

# **Economic Roundtable Procedures for De-Identifying Linked Administrative Records of Los Angeles County Residents Who Received Public Benefits**

## **Goals and Intended Uses of De-Identified Data**

This data set has been de-identified to preserve it for continued research into problems of poverty, such as homelessness, while also ensuring that the privacy of individuals represented in the data is protected.

This is the largest, most multi-faceted and extended dataset of this kind that we know of in the United States. It is a powerful research tool for combatting homelessness in Los Angeles County and across the country. It would be a great loss to both homeless individuals and taxpayers for this dataset to no longer be available due to concerns about protecting the privacy of individuals.<sup>1</sup>

The data set is important for policy research because it covers a fifteen-year time window, making it possible to describe large segments of individuals' life course. It is possible to identify long-term outcomes for individuals with specific life experiences or conditions, or to look backwards and identify the antecedent experiences and conditions of individuals who have specific types of outcomes, such as being persistently homeless or high-cost users of public services.

The data set is also important because it includes the complete universe of low-income county residents who received any type of public benefits. This eliminates the difficult problem of estimating how a subgroup within that population fits into the overall universe of poverty. Because the data set includes the total low-income population and is more diverse, it can provide information about a wider range of population groups and identify where each group fits in the overall social framework.

Another reason why the data set is important is because it provides information about many types of local public services people received. This illuminates individual and family attributes, economic needs, health and mental health conditions, substance abuse, social service needs, justice system involvement, homelessness, and cost of public services received.

The Economic Roundtable seeks to continue using this data set to support efforts to combat homelessness. Planned uses include continuing to develop more predictive analytic screening

---

<sup>1</sup> William W. Stead, M.D., Chair of the National Committee on Vital and Health Statistics, outlines the values of making de-identified health data available for research to support ongoing federal policy improvements. See Stead, William W. (2017) *Letter to U.S. HHS Secretary Price, Re: Recommendations on De-identification of Protected Health Information under HIPAA*, National Committee on Vital and Health Statistics.

tools for identifying individuals who are likely to have specific types of homeless outcomes, such as persistent homelessness, so that they can receive immediate help to avoid those outcomes. The Roundtable also seeks to use this data to develop updated typologies of outcomes from homelessness that strengthen the connection between homeless classifications and the services needed by individuals in each classification group to escape homelessness through early interventions. Because this data set is de-identified it can be made accessible to other researchers who are investigating related problems.

### **Description of the De-Identified Data Set**

The data set consists of linked administrative records for 7,220,267 unduplicated individuals who received any type of public benefits from the Los Angeles County Department of Public Services (DPSS) from 2002 through 2015. Of these individuals, 1,007,530 experienced homelessness, as indicated by using the address of a DPSS office as their mailing address. Records from the Los Angeles County departments of Health Services (DHS), Mental Health (DMH), Public Health (DPH) for substance abuse services, Children and Family Services (DCFS), Sheriff, and Probation are linked to the records of recipients of public benefits provided by DPSS. The data set contains the following 37 types of information, with data suppressed, as indicated below, when a breakout by age, gender, ethnicity, and a variable that may contain sensitive information creates a subgroup containing fewer than 11 individuals.<sup>2</sup>

1. *Random person identifier*: a unique random numerical identifier for each individual that has no relationship to any attribute, value or sequence of records in the data set. De-duplication was carried out for this data set to remove duplicate records for the same individuals in the source records. This further reduces the possibility that the source records could be used to re-identify individual records.
2. *Year of Birth*: Date of birth aggregated to year of birth, with multi-year aggregation for persons born 1900 to 1970, and 2014-2015.
3. *Sex*: Binary variable using the two categories of male and female.
4. *Ethnicity*: Ethnicity and race aggregated into the five categories of African American, Asian American / Pacific Islander, Latino, European American, and Other.
5. *Homeless in any month*: binary variable for whether the individuals were shown to be homeless in any month from 2002 through 2015 based on DPSS records. This data is suppressed for subgroups based on age, gender, ethnicity, and this variable that contain fewer than 11 individuals.
6. *Year of first homeless stint*: Year when first recorded homeless episode in the 2002-2015 time window began. This data is suppressed for subgroups based on age, gender, ethnicity, and this variable that contain fewer than 11 individuals.

