

Administrative Records in the ACS: Discussant Presentation

Barbara 'babs' Battenfield
University of Colorado – Boulder

30 March 2018

Background on Administrative Records Use for ACS

The American Community Survey (ACS) is conducted every year.

- Provides up-to-date information about social and economic needs of community.
- ACS shows how people live -- education, income, housing, jobs. etc.
- Results may be used to decide where new schools, hospitals, and emergency services are needed.

(Information taken from the Census ACS website)

Background on Administrative Records Use for ACS

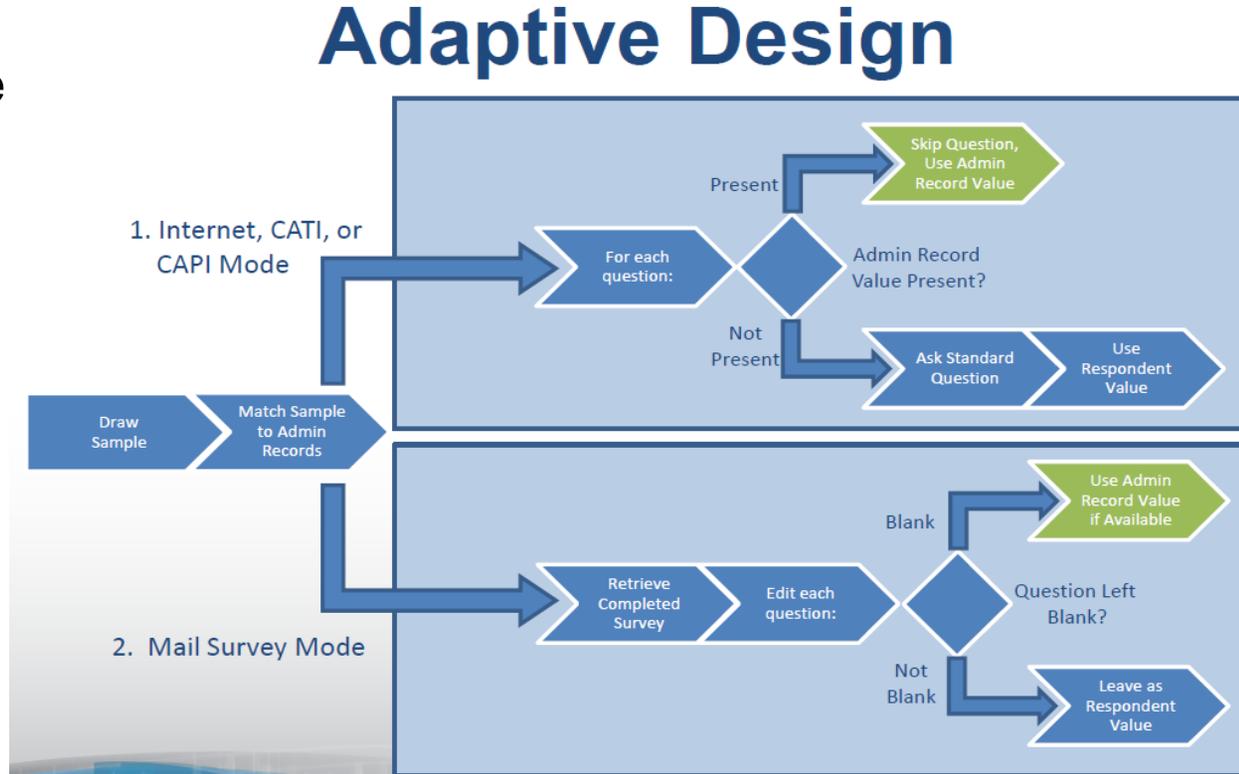
- Concerns about respondent burden and perceived intrusiveness led to Congressional efforts to make ACS voluntary
 - Census testing shows this would lower response rates
 - Would increase costs to maintain sample size (estimate \$90 million per year)
 - Would increase error and bias rates (Margins of Error)
- Census wants to minimize concerns about respondent burden while maintaining quality of data collection
 - Solicited input from NAS Cttee. for National Statistics (CNSTAT) Spring 2016
 - Meetings with Harvard Behavioral Insight Group December 2016
- Changes described in series of *Agility in Action* reports, including the use of alternative data sources (Administrative Records, or ARs)

Administrative Records (ARs) for Housing

- Use ACS 2015 responses as AR source for mail, Internet, CATI, CAPI modes of data collection
- Simulate 1-year and 5-year data products on property tax, property value, year built, and acreage
 - These four topics found to be a promising use of ARs for housing data collection
- Testing (for these four topics) :
 - Create simulated products and match to 2015 ACS responses
 - Explore production system changes needed to use ARs effectively

Questions about Adaptive Design

When response missing for mail mode (lower half of diagram), use AR – that seems clear.

