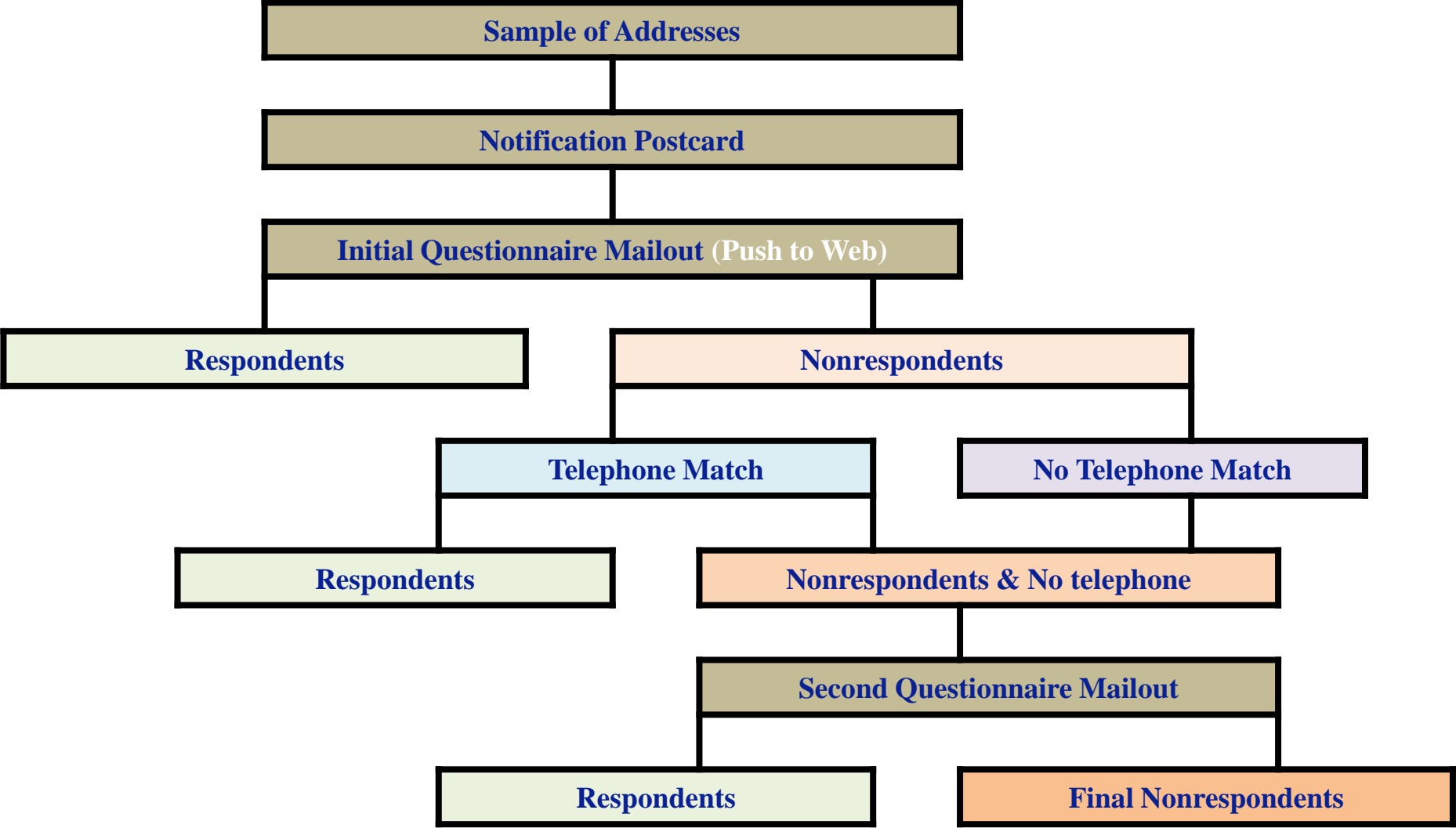
SURVEY ADMINISTRATION OPTIONS

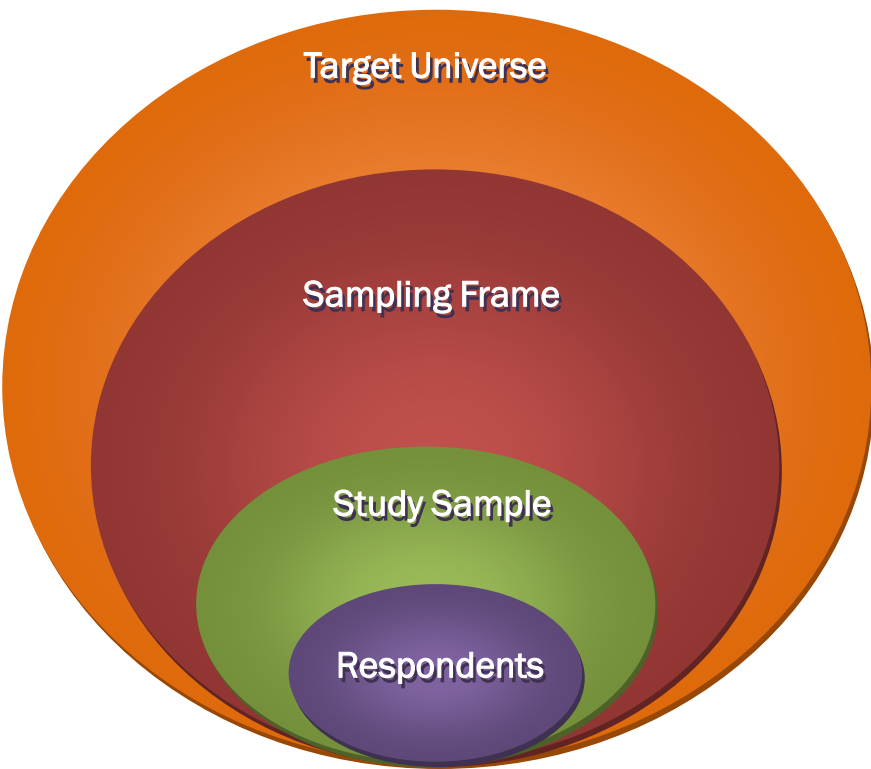


SURVEY ADMINISTRATION OPTIONS



SURVEY ADMINISTRATION OPTIONS





WEIGHTING CONSIDERATIONS

- ❖ Single frame for sample selection offers unambiguous selection probabilities and eliminates all hocus-pocus from the weighting process
- ❖ Hundreds of ancillary data on the MSG-refined ABS frame provides an unparalleled framework for nonresponse bias analysis and adjustments
- ❖ Dedicated *GeoDemo* team provides comprehensive access to public and commercial data sources for population benchmarks at various levels of aggregation

WHY ABS?

- Gold standard sampling frame providing complete coverage of all delivery points
- A single frame provides:
 - ✓ Unambiguous selection probabilities
 - ✓ Coherent weighting & analytics without fuzzy extrapolations
- Availability of ancillary data from:
 - ✓ Public sources (CPS, ACS, and other government surveys)
 - ✓ Commercial sources including name and phone numbers
- Enhanced sampling frame supports complex design options
- Surgical accuracy for targeting small areas and rare subgroups
- Conducive to all modes of contact for effective data collection

ABS and Covid-19

- Obtaining reliable and measurable estimates of infection rates in the country:
 - ✓ Invitation to testing sites
