Use of Administrative Records to Reduce Burden and Improve Quality: A Discussion

Michael Davern, Ph.D.

March 8, 2016

NORC at the UNIVERSITY of CHICAGO

Disclaimer

NORC at the UNIVERSITY of CHICAGO

Opinions and conclusions expressed here are those of the discussant only and do not necessarily represent the views of NORC at the University of Chicago.

Outline of Discussion

- How you envision a survey will impact how you think about direct substitution
- Substitution is a long run solution for the ACS to reduce burden
- Other more immediate solutions for the ACS are available that will also increase quality and potentially reduce burden
- Work administrative data into ACS data products now!

How do you envision a survey?

- Do you see a series of static questions on piece of paper?

ice

ealth insurance plan – Specify

or does he/she have aring?

does he/she have ng even when wearing

□ Never married → SKIP to 1

21 In the PAST 12 MONTHS did this person get –

	Yes	No
a. Married?	□	□
b. Widowed?	□	□
c. Divorced?	□	□

22 How many times has this person been married?

□ Once
□ Two times
□ Three or more times

23 In what year did this person last get married?

Year

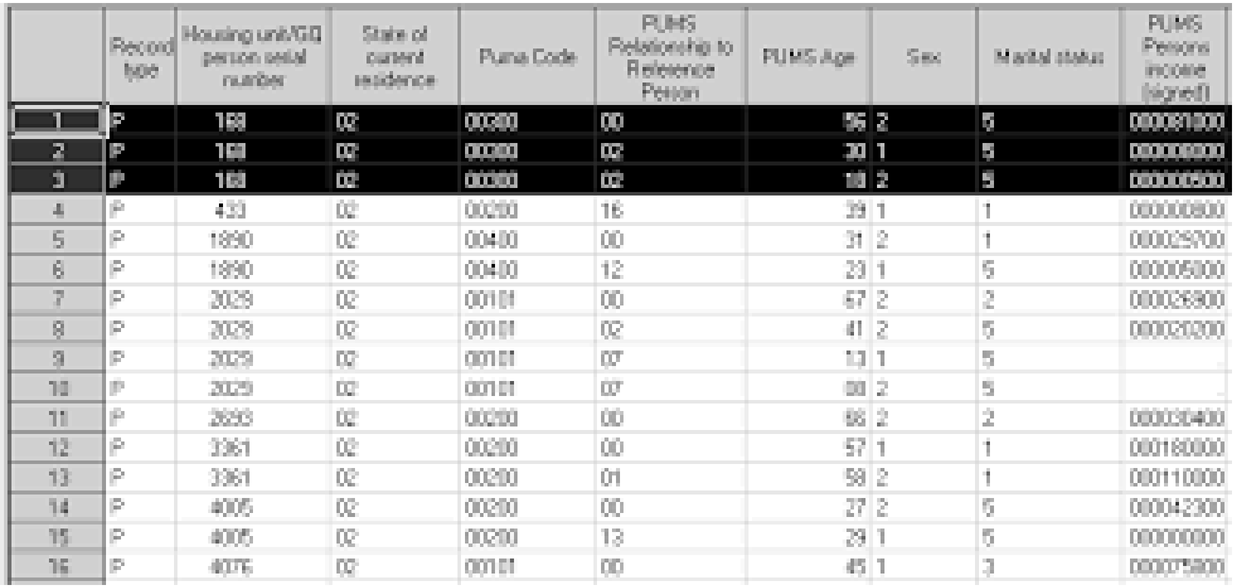
□ Neve ques
□ Only or Na
□ Now o
□ On act

27 When did thi U.S. Armed F in which this p period.

□ Septemb
□ August 19 Persian G
□ May 1975 t
□ Vietnam er
□ Februar

How do you envision a survey?

- Or do you see static variables in columns and people and establishments as rows?



	Record type	Housing unit/GQ person serial number	State of current residence	Puma Code	PUMS Relationship to Reference Person	PUMS Age	Sex	Marital status	PUMS Persons income (signed)
1	P	168	02	00300	00	56	2	5	000081000
2	P	168	02	00300	02	30	1	5	000008000
3	P	168	02	00300	02	10	2	5	000000500
4	P	433	02	00200	16	39	1	1	000000800
5	P	1890	02	00400	00	31	2	1	000029700
6	P	1890	02	00400	12	23	1	5	000005000
7	P	2029	02	00101	00	67	2	2	000026900
8	P	2029	02	00101	02	41	2	5	000020200
9	P	2029	02	00101	07	13	1	5	.
10	P	2029	02	00101	07	00	2	5	.
11	P	2693	02	00200	00	66	2	2	000030400
12	P	3361	02	00200	00	57	1	1	000180000
13	P	3361	02	00200	01	58	2	1	000110000
14	P	4005	02	00200	00	27	2	5	000042300
15	P	4005	02	00200	13	29	1	5	000000000
16	P	4075	02	00101	00	45	1	3	000075800

I See a Dynamic Complex Web of Inter-related Parts

- In the ACS everything from survey questions to the final released micro data is interconnected in a complex web held together by a large investment in operations, processing and computer programs
  - There are complicated instruments with complicated skip patterns (some electronic some paper)
  - There are complicated processing systems that scan and ingest data in from the web, paper, telephone and in-person interviews and turns it into initial electronic formats with associated metadata
  - Variables have 100 page edit specifications and associated computer code implementing the specifications; these edit specifications call on many other variables from the survey for editing and imputation
  - There are millions of data table specifications and associated computer code
  - There are review procedures for data products review
  - There is public use file data creation, weight creation and data product review
  - There is disclosure editing, and loading the data into dissemination systems
  - Etc.

