

MARKETING | SYSTEMS | GROUP

Leadership Through Innovation

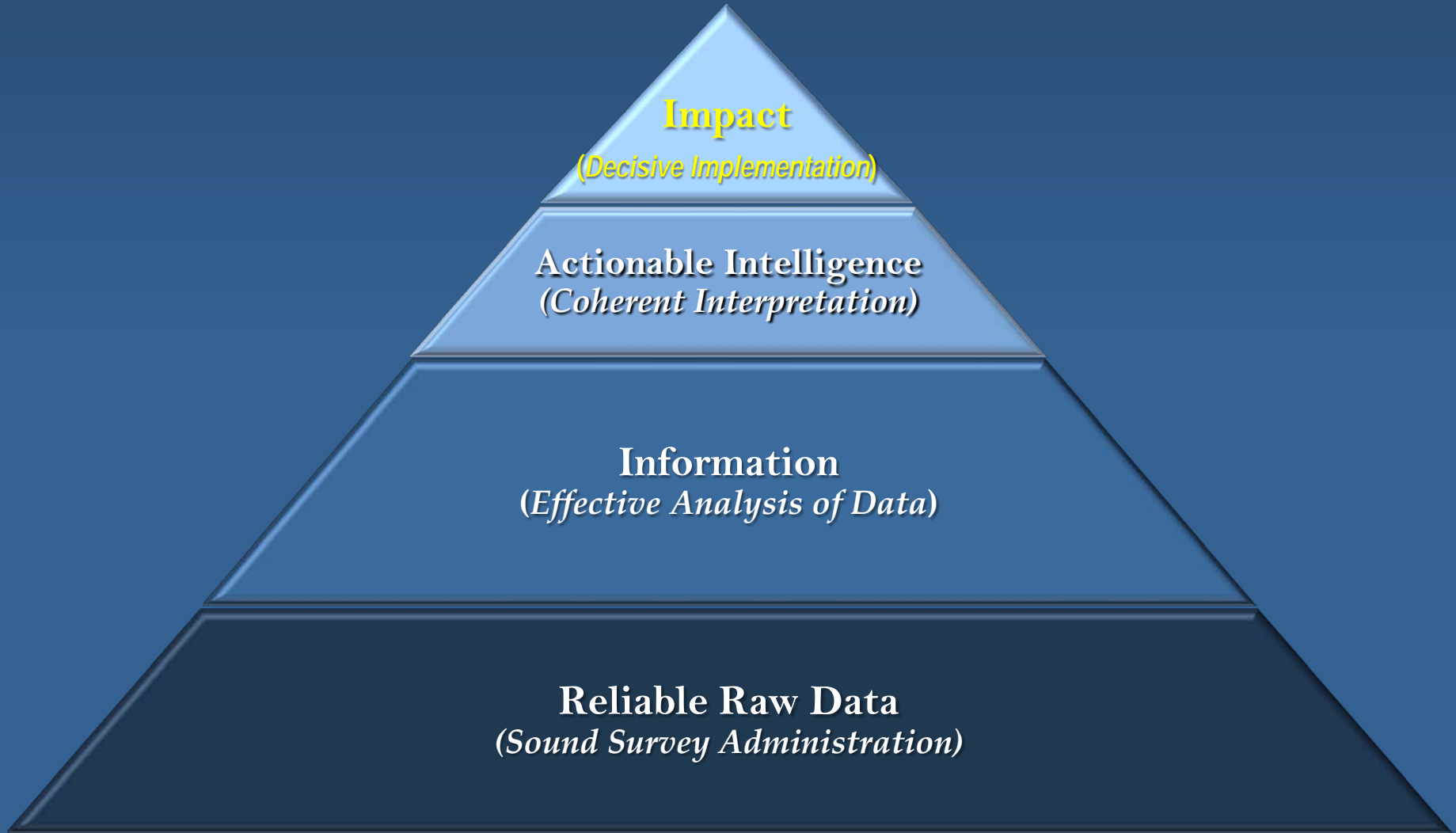
Address-Based Sampling (ABS) Merits, Design, and Implementation