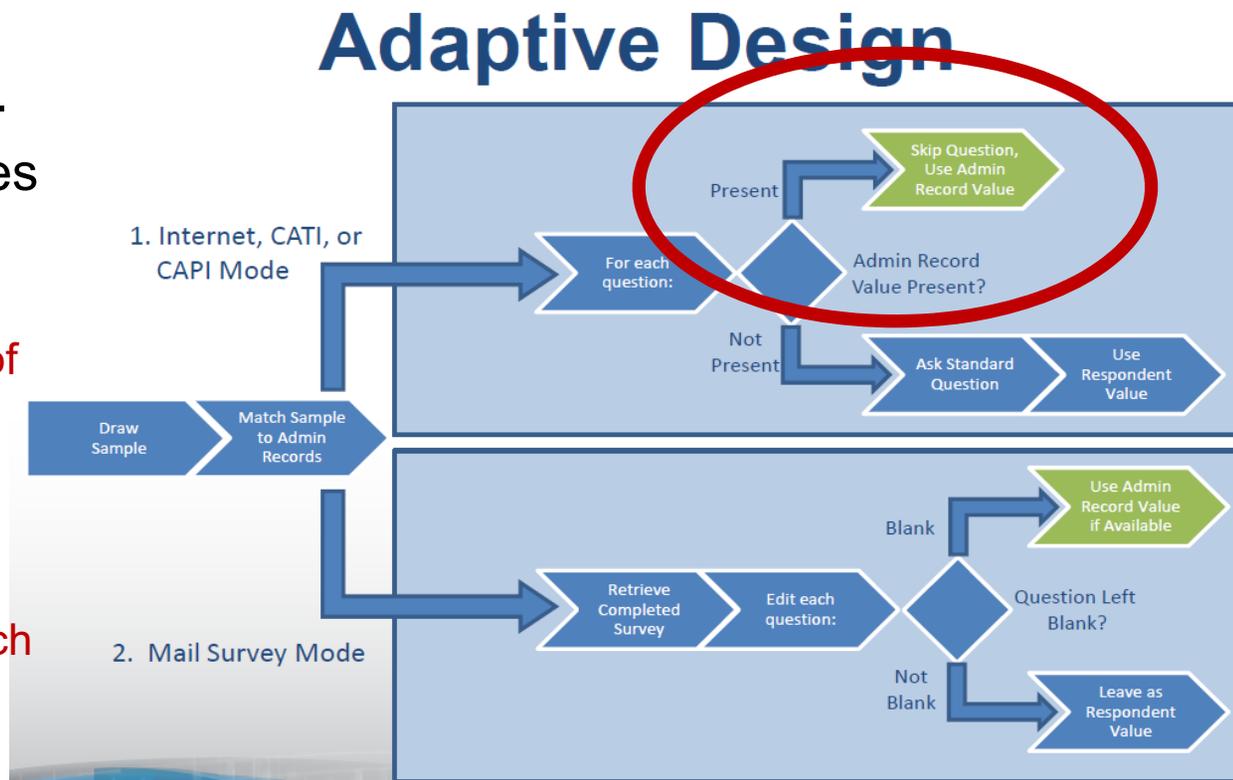


Questions about Adaptive Design

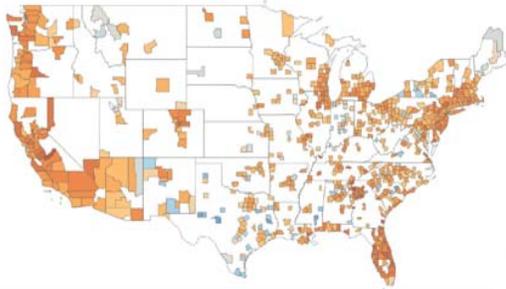
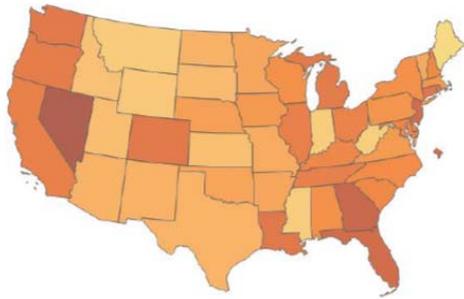
When response missing for mail mode, use AR – that seems clear.

But for Internet, CATI, CAPI modes -- is AR used instead of soliciting actual response?

- Why not ask for confirmation instead of assuming the AR is correct?
- Could reduce respondent burden but could also reduce quality – which is higher priority?
- What percentages of responses in each mode?