---

<sup>2</sup> The cut-off of 11 individuals is established in guidelines set by the California Department of Health Care Services, *Data De-identification Guidelines* (2016), p. 15.

7. *Age when first homeless*: Roll-up of age when first homeless into eight categories. Some gaps in this data were filled through statistical interpolation, making these records different than the source records and further reducing the possibility that the source records could be used to re-identify individual records. This data is suppressed for subgroups based on age, gender, ethnicity, and this variable that contain fewer than 11 individuals.
8. *Total homeless months*: Roll-up into nine categories of the total number of months homeless during the 2002-2015 time window. Some gaps in this data were filled through statistical interpolation, making these records different than the source records and further reducing the possibility that the source records could be used to re-identify individual records. This data is suppressed for subgroups based on age, gender, ethnicity, and this variable that contain fewer than 11 individuals.
9. *Total years with homeless months*: Roll-up into five categories of the number of years with homeless months, with the first year beginning with the first month of homelessness in the 2002 to 2015 time window. Some gaps in this data were filled through statistical interpolation, making these records different than the source records and further reducing the possibility that the source records could be used to re-identify individual records. This data is suppressed for subgroups based on age, gender, ethnicity, and this variable that contain fewer than 11 individuals.
10. *Months homeless in year*: Roll-up into four categories of the number of months homeless in each year from 1 to 8 following the first month of homelessness in the 2002 to 2015 time window. Some gaps in this data were filled through statistical interpolation, making these records different than the source records and further reducing the possibility that the source records could be used to re-identify individual records. This data is suppressed for subgroups based on age, gender, ethnicity, and this variable that contain fewer than 11 individuals.
11. *Summary of household type history*: Roll-up of household structure in the 2002-2015 time window into three categories. This is not a fixed attribute, it changes over time, and is common to all people rather than a specific group of individuals, so it has very little value as an identifier. It is not sensitive information that creates any risk of harm or embarrassment.
12. *Ever in family household*: Binary variable for whether the individuals was ever in a family household in the 2002 to 2015 time window. This attribute often changes and is common to most people rather than a specific group of individuals, so it has very little value as an identifier. It is not sensitive information that creates any risk of harm or embarrassment.
13. *Ever in single household*: Binary variable for whether the individuals was ever in a family household in the 2002 to 2015 time window. This attribute often changes and is common to most people rather than a specific group of individuals so it has, so it has very little value as an identifier. It is not sensitive information that creates any risk of harm or embarrassment.