*Mansour Fahimi, Ph.D.
VP, Statistical Research Services*

**National Conference on Health Statistics
National Center for Health Statistics (NCHS)**

August 16 - 18, 2010

FROM DATA TO IMPACT



SOURCES OF SURVEY ERRORS

Total Survey Error

Errors of
Non-observation

Sample Coverage

Response Rates

Errors of
Observation

Instrument

Data Collection

Errors of
Processing

Data Cleaning &
Editing

Imputation &
Weighting

Errors of
Dissemination

Analysis of
Survey Data

Interpretation &
Conclusion

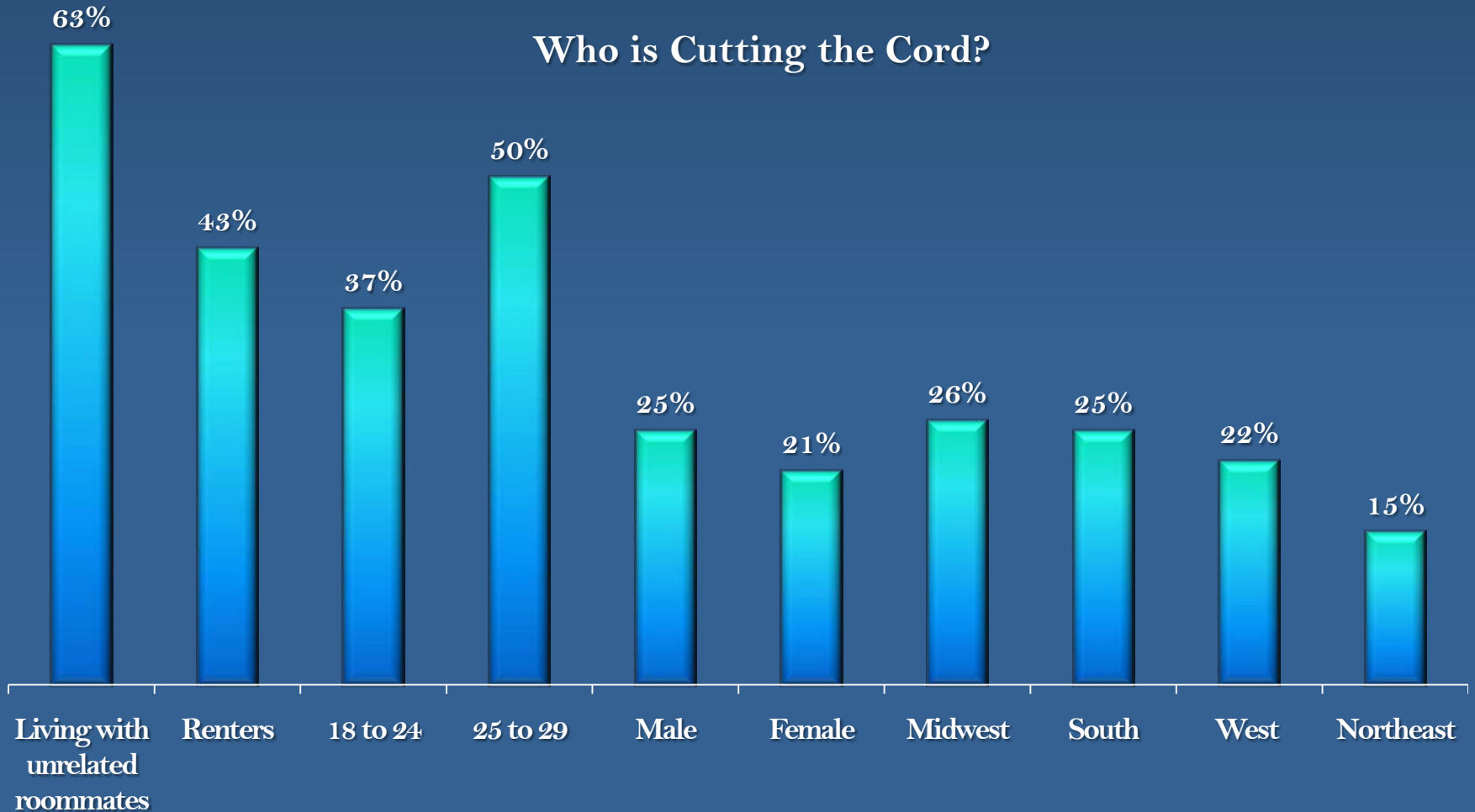
REASONS FOR EMERGENCE OF ABS

- ▣ Evolving coverage problems associated with RDD samples
- ▣ Eroding rates of response to single modes of contact and the increasing costs of refusal conversion
- ▣ Convoluted sampling/weighting/estimation implications of interim alternatives via dual-frame methodology
- ▣ ABS provides a versatile platform for creative strategies to improve coverage and response rates
- ▣ Availability of the Computerized Delivery Sequence File (**CDSF**) of the USPS for sampling purposes