Questions about Results: % Difference in Property Value

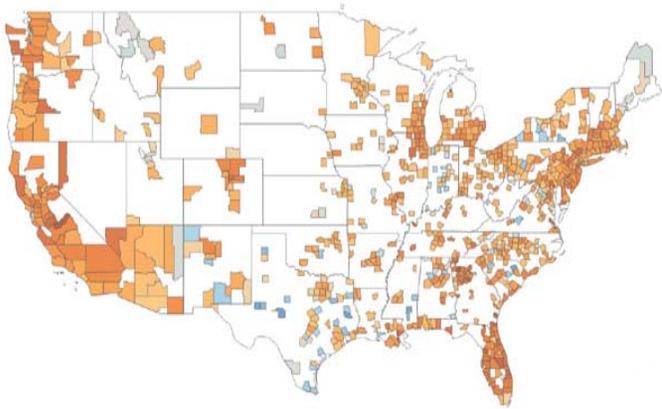


- What year is Test, what year is Control? Are property values standardized for inflation?
- Are blank areas not statistically different?
- For places, are circle sizes population counts? Is size accounted for in simulation?

- State Median Range = -12 to 0 (rough)
Interquartile = ~ 4.0
- County Median Range = -19 to 11 (rough)
Interquartile = ~ 6.0
- Place Median Range = -25 to 22 (excl. highest)
Interquartile = ~ 6.0

Surprising that IQ Range unchanged for smallest enumeration units

Questions about Results: % Difference in Property Value



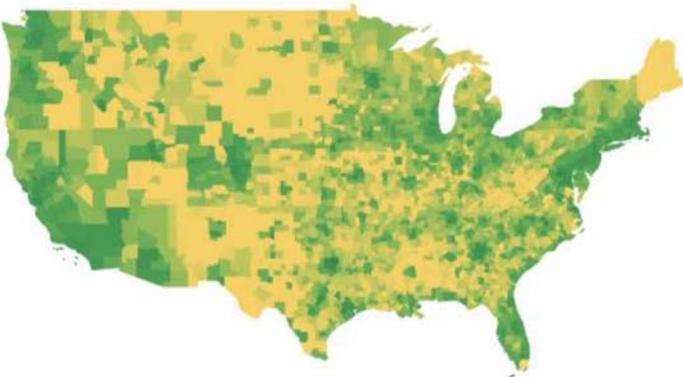
Compare differences for Counties with matching ARs...

% ACS households with % AR match (gold is low, green is high):

Areas with higher match (green) but few significant differences – eastern Oregon, Nevada, Idaho, Utah, west Colorado

Areas with lower match (gold) but more significant differences – New Mexico, west Texas, eastern Montana, Dakotas, Appalachia

Are these regions distinguished by one or more demographic characteristics that could be incorporated into the simulations?



Challenge: single outside vendor (CoreLogic)

AR Working Group was asked to identify possible vendors (thanks to Krishna Rao!)

	CoreLogic (CL)	Realty Trac	Black Knight (BK)	Zillow	First American
Notes	Real Estate (RE) leader	RE foreclosed properties	Mortg sol'ns and RE industry srvc	Adding internal acquisition	Title insurance, RE services
Database (DB) Source	Built own DB >4.5B recs, spans 50 yrs	Licensed from CL	Built own DB	Leveraged BK, adding to it	Acquired from BK
Geogr. Coverage	3100+ cnty, 99% prop'ty recs, 95% tax dist, 80% mrtg's	2575 cnty, aggregated prop'ty recs	3000+ cnty, 99% prop'ty recs, 70% mrtg's	3100+ cnty, 99% prop'ty recs, 70% mrtg's	3100+ cnty, 95% parcels, 94% f'close, 99% mrtg's
Products / Services	Bulk License, AVM	Bulk License	Bulk License	API flexible, AVM	Bulk License
Issues	Proprietary methd, API probs	Relies on CL	\$\$\$, leaving market?	Relies on BK for now	Relies on BK

Questions Posed to CSAC

- *What criteria should determine AR fitness for use?*
 - From Velkoff slides: reliability, accuracy, consistency, timeliness, (quality)
 - Run statistical models to determine contribution of each criterion
- Confession – not enough time to poll Working Group or CSAC, so these suggestions are open for discussion
 - Reduce respondent burden
 - Higher match rates with ARs
 - Accuracy → Reduction in MoE values
 - Currentness or closest temporal match with ARs
 - Geographic / demographic characteristics
 - Others?

Questions Posed to CSAC

- *Best way to introduce ARs into ACS production?*
 - Run in parallel temporarily?
 - Run for selected parts of the country? (Base these on prioritized criteria, and / or regions with similar characteristics, as shown in slide 8)
 - Run for selected questions or topics based on strongest results
- *How to communicate the change in estimates that use ARs?*
 - *Flag records for AR fields*
 - *Add field containing modified MoE*