14. *Primary household structure*: Binary variable for primary household structure (single or family) in each year from 2002 through 2015. This attribute often changes and is common to most people rather than a specific group of individuals so it has, so it has very little value as an identifier. It is not sensitive information that creates any risk of harm or embarrassment.
15. *Emergency room visit*: Binary variable for whether the individual visited a DHS emergency room anytime in the 2002 to 2015 time window. This is not a stable event, it changes from year to year, so it has very little value as an identifier. Visits to emergency rooms are universal for all population groups. This is not sensitive information that creates any risk of harm or embarrassment.
16. *Hospital inpatient*: Binary variable for whether the individual was admitted as an inpatient to a DHS hospital in the 2002 to 2015 time window. This data is suppressed for subgroups based on age, gender, ethnicity, and this variable that contain fewer than 11 individuals. This is not a stable event, it changes from year to year, so it has very little value as an identifier. Hospital inpatient admissions are universal for all population groups. This is not sensitive information that creates any risk of harm or embarrassment.
17. *Medical diagnosis*: Binary variable for whether the individual ever received any medical diagnosis from either DHS or DMH. This data is suppressed for subgroups based on age, gender, ethnicity, and this variable that contain fewer than 11 individuals.
18. *Body system diagnosis*: Binary variable for whether the individual ever received a medical diagnosis from either DHS or DMH for each of 13 diagnoses at the 2-digit level of the ICD-9 CM medical diagnostic classification system (the most highly aggregated level of ICD-9 CM medical diagnoses, which is a 6-digit classification system). This data is suppressed for subgroups based on age, gender, ethnicity, and each diagnostic category that contains fewer than 11 individuals. The 13 diagnostic categories are:
  - a. Infectious and Parasitic Diseases (ICD-9 CM 001-139)
  - b. Endocrine, Nutritional and Metabolic Diseases, and Immunity Disorders (ICD-9-CM 240-279)
  - c. Mental Disorders (ICD-9-CM 290-319)
  - d. Diseases of the Nervous System and Sense Organs (ICD-9-CM 320-389)
  - e. Diseases of the Circulatory System (ICD-9-CM 390-459)
  - f. Diseases of the Respiratory System (ICD-9-CM 460-519)
  - g. Diseases of the Digestive System (ICD-9-CM 520-579)
  - h. Diseases of the Genitourinary System (ICD-9-CM 580-629)
  - i. Diseases of the Skin and Subcutaneous Tissue (ICD-9-CM 680-709)
  - j. Diseases of the Musculoskeletal System and Connective Tissue (ICD-9-CM 710-739)
  - k. Symptoms, Signs, and Ill-Defined Conditions (ICD-9-CM 780-799)
  - l. Injury and Poisoning (ICD-9-CM 800-999)

m. Supplementary Classification of Factors Influencing Health Status and Contact with Health Services (ICD-9-CM V01-V90)

19. *Psychosis diagnosis*: Binary variable for whether the individual was ever diagnosed by either DHS or DMH as having a psychotic disorder, ICD-9 CM 290-299. This data is suppressed for subgroups based on age, gender, ethnicity, and each diagnostic category that contains fewer than 11 individuals.
20. *Schizophrenia diagnosis*: Binary variable for whether the individual was ever diagnosed by either DHS or DMH as having a schizophrenic disorder, ICD-9 CM 295. This data is suppressed for subgroups based on age, gender, ethnicity, and each diagnostic category that contains fewer than 11 individuals.
21. *Outpatient visit*: Binary variable for whether the individual visited a DHS medical clinic or the clinic of a private medical partner in each year from 2006 to 2014. Visits to doctors' offices are universal for all population groups. This is not sensitive information that creates any risk of harm or embarrassment.
22. *Emergency room visit*: Binary variable for whether the individual visited a DHS emergency room in each year from 2006 to 2014. Visits to emergency rooms are universal for all population groups. This is not sensitive information that creates any risk of harm or embarrassment.
23. *Inpatient days*: Roll-up into four categories of the number of inpatient days in a DHS hospital in each year from 2006 to 2014.
24. *Foster care ever*: Binary variable for whether the individual ever received foster services. This data is suppressed for subgroups based on age, gender, ethnicity, and this variable that contain fewer than 11 individuals.
25. *Foster care*: Binary variable for whether the individual received foster care in each year from 2002 to 2015, with years 2002 to 2004 rolled-up into a single category. This data is suppressed for subgroups based on age, gender, ethnicity, and this variable that contain fewer than 11 individuals.
26. *Public benefits*: Binary variable for whether the individual received any public benefits from DPSS – food stamps, Medi-Cal or cash aid – in each year from 2002 to 2015. This data is suppressed for subgroups based on age, gender, ethnicity, and this variable that contain fewer than 11 individuals.
27. *Cash aid*: Binary variable for whether the individual received any cash aid from DPSS – CalWORKs or General Relief – in each year from 2002 to 2015. This data is suppressed for subgroups based on age, gender, ethnicity, and this variable that contain fewer than 11 individuals.
28. *Jail days*: Roll-up into three categories of the number of days the individual spent in jail in each year from 2005 to 2015. This data is suppressed for subgroups based on age, gender, ethnicity, and this variable that contain fewer than 11 individuals.
29. *Jail medical*: Binary variable for whether the individual spent any time in a jail medical facility anytime from 2005 to 2015. This data is suppressed for subgroups based on age, gender, ethnicity, and this variable that contain fewer than 11 individuals.