COVERAGE PROBLEMS FOR RDD SAMPLES

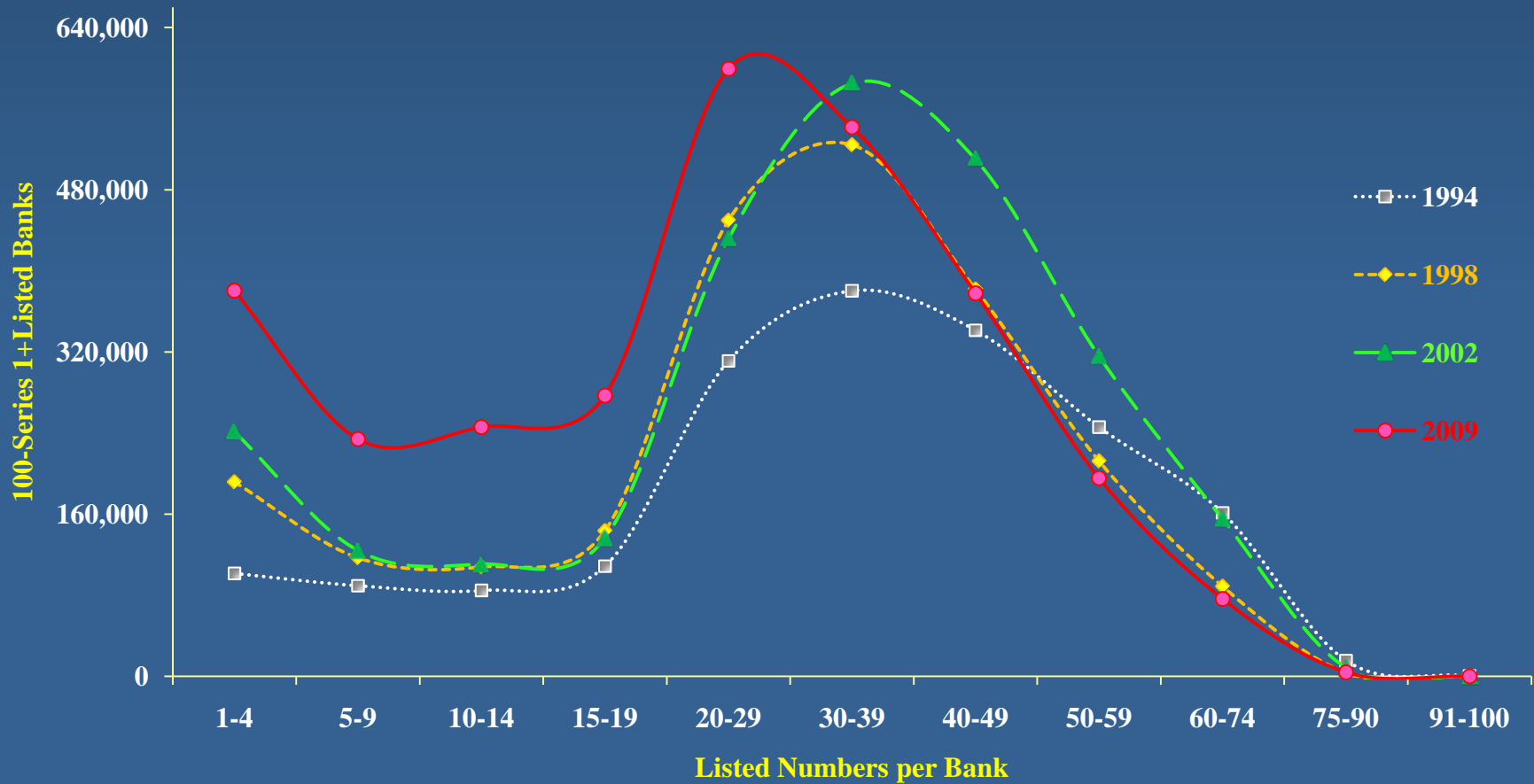
(A growing percentage of adults are becoming cell-only)

Who is Cutting the Cord?