 - ✓ Mailing self-administered tests
 - ✓ Home visit testing
- Scientific survey sampling protocols require:
 - ✓ Probability-based samples
 - ✓ Full coverage of the US
 - ✓ Geographic targeting
 - ✓ Demographic targeting
 - ✓ Multi-mode of contact
 - ✓ Rigorous projections and analytics

- ❖ AAPOR Taskforce Report on ABS: <https://www.aapor.org/Education-Resources/Reports/Address-based-Sampling.aspx>.
- ❖ Fahimi, M. & D. Kulp. Address-Based Sampling. *Quirk's Marketing Research Review*, May 2009.
- ❖ Staab, M., & Iannacchione, V. Evaluating the use of residential mailing addresses in a national household survey. *Proceedings of the ASA, Survey Methodology Section*, pp.4028- 4033.
- ❖ Dohrmann, S. & Mohadjer, L. Residential Address Lists vs. Traditional Listing. *Proceedings of the ASA, Survey Methodology Section*, Seattle, WA. pp. 2959- 2964.
- ❖ Montaquila, Jill, V. Hsu, J. M. Brick, N. English, and C. O'Muircheartaigh. "A comparative evaluation of traditional listing vs. address-based sampling frames." *Proceedings of the JSM*, 2009.
- ❖ Link, M. et al. Addressed-based vs. RDD Surveys: Comparison of Key Health and Risk Indicators. *American Journal of Epidemiology*, 164, 1019 - 1025.
- ❖ Dillman, D. The Design & Administration of Mail Surveys, *Annual Review of Sociology*, 17, 225-249.



The wound is the place where the *Light* enters you.

Rumi