30. *Jail mental health*: Binary variable for whether the individual spent any time in a jail mental health facility anytime from 2005 to 2015. This data is suppressed for subgroups based on age, gender, ethnicity, and this variable that contain fewer than 11 individuals.
31. *Adult Probation*: Binary variable for whether the individual spent any time in adult probation in each year from 2005 to 2013. This data is suppressed for subgroups based on age, gender, ethnicity, and this variable that contain fewer than 11 individuals.
32. *Mental health services*: Binary variable for whether the individual received services from DMH in each year from 2006 to 2015. This data is suppressed for subgroups based on age, gender, ethnicity, and this variable that contain fewer than 11 individuals.
33. *Mental health inpatient*: Binary variable for whether the individual received acute inpatient care from DMH anytime from 2009 to 2015. This data is suppressed for subgroups based on age, gender, ethnicity, and this variable that contain fewer than 11 individuals.
34. *Substance abuse services*: Binary variable for whether the individual received substance abuse services from the Department of Public Health anytime from 2002 to 2015. This data is suppressed for subgroups based on age, gender, ethnicity, and this variable that contain fewer than 11 individuals.
35. *Substance abuse flag*: Binary variable for whether substance abuse is indicated any time from 2002 to 2015 based on a medical diagnosis from the either DHS or DMH, substance abuse services from the Department of Public Health, or a substance abuse flag from DPSS. This data is suppressed for subgroups based on age, gender, ethnicity, and this variable that contain fewer than 11 individuals.
36. *Disabling condition*: Binary variable for whether a disabling condition was identified by DPSS, DHS or DMH in each year from 2002 to 2015. This data is suppressed for subgroups based on age, gender, ethnicity, and this variable that contain fewer than 11 individuals.
37. *Average monthly cost*: Average annual monthly cost for all county services in each year 2002 to 2015. This includes costs for services from DHS, DMH, DPH, DPSS, DCFS, Sheriff, Probation, estimated costs for arrests and court hearings based on each episode of incarceration, and estimated costs for emergency medical transportation (EMT). Costs are adjusted to 2017 dollars and rounded to the nearest whole dollar.

## **Public Benefits from Use of Data for Research**

Problems growing out of poverty and social marginalization damage the fabric of society and create ongoing public costs from lost productivity and social safety net costs for health and mental health care, income maintenance, housing, and justice system services.

This data set is a powerful tool for understanding the needs of homeless individuals and developing operational tools for helping people avoid long-term homelessness through early intervention. This is far less difficult and costly than dealing with the problems that come with

persistent homelessness. Homelessness takes a heavy toll on those who experience it, especially on their health. The costs of persistent homelessness are also high for taxpayers. As periods of homelessness persist, problems become more serious and the cost of treating them more expensive.

The problem is that it is not easy to identify those people who are at high risk of becoming persistently homeless so that resources can be targeted toward early interventions that work for them.

The important benefit of screening tools that have already been developed using this data set, and new tools that will be developed, is that they enable high-risk individuals to avoid the misery and harm of persistent homelessness by identifying them when they first become homeless, or even before they are homeless, so that they can immediately get the help they need to escape homelessness. In contrast, the current system of prioritization effectively forces many people to spend at least a year on the street, or to become so ill that they become extremely expensive for the public, before they can receive the help they need.