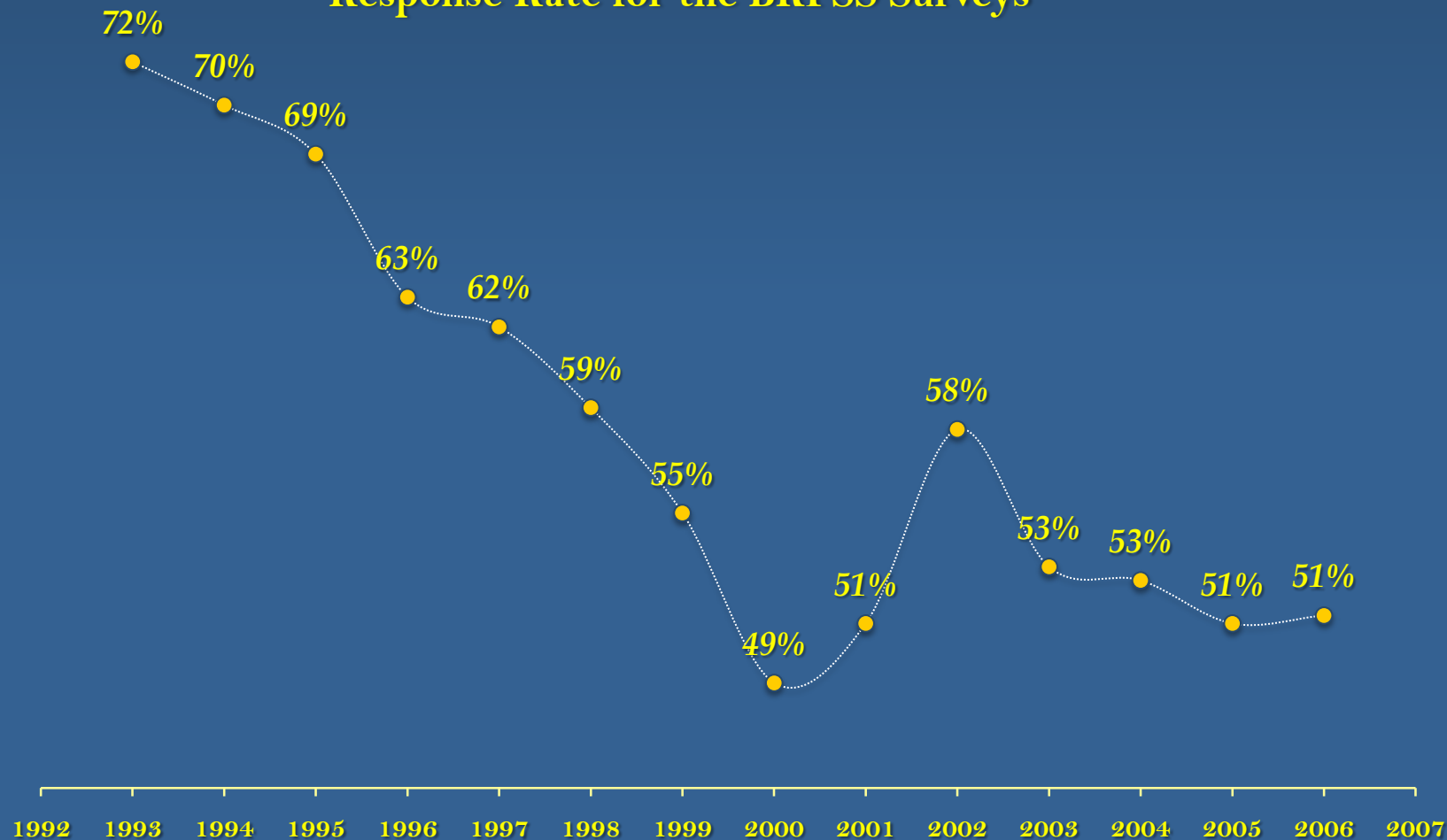
COVERAGE PROBLEMS FOR RDD SAMPLES (Beyond Cell Phones)

Distribution of 1+Listed 100-Series Banks by Residential Density



ERODING RATES OF RESPONSE TO SINGLE MODES OF CONTACT

Response Rate for the BRFSS Surveys



IMPROVEMENTS IN DATABASES OF HOUSEHOLD ADDRESSES

- ▣ With over 135 million addresses the CDSF is the most complete address database
- ▣ CDSF improves *address hygiene*:
 - ▣ Reduce undeliverable-as-addressed mailings
 - ▣ Increase delivery speed
 - ▣ Reduce cost
- ▣ Continuous database update via daily feedback from thousands of letter carriers

SAMPLING CANVAS VIA ABS



TOPOLOGY OF THE CDSF

(Delivery Point Types)

- ▣ **Business:** Indicates the delivery point is a business address
- ▣ **Central:** The delivery point is serviced at a mail receptacle located within a centralized unit
- ▣ **CMRA (Commercial Mail Receiving Agency):** A private business that acts as a mail-receiving agent for specific clients
- ▣ **Curb:** The delivery point is 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 by students

