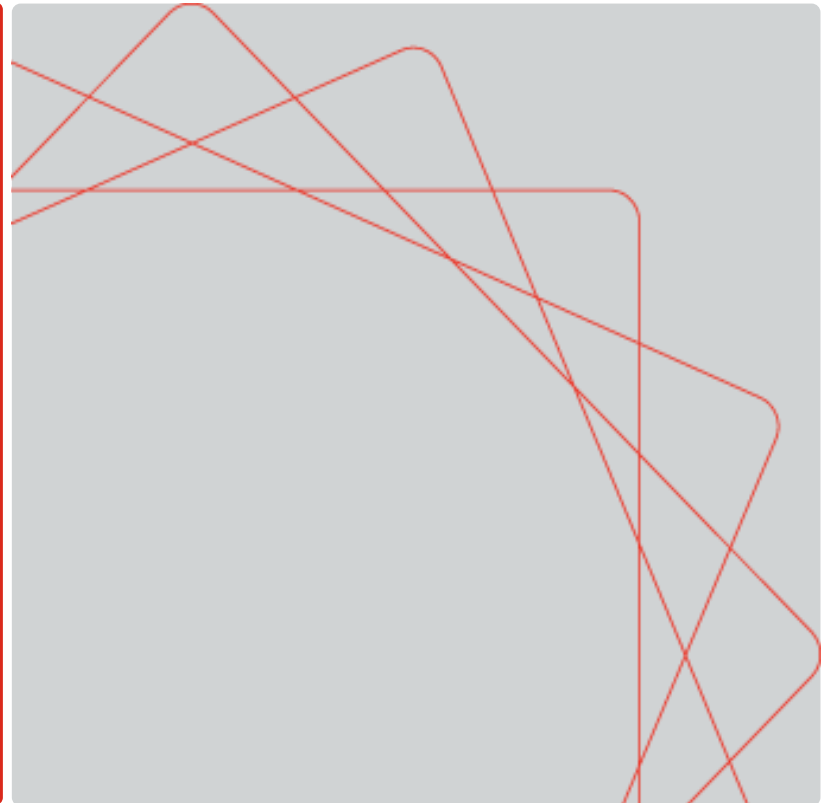




**One Outcome, Two
Methods: Comparing the
Cost, Quality and Utility of
Survey and Administrative
Data Sources**

Matthew Zeidenberg
Eric Friedman

CNCS Research Summit
December 15, 2016



Abt's Recent Work for CNCS



- In Summer 2016, Abt conducted a survey of AmeriCorps alumni from three cohorts (2005, 2010, and 2013)
 - Collected wide range of information about alumni pre- and post-service, including information about their current educational attainment and employment
- Abt prepared a report on educational outcomes of a similar sample of alumni using data from the National Student Clearinghouse (NSC)
 - NSC data is administrative data collected from institutions of higher education on periods of enrollment and degrees awarded

Surveys Versus Administrative Data



- It is much more expensive to conduct a survey than to obtain administrative data
- However, surveys can obtain data that is not collected administratively
- This presentation considers the relative advantages of each data source, based on our recent work studying the educational outcomes of AmeriCorps alumni
- We compare the data sources on information availability, cost, coverage, and data quality and statistical precision and close with recommendations

Bias and Information Availability in Surveys



- Surveys almost always have a response rate less than 100 percent, and often much less
 - This leads to non-response bias or survey sampling bias as well as a loss of precision in estimates
- Surveys also involve response bias (also called reporting bias)
- Surveys for the federal government require OMB approval
 - Research efforts cannot be duplicative or too burdensome for respondents

Information Availability from Administrative Sources



- Data is limited to what is actually collected by sources such as agencies or colleges
- Abt explored two sources relevant to AmeriCorps alumni:
 - The National Student Clearinghouse (NSC)'s enrollment and degree attainment data
 - The Longitudinal Employer-Household Dynamics (LEHD) program of the U.S. Census, which has employment data

Information Availability from Administrative Sources: the NSC



- The NSC is the leading provider of data on postsecondary enrollment and degree attainment
- Data are available to researchers for a fee if they agree to a standard-form contract
- Additional consent from AmeriCorps alumni was determined to be not required for the NSC data match

Information Availability from Administrative Sources: LEHD



- LEHD is a cooperative program between the states and the U.S. Census
- States have agreed to share microdata from the Unemployment Insurance (UI) program, which covers most workers
- UI data is quarterly earnings
- Data are made available to researchers at the Federal Statistics Research Data Centers (RDCs)
- Application process to access RDCs; researcher must go in person to carry out research and be sworn to confidentiality
- Only 17 states plus DC currently allow their data to be shared with researchers outside Census

Survey Costs



- Costs of collecting data for a survey is high
- Cost is not fixed
 - If one is surveying 1,000 people to get 800 responses, the first 100 responses are cheaper than the last 100
 - Later responses often require multiple contact attempts
 - Tradeoff between high response rate and limiting cost
- Strategies to increase response rates include:
 - Accurint lookup to improve contact information
 - Panel maintenance and regular alumni engagement
 - Completion incentive (monetary incentives typically not approved by OMB)
 - Telephone follow-up (costly)
 - In person follow-up (very costly)