Problems that can be investigated using this data set include:

1. Mapping the complete range of long-term outcomes for people who experience homelessness (i.e., developing new homeless typologies)
2. Screening tools to predict people's long-term trajectory when they are newly homeless so that they can immediately get the right kind of help
3. Long-term outcomes for children who experience homelessness
4. Risk factors that differentiate outcomes for children who experience homelessness
5. Long-term public costs from childhood homelessness
6. Public services used by individuals and families after cash aid ends
7. Total public costs for poor individuals and families who do not receive cash aid
8. Mapping the use of public services following jail incarceration
9. Long-term outcomes of children whose mothers were incarcerated
10. Life course of jail inmates with mental disabilities

## **Evaluation of Risks Arising from Release of De-Identified Data**

### *Probability of Re-Identification*

The probability of re-identification of persons in this data set is very low for the following reasons:<sup>3</sup>

---

<sup>3</sup> These risk factors of data source availability, replicability and distinguishability are specified by the U.S. Department of Health and Human Services (2012), *Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule*. For further explication of these risk factors see B. Malin, D. Karp, and R. Scheuermann (2010), "Technical and

1. The data set is comprised of administrative records. The data fields do not contain information found in any publicly available data that be used to re-match the records and identify specific individuals. There is a very low probability of re-matching records for the following reasons:
  - a. *Data source availability*: The only foreseeable way in which re-matching might occur would be if another researcher who had received many of the same administrative records from the county, plus additional information that identified or could help identify individuals, attempted to match those records with these de-identified records. The source data is held only by Los Angeles County and a small number of researchers.
  - b. *Replicability*: Every data field has been aggregated to create binary or categorical variables that mask specific patterns or dates of service use or individual attributes.<sup>4</sup>
  - c. *Distinguishability*: Demographic attributes that might serve as quasi-identifiers are broad and contain large populations. The smallest gender group (males) includes 3,435,679 individuals. The smallest ethnic group (Other, which includes indigenous persons and individuals with multiple ethnicities) includes 349,290 individuals. The next larger ethnic group (Asian American and Pacific Islander) includes 554,547 individuals. The smallest age group (individuals born in 2013) includes 84,664 individuals. There are not any geographic identifiers in the data set.
2. The data set is very large. It includes 7,220,267 unduplicated individuals. Because record population is so large, most combinations of individual attributes and service use are repeated in many records and do not serve to identify specific individuals.
3. Data is suppressed for subgroups based on age, gender, ethnicity, and any variable containing sensitive information when there are fewer than 11 individuals with those combination of characteristics. This rule about suppression of small cells with sensitive information applies to 120 of the 183 variables in the data set.

### *Adverse Impacts Resulting from Re-Identification*

Re-identification could cause embarrassment or even financial harm to individuals with sensitive information in their records. Information that might cause social embarrassment includes:

---

policy approaches to balancing patient privacy and data sharing in clinical and translational research," *Journal of Investigative Medicine*, 2010; 58(1): 11-18."

<sup>4</sup> Garfinkel, Simson L. (2016), *De-Identifying Government Datasets*, NIST Special Publication 800-188. U.S. Department of Commerce, National Institute of Standards and Technology. See section 4.3.3 on de-identifying dates. See also: Joint Task Force Transformation Initiative, Information Technology Laboratory, National Institute of Standards and Technology (NIST) (2012), *Guide for Conducting Risk Assessments*. U.S. Department of Commerce, National Institute of Standards and Technology (NIST).

- *Homeless experience:* 14 percent of individuals experienced homelessness. It should be noted, however, that some gaps in this data were filled through statistical interpolation, making these records different than the source records and further reducing the possibility that the source records could be used to re-identify individual records.
- *Mental health services:* 5.2 percent of individuals received services from the Department of Mental Health
- *Diagnosis of mental disorder:* 5.9 percent of individuals were diagnosed as having a mental disorder
- *Diagnosis of psychosis:* 2.5 percent of individuals were diagnosed as having a psychosis
- *Diagnosis of schizophrenia:* 0.3 percent of individuals were diagnosed as having a schizophrenic disorder
- *Mental health hospitalization:* 0.3 percent of individuals received acute mental health inpatient care
- *Foster care:* 0.4 percent of the individuals received foster care services
- *Public assistance benefits:* all of the individuals received some type of public benefits, including food stamps, Medi-Cal or cash aid for some amount of time.
- *Cash aid public assistance:* 31.5 percent of individuals received cash aid for some amount of time
- *Disabling condition:* 19 percent of individuals were identified as having a disabling condition, either temporary or permanent

