



## Great potential for linked data

- Potential for linked data:
  - Improving accuracy of survey data collection of enrollment data (Medicaid, SSI, etc.)
  - Improve survey sample frames (Census MAF)
  - Using linked data to create small area estimates
  - Improve administrative data race/ethnicity information
  - Great benefit to using information in imputation models and editing
  - Improve policy simulations by allowing researchers to better engage errors and appropriately model them





	Item non-response and missing data
	• Non-response error (or missing data)
	• Item non-response can be a major issue in administrative data and when linking
	• Important data for research can be missing (e.g., age, address, program codes, or race/ethnicity)
	• Some of these data can be missing systematically
	• E.g, VA study
	• Linking data can also be missing systematically
	• Can be a large source of sample loss when matching survey and administrative data
	<ul> <li>An example of linking the Medicaid data to the Current Population Survey</li> </ul>
SHA DAC	6











Timeliness and Data Access	
<ul> <li>Timeliness of linked data files <ul> <li>Linking takes time</li> <li>Most recent linked data for our CPS-Medicaid link is 2004</li> </ul> </li> <li>Data access <ul> <li>Due to sensitive nature of both, linking has to be done in a very restricted environment</li> <li>Access to linked data files is, by necessity, very limited</li> <li>Public use linked data files are unlikely <ul> <li>US Census Bureau Research Data Centers and/or synthetic data hold promise in this area</li> </ul> </li> </ul></li></ul>	
	12