Costs of Obtaining Administrative Data



- Main cost to the administrative agency is in initiating the search; there is typically minimal marginal cost to increase the number of records pulled
- Per record charge, if it exists, is typically much lower than the cost associated with collecting a survey response
- Search is typically automated to pull records from a larger database; often it is possible to get large samples, possibly up to the entire universe of interest
- Uses Social Security Number or name and DOB

Comparing Costs: Abt's Experience



- Survey cost roughly ten times the cost of NSC data match
 - Tried to survey 17,915 alumni, got 3,772 responses for a 21 percent response rate; the rate varied by cohort and program
 - Additional costs: IRB, OMB clearance, data security plan
- Total cost of the NSC match: \$2,750 fee to NSC plus labor costs for data analysis and reporting
 - Sent 3150 alumni to NSC; got 2582 back; 230 were missing and 331 were not returned due to a “FERPA block”
 - Additional costs: IRB and data security plan

Coverage



- Survey coverage issues stem from non-response and reporting bias
- NSC has known coverage issues:
 - NSC currently collects enrollments of 98 percent of the students in the country
 - However, not all institutions report degrees as report enrollments; NSC currently collects about 84 percent of degrees
 - Coverage has been worse in the past
 - No information on credit load or individual courses
 - Students can and sometimes do refuse to disclose their data (FERPA block)
 - FERPA block rate for AmeriCorps alumni sample was 11%, higher than the average of 4% across NSC database

Data Quality and Statistical Precision

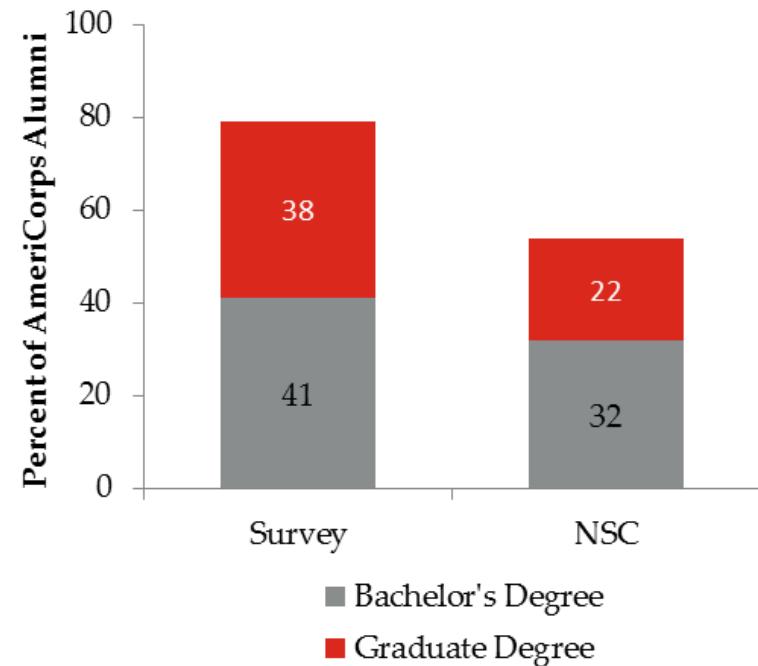


- Main data quality issues with surveys involve reporting and response bias
- In administrative data such as that of the NSC, main source of bias is missing data due to matching errors and data suppression
- Precision of estimates increases as the size of the sample does; it is cheaper to achieve large samples with administrative data
- Universal or near-universal matching is sometimes possible with administrative data, which could eliminate or reduce bias and lead to high precision

An Example: Survey and NSC Information on Degree Completion



- Lower rates for NSC are due to two factors:
 - Incomplete degree coverage: 84 percent of enrollments were at institutions that provide degree information to the NSC
 - FERPA (privacy) blocks: 11 percent



- Response bias to the survey may be inflating the survey estimate if alumni with more education were more likely to respond to the survey

Summary and Recommendations



Comparison Factor	AmeriCorps Alumni Outcomes Survey	National Student Clearinghouse Administrative Data Match
Information Availability	Survey efforts must be approved by OMB and must induce enough responses to satisfy design quota	Limited to data collected by program administrators
Cost	10x	x
Coverage	21% response rate introduces non-response bias	Degree info for 84% of enrollments, FERPA block rate of around 11% Small percentage of false negatives
Data Quality and Statistical Precision	Standard error of 1-2%	Standard error of 1-3%

- The decision to collect data via survey or an administrative source depends on budget and needs
 - Surveys are costly to administer but collect richer and more specific data, provided there are no legal, ethical, or logistical obstacles
 - Administrative data limits one to data that is collected to serve a social program or purpose
- Administrative data sources may prove to be adequate and cost effective means of obtaining simple educational and employment outcomes for program alumni, on even an annual basis
- If possible, collecting data from both sources allows one to compare the results, assess the reasons for any differences, and arrive at the best estimate for the study population

Questions?



**BOLD
THINKERS
DRIVING
REAL-WORLD
IMPACT**