TOPOLOGY OF THE CDSF

(Delivery Point Types)

- ▣ **NDCBU (Neighborhood Delivery Collection Box Unit):** Services at a mail receptacle located within a cluster box
- ▣ **No-Stat:** Indicates address is not receiving delivery and is not counted as a possible delivery point for various reasons
- ▣ **Seasonal:** Receives mail only during a specific season and the months the seasonal addresses are occupied are identified
- ▣ **Throwback:** Address associated with this delivery point is a street address but the delivery is made to a P.O. Box address
- ▣ **Vacant:** Was active in the past, but is currently vacant (in most cases unoccupied over 90 days) and not receiving delivery

TOPOLOGY OF THE CDSF

(Counts of Delivery Points)

Delivery Type	Count
City Style/Rural Routes	114,135,810
PO Box	14,936,080
Seasonal	890,488
Educational	110,914
Vacant	4,071,036
Throwback	291,302
Drop Points	786,896
Augmented City Style/Rural Route (MSG)	192,443
Augmented PO Boxes (MSG)	395,307
Total	135,810,276

CDSF IS NOT A SAMPLING FRAME

(Possible Enhancements for ABS)

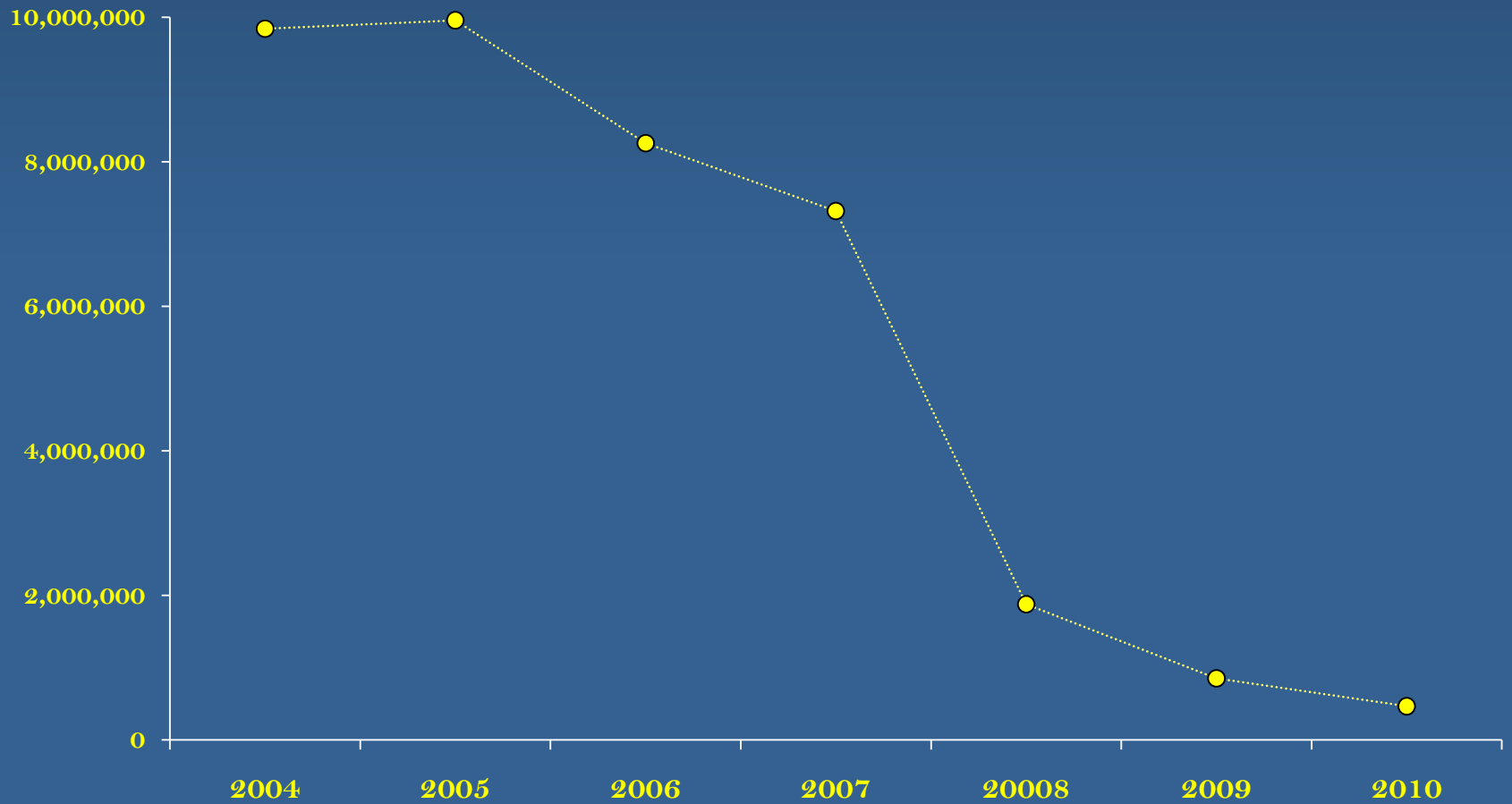
- ▣ CDSF does not include effective stratification variables
 - ▣ Detailed geodemographic data appendage
- ▣ Certain delivery points are more likely to be excluded
 - ▣ Simplified address resolution
 - ▣ Predicting areas of poor coverage (need for listing)
- ▣ Certain dwellings have multiple chances of selection
 - ▣ Methods for reducing frame multiplicity