Information that might cause both social embarrassment and financial harm by making it more difficult for individuals to obtain employment includes:

- *Jail history:* 3.3 percent of individuals spent time in jail
- *Jail medical:* 0.6 percent of individuals were incarcerated in a jail medical facility
- *Jail Mental Health:* 0.5 percent of individuals were incarcerated in a jail mental health facility
- *Adult probation:* 2.5 percent of individuals spent time on probation
- *Substance abuse services:* 1.8 percent of individuals received substance abuse services from DPH
- *Substance abuse flag:* 2.1 percent of individuals have an identifiable history of substance abuse based on a medical diagnosis, DPH services, or a DPSS flag.

### **Risk of Inferential Disclosure**

This refers to being able to make an inference about an individual with high probability, even if the individual was not in the data set prior to de-identification. For example, being able to infer something about an individuals because she or he is not in the data set. The only inference that could be drawn from the absence of an individual from the data set is that she or he did not

receive public assistance benefits from 2002 to 2015. This information is unlikely to cause social embarrassment or harm.

### **Loss of Data Integrity**

This refers to the risk that the de-identification process might introduce bias or inaccuracies into the dataset that result in incorrect decisions.<sup>5</sup> The data set was analyzed to identify the optimal balance between data specificity and record aggregation to preserve anonymity. Data suppression was then applied to 120 of the 183 variables in the data set. This means that when demographic variables (age, gender and ethnicity) overlaid on one of the 120 variables containing sensitive or potentially sensitive information created a cell with fewer than 11 records, the values for that variable in that cell became “0”, indicating that the data was missing.<sup>6</sup>

The suppression of sensitive data has resulted in some loss of information for small populations. For example, there is some loss of information about the jail histories of Asian women in some age categories because there are comparatively few of these individuals in the population that has been incarcerated. There is loss of information about substance abuse among some older individuals, who receive substance abuse services less often than younger individuals. And there is loss of medical diagnostic information for some groups of children because they are diagnosed with conditions such as circulatory disorders less frequently than older individuals.

Overall, most information has been preserved in the de-identified data set and it will support accurate research findings. Preserving this data in de-identified form averts the risk of incorrect decisions that might result from having insufficient information were this data set not available.

An electronic file showing the results from testing each of the 120 sensitive or potentially sensitive variables for cell size for each demographic group along with the impact of data suppression on each variable has been retained by the Economic Roundtable. Because there is a small probability that this file could contribute to record re-identification, the Roundtable’s data security protocol has been applied to safeguard the file as confidential data.

---

<sup>5</sup> The National Institute of Technical Standards states that the risk of harm to human subjects from record re-identification should be weighed against the risk of harm to the public from withholding the information. “There may be a tendency for government agencies to either overprotect or under-protect their data. . . . For example, absent the data release, external organizations will suffer the cost of re-collecting the data (if it is possible to do so), or the risk of incorrect decisions that might result from having insufficient information.” *De-Identifying Government Datasets, NIST Special Publication 800-188* (2016), p. 14.

<sup>6</sup> Masking sensitive data with a value of “0” is recommended by the California Department of Health Care Services “since a non-event cannot be identified,” *Data De-identification Guidelines* (2016), p. 15.

## De-Identification Standards

The objective of the steps taken to de-identify the data set was to remove all association between data subjects in the original identifying dataset and records in the de-identified data set. This was achieved by removing all identifiers, and transforming all quasi-identifiers through aggregation into broad categories, and suppression of sensitive or potentially sensitive data in small demographic cells within 120 variables. These steps were followed because the linkage of multi-agency information creates high-dimensionality data that theoretically could be used to single out data records and identify unique patterns that could be identifying, if these values exist in a secondary source to link against.<sup>7</sup>

### *Threat modeling*

The Economic Roundtable analyzed the additional information that would be required and is available for re-identification, including both quasi-identifiers and non-identifying values that an adversary might use for re-identification. The only viable method for re-identification that was identified was through use of the source data held by the County of Los Angeles or a research organization that has received portions of records contained in this data set. This is a small risk. The steps taken to de-identify this data set make that risk exceedingly small.