The Challenges to Making Substitution Work

- Multi-mode surveys with extremely complicated processing systems like the ACS are very hard to change and/or substitute new sources of data administrative for old self-reported data
- Amy O'Hara pointed to several immediate steps to make substitution work by 2018 in her thought experiment.
- For me the three significant challenges are:
  - Legal issues for administrative data agencies and Title 13 disclosure concerns with substitution
  - It will alter trends in ACS time series data as the error structure of the data will vary and there will likely be unanticipated secondary impacts
    - Order effects, processing, editing and/or imputation changes. Also potential need for new disclosure editing
  - Census would not control the administrative source data
    - Administrative agencies could alter the database as they sees fit.

Substitution Can Reduce Burden

- Substitution will require significant commitment from ACS and Census leadership
  - To be successful it needs to be:
    - Very well integrated within ACS operations
    - Very well funded
    - And it can not be treated as separate *ad hoc* research task
  - There will be many intermediate complications
    - leadership will need to quickly separate those complications into those that are fixable from those which are not
- Amy O'Hara's presentation offers good initial candidates including Year built and phone status
- In addition to considering more variables for substitution I hope ACS considers adding supplemental administrative data to the ACS

Other Ways to Improve Quality and Reduce Burden

- There are more immediate things the Census Bureau/ACS can do that are *post-processing steps* to increase quality and potentially reduce burden over the long run
- As these are post-processing they will not disrupt the current ACS processing system and will be easier to implement

Move Linked Data Research Into Data Products

- Federal Statistics revolve around data products as Congress and Agencies fund data products and surveys
  - Data linkage activities are often viewed as an *ad hoc* study done as part of a survey
  - We have to make sure this *ad hoc* research has as its end the goal of adding the knowledge created to the data products in some way
- Create data products that benefit from insights gathered from linked data research

Data Sources

- Amy O'Hara mentioned agency data including IRS, HUD, CMS, HHS, SSA, to which I would hope we can consider DOD, VA and potential commercial sources.
- Housing data from property value sources, taxes and deeds
- Specific programs include: Social Security, WIC, SNAP, TANF, Medicaid, Medicare, VA service connected disability

Create Restricted Use Linked Data Products

- First, supplement survey data by creating a linked administrative-survey microdata product that is:
  - – Well documented
  - – Cleaned and edited
  - – Should have weights created adjusting for non-linkage issues
    - – Refusal to link, missing identifying information, etc.
  - – Be made accessible to researchers through RDC and other avenues
  - – Created at the end stage of ACS processing so as to not interfere with the current system
- Creating and disseminating these linked data products will lead to increased research on how to use to create higher quality estimates for specific policy related purposes
  - These data will likely need to be restricted to RDC use

Create Blended Estimates for Public Use Files

- Second, Create Blended (imputed or modelled) estimates
  - This can help increase quality by reducing measurement error in survey reported items
  - Can be done at the end of processing and does not interfere with the complex and dynamic web of production
  - The end goal can be either a fully blended or imputed estimate or just simply model coefficients for outside data users

Two Substantive Examples

- Work done on the ACS linked to Medicaid administrative data has found 22% of those ACS cases linked to administrative data showing Medicaid coverage did not report having the coverage.
- Research on Food Stamps (SNAP) for New York found 26 percent of cases showing receipt in administrative data did not report receipt in the ACS.

These Data are Used for Important Policy Research

- Medicaid and SNAP are important non-cash benefits that policy researchers would like to adjust for in estimating supplemental poverty measures
  - In general underestimating receipt will over-estimate poverty for supplemental measures
- Simulation modelling done by the Congressional Budget Office, ASPE and CMS relies on survey reported data to evaluate and price policy options

Use Linked Data to Partially Correct Survey Error

- Build a model on linked data and use it to create blended estimates
  - For Medicaid enrollment in the U.S. the root mean squared error (RMSE) for the model based blended estimate is 81% lower than the RMSE for the direct estimate (this is a CPS model)
  - For SNAP receipt research on New York states showed that the model based blended procedure reduces RMSE by 93 percent.
- For these models to work well its essential to keep the self-reported enrollment/receipt status in the surveys as the SNAP and Medicaid models is the strongest predictor
  - And to reduce burden the amount dollar value information could be dropped and imputed/modelled
- The modelling greatly reduces confidentiality concerns, and can be extrapolated as needed from one geography to another or one time period to another
- These types of blending (or imputation or modeling) approaches have been used in survey research but rarely in production estimates

Discussion

- ACS data products should start incorporating administrative data into the data products to reduce burden and improve quality
  - Longer term solution of substitution
  - Shorter term solution of creating linked data files and modelled estimates
- Other methods used to improve survey quality can be expensive and/or do not have direct evidence of estimate improvement
  - For example, much is spent on improving response rates through incentives and using many attempts to get an interview but little evidence shows this leads to bias reduction
    - Also linked data can be very cost-effective way to adjust for any non-response bias

The Time is Now!

- I believe it is an imperative for the Federal Statistical system to start using linked data in the creation of data products because
  - The foundational research for use of linked administrative data and survey data has been conducted for several potential sources
  - There is clear evidence from these research projects studying linked survey and administrative data that the amount of bias due to measurement error in the survey data responses could be significantly reduced
  - The necessary infrastructure for sharing data among federal agencies and directives have been supplied by the Office of the Management and Budget
  - For the short term solutions there is no need to change the surveys only to add the linked data products at the end

Conclusion

- The data world is changing and the ACS needs to keep pace by finding ways to improve the relevancy of their data
- The days of surveys existing in a vacuum and not being merged with other sources of relevant information are over
- The ACS needs to incorporate data from other sources into its data products that will reduce burden and increase quality of the data for the critical uses they are put to