POSSIBLE ENHANCEMENTS OF THE CDSF

(Appending Information)

- ▣ Geographic Information Enhancements:
 - ▣ Census geographic domains
 - ▣ Marketing and media domains
- ▣ Demographic Information Enhancements:
 - ▣ Direct household data from commercial databases
 - ▣ Molded household statistics at various levels of aggregation
- ▣ Name and Telephone Number Retrievals:
 - ▣ Append a name associated with the address
 - ▣ Retrieve listed telephone number associated with the name
- ▣ Simplified Address Resolution

SIMPLIFIED ADDRESSES BY YEAR



POSSIBLE ENHANCEMENTS OF CDSF

(Resolution Summary for CDSF-Based Samples)

- ▣ There are about 135 million residential addresses:
 - ▣ Simplified addresses account for 467,375 addresses
 - ▣ MSG can augment the majority of simplified addresses
 - ▣ Augmented sampling frame covers over 99% of all residential addresses in the U.S.

- ▣ Percent name append on average is about 90 and more

- ▣ Percent phone append on average is about 60

- ▣ Match rates will vary with geography and inclusion of P.O. Boxes as they tend to drive down the rates

POSSIBLE ENHANCEMENTS OF CDSF

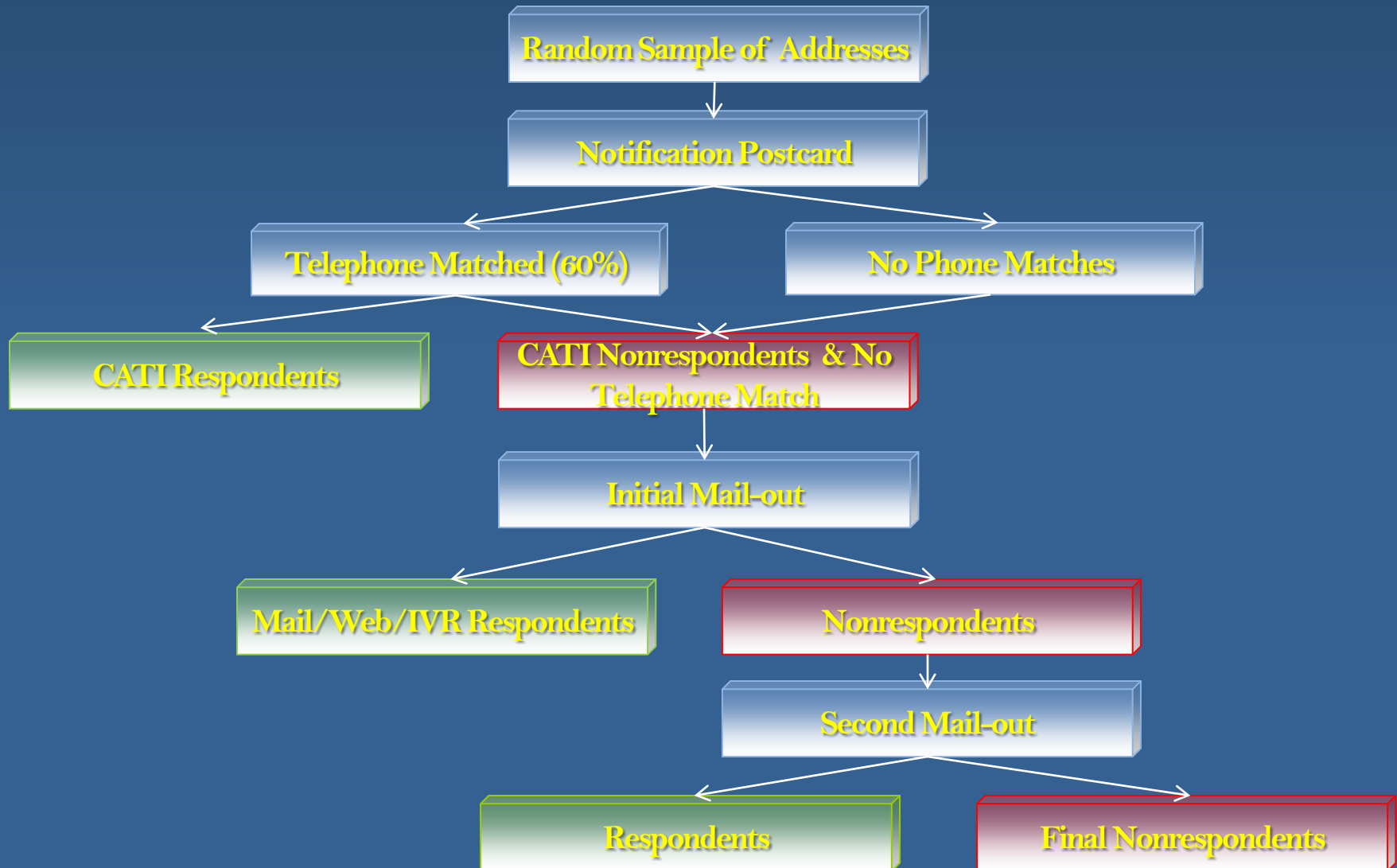
(Reducing the Frame Multiplicity)