### *Minimal acceptable data quality*

Conservative guidelines were adopted for protecting against record re-identification, and within those guidelines trade-offs between levels of aggregation for different variables were optimized to preserve data quality. The guidelines were:

- Suppress data for potentially embarrassing information, even if it is a public record or a data type that has been identified as non-sensitive. This includes whether individuals spent time in jail and broad categories of medical diagnoses.
- Suppress data for small cells with fewer than 11 records when demographic information is overlaid on a sensitive variable, even if guidelines for that type of variable only require record suppression for cells with fewer than 5 records, as is the case with public assistance data.<sup>8</sup>

---

<sup>7</sup> High-dimensionality data is identified as a possible risk for re-identification by the National Institute of Technical Standards, *De-Identifying Government Datasets, NIST Special Publication 800-188* (2016), p. 39. As stated earlier, there are no secondary sources in the public domain that could be used to link against this data set. The source data is held by the County of Los Angeles and portions of it have been shared with several research organizations.

<sup>8</sup> California Department of Social Services, *Confidentiality and Security Requirements, Rules of Aggregation*, p. 8. [http://www.cdss.ca.gov/calfreshoutreach/res/XX-XXXX\\_Exhibit\\_E-Attachment\\_1.pdf](http://www.cdss.ca.gov/calfreshoutreach/res/XX-XXXX_Exhibit_E-Attachment_1.pdf)

The demographic variable that had the greatest impact on cell size was year of birth. This age identifier was aggregated to achieve an average of 131,019 records per age category, with the smallest category containing 84,664 records.

There are a small number of records with missing ethnicity, gender, or year of birth information. Since nonevents cannot be identified, this missing information reduces the probability of re-identification for these records. Even though this missing information creates small cell sizes, data suppression procedures were not applied to these small cells with missing demographic information.

## **Conclusion**

The Economic Roundtable aggregated all quasi-identifiers, reviewed the results of the trial de-identification data run, increased the level of aggregation of year-of-birth data to reduce record suppression, verified that the record suppression coding was correct, and transformed all sensitive information for the entire dataset to eliminate information in cells with fewer than 11 records.

The risk of re-identification was evaluated using the California Department of Health Care publication scoring criteria for determining if statistical masking is necessary.<sup>9</sup> The score for this de-identified data set is 9, which is less than the score of 12 that is the threshold that would indicate further data masking is needed. These scoring results are attached.

Economic Roundtable researchers analyzed every aspect of the de-identified data set in light of the question, “can this information be used to identify someone?” The result of this risk assessment is that the data set is not vulnerable to record re-identification. Every variable in the data set has been aggregated to a level of generalization that effectively eliminates the individual specificity of each dimension of information and data has been suppressed in all small cell of sensitive information to further reduce the risk of harm to individuals should any record be re-identified. In conclusion, a conservative, comprehensive, methodical, and tested approach has been taken to record de-identification. The data set is not vulnerable to re-identification.

## **Attachments**

There are two attachments. The first second is the publication scoring results shown in the format of the California Department of Health Care. The second is a data layout table showing variables, data structure, and data suppression for the de-identified data set.

---

<sup>9</sup> California Department of Health Care Services, *Data De-identification Guidelines* (2016), p. 18.

## Publication Scoring Criteria

### Criteria for determining if statistical masking is necessary

Data De-identification Guidelines (DDG), California Department of Health Care Services, November 22, 2016, Version 2.0

*Score of 12 or more points requires masking*

*Scores applicable to Economic Roundtable anonymized data set are highlighted in yellow*

Variable	Characteristics	Score
