

Reducing the Revolving Door of Incarceration and Homelessness in the District of Columbia: Data Sources, Methods, and Limitations

Sam Hall, Martha Burt, Caterina G. Roman, and Jocelyn Fontaine
March 16, 2009

The data used in our analyses include:

- **Jail Use**—Department of Corrections (DOC) data on all people with one or more jail episodes between October 1, 2004 and March 31, 2008 but who were *not* in jail on March 31, 2008 (i.e., all episodes for each person are completed).
- **Shelter Use**—Homeless Management Information System (HMIS) data for people using the public emergency shelter system between October 1, 2005 and September 30, 2007.
- Fire and Emergency Medical Services (FEMS) data for all calls during an 8 month period from January 1 through August 31, 2008 for which an Electronic Patient Care Report was filled out and contained a full name.

Specific Data Used, Matching Techniques, and Limitations

The basic technique used to produce the results we report is **matching**—comparing information on people in one system to information on people in another system to see if they are the same people. Cross-system data matching is at least as much an art form as a simple technique, due to missing data, misspellings, different forms of the same name, and other issues. Usually about 75-80 percent of matches are clear and straightforward, but it is not unusual for 20-25 percent to require some degree of judgment to decide whether someone in one system is the same person as appears in another system. The reader may want to know the decision rules we followed for the matching we report; they are given below.

Due to missing observations for many of the descriptive demographic variables in the data sets, most notably HMIS and FEMS, we were forced to merge on first and last names only—an inefficient and imperfect matching method. This posed two limitations. First, when a first or last name is missing or misspelled we will likely not pick up the overlap across data sets. For FEMS, we were forced to disregard all those with “Jane” or “John Doe” name inputs, many of whom could be homeless individuals unable or unwilling to give identification information. Second, merging on names also links different people with the same first and last name if both appear in the data. This type of matching error is difficult or impossible to fix if the two data sets do not contain other identifiers (e.g., age, race) on which to match, and records for many people in the HMIS system are missing these identifiers. Manually sifting through all 200,000+ data points for questionable spelling and individuals with the same name was prohibitive in the time we had with the amount of missing or inaccurate variables across FEMS and HMIS data

sets (DOC data was generally complete and accurate). For now, our analysis can only be used as a close but useful estimation of the overlap.

When to consider someone in one database the same as someone in another database: We used the following criteria to declare two people from different databases a match:

- 1) If an individual had the same name in two different data sets but had no other identifying information, he or she was considered matched.
- 2) If first and last name and race and gender variables matched but no other demographic information was given, he or she was considered matched.
- 3) If dates of birth for two observations with the same name were different but had the same day and month or were within 365 days of each other, the difference in dates was assumed to be due to human error (either by those giving the information or by those taking it) and the observations were considered matched.
- 4) Between FEMS and DOC data sets, if two observations with the same name had matching social security numbers, they were a match.
- 5) Between FEMS and DOC data sets, if social security numbers were missing and dates of birth for two observations with the same name were different but had the same day and month or were within 365 days of each other, the difference in dates was assumed to be due to human error and the observations were considered matched.

When to consider someone homeless:

- 1) If he or she appears in the HMIS database
- 2) If the FEMS or DOC address is:
 - a. FEMS
 - i. "No fixed address" (and variants)
 - ii. "Homeless" (and variants)
 - iii. Shelter address – e.g., "1355 NEW YORK AVE NE"
 - b. DOC
 - i. "Homeless" (and variants)
 - ii. "No fixed address" (and variants)
 - iii. 00000 (and variants)
 - iv. "Shelter"
 - v. Shelter address or name – e.g., "La Casa Shelter"

When to consider someone DMH eligible (only possible to estimate if the person appears in the jail data):

To make a proxy for DMH eligibility, UI researchers created a dummy variable to mark any instance of a serious and persistent mental illness in jail mental health assessments. Using information gained from interviews with DMH staff, UI considered the following types of disorders as eligible for broad DMH services: schizophrenia, bipolar disorders (I and II), borderline personality disorders, and any psychotic disorders.