PO Boxes (Including Augmented)	Count
PO Box	15,331,387
Only Means of General Delivery	5,256,279
Non-vacant PO Boxes	3,639,618
Potential Duplicates (Box & Address)	10,075,108

POSSIBLE ABS IMPLEMENTATION PROTOCOL (Option One)



POSSIBLE ABS IMPLEMENTATION PROTOCOL (Option Two)



PROS & CONS OF MULTI-MODE ALTERNATIVES

- ▣ In comparison to single-mode methods ABS with multiple modes for data collection can (Link 2006, 2007,2009):
 - ☺ Improve coverage
 - ☺ Boost response rates
 - ☺ Reduce cost (hard & soft)
- ▣ Multi-mode methods that include mail as an option can entail:
 - ☹ Compromised ability to conduct quick turnaround studies
 - ☹ Compromised instruments with respect to length and complexity
 - ☹ Need for additional infrastructure
- ▣ There are concerns about systematic differences when collecting similar data using different modes (Dillman 1996):
 - ▣ Higher likelihood for socially desirable responses to sensitive questions in interviewer-administered surveys (Aquilino 1994)
 - ▣ More missing data in self-administered surveys (Biemer 2003):

CLOSING REMARKS

- ▣ Telephone surveys based on landline RDD samples are subject to non-ignorable coverage bias
- ▣ Dual-frame RDD alternatives are costly and complicated
- ▣ Single-mode methods of data collection are problematic for response rate, coverage, and cost reasons
- ▣ Multi-mode methods of data collection can reduce some of the problems associated with the conventional methods
- ▣ CDSF provides a natural and efficient framework for design and implementation of multi-mode surveys
- ▣ Enhancing the CDSF can significantly improve its coverage and expand its utility for design and analytical applications

REFERENCES

- ▣ Aquilino, W.S. (1994). Interview mode effects in surveys of drug and alcohol use: a field experiment. *Public Opinion Quarterly*, 58, 210-40.
- ▣ Biener, L., Garrett, C.A., Gilpin, E.A., Roman, A.M., & Currivan, D.B. (2004). Consequences of declining survey response rates for smoking prevalence estimates. *American Journal of Preventive Medicine*, 27(3), 254-257.
- ▣ Biemer, P.P. & Lyberg, L.E. (2003). *Introduction to Survey Quality*, New York: John Wiley & Sons, Inc.
- ▣ Blumberg, S. J. and Luke, V. J. (2007). "Wireless Substitution: Early Release of Estimates from the National Health Interview Survey."
- ▣ Brick, J. M., J. Waksberg, D. Kulp, and A. Starer. 1995. "Bias in List-Assisted Telephone Samples." *Public Opinion Quarterly*, 59: 218-235.
- ▣ Curtin, R., Presser, S., & Singer, E. (2005). Changes in telephone survey nonresponse over the past quarter century. *Public Opinion Quarterly*, 69, 87-98.
- ▣ de Leeuw, E. & de Heer, W. (2002). Trends in household survey nonresponse: a longitudinal and international comparison. In R. M. Groves, D. A. Dillman, J. L. Eltinge (Eds.), *Survey Nonresponse* (pp. 41-54). New York: John Wiley & Sons, Inc.
- ▣ Dillman, D. A. 1991. The Design and Administration of Mail Surveys, *Annual Review of Sociology*, 17, 225-249.
- ▣ Dillman, D., Sangster, R., Tanari, J., & Rockwood, T. (1996). Understanding differences in people's answers to telephone and mail surveys. In Braverman, M.T. & Slater J.K. (eds.), *New Directions for Evaluation Series: Advances in Survey Research*. San Francisco: Jossey-Bass.

REFERENCES

- ▣ Dohrmann, S., Han, D. & Mohadjer, L. (2006). Residential Address Lists vs. Traditional Listing: Enumerating Households and Group Quarters. *Proceedings of the American Statistical Association, Survey Methodology Section*, Seattle, WA. pp. 2959- 2964.
- ▣ Groves, R.M. (2005). *Survey Errors and Survey Costs*, New York: John Wiley & Sons, Inc.
- ▣ Fahimi, M., M. W. Link, D. Schwartz, P. Levy & A. Mokdad (2008). “Tracking Chronic Disease and Risk Behavior Prevalence as Survey Participation Declines: Statistics from the Behavioral Risk Factor Surveillance System and Other National Surveys.” *Preventing Chronic Disease (PCD)*, Volume 5: No. 3.
- ▣ Fahimi, M., D. Creel, P. Siegel, M. Westlake, R. Johnson, & J. Chromy (2007b). “Optimal Number of Replicates for Variance Estimation.” *Third International Conference on Establishment Surveys (ICES-III)*, Montreal, Canada.
- ▣ Fahimi, M., Chromy J., Whitmore W., & Cahalan M. Efficacy of Incentives in Increasing Response Rates. (2004). *Proceedings of the Sixth International Conference on Social Science Methodology*. Amsterdam, Netherlands.
- ▣ Fahimi, M., D. Kulp, and M. Brick (2009). “A reassessment of List-Assisted RDD Methodology.” *Public Opinion Quarterly*, Vol. 73 (4): 751–760.
- ▣ Gary, S. (2003). *Is it Safe to Combine Methodologies in Survey Research?* MORI Research Technical Report.
- ▣ Iannacchione, V., Staab, J., & Redden, D. (2003). Evaluating the use of residential mailing addresses in a metropolitan household survey. *Public Opinion Quarterly*, 76:202-210.

REFERENCES

- ▣ Link, M., M. Battaglia, M. Frankel, L. Osborn, & A. Mokdad. (2006). Addressed-based versus Random-Digit-Dial Surveys: Comparison of Key Health and Risk Indicators. *American Journal of Epidemiology*, 164, 1019 - 1025.
- ▣ Link, M.W., Battaglia, M.P., Frankel, M.R., Osborn, L. and Mokdad., A.H. (2008). Comparison of address based sampling (ABS) versus random-digit dialing (RDD) for general population surveys. *Public Opinion Quarterly*.
- ▣ O'Muircheartaigh, C., Eckman, S., & Weiss, C. (2003). Traditional and enhanced field listing for probability sampling. *Proceedings of the American Statistical Association, Survey Methodology Section* (CD-ROM), Alexandria, VA, pp.2563- 2567.
- ▣ Staab, J.M., & Iannacchione, V.G. (2004). Evaluating the use of residential mailing addresses in a national household survey. *Proceedings of the American Statistical Association, Survey Methodology Section* (CD-ROM), Alexandria, VA, pp.4028- 4033.
- ▣ Voogt, R. & Saris, W. (2005). Mixed mode designs: finding the balance between nonresponse bias and mode effects. *Journal of Official Statistics*. 21, 367-387.
- ▣ Wilson, C., Wright, D., Barton, T. & Guerino, P. (2005). "Data Quality Issues in a Multi-mode Survey" Paper presented at the *Annual Meeting of the American Association for Public Opinion Research*, Miami, FL.

CONTACT INFORMATION

Mansour Fahimi

mfahimi@m-s-g.com

240-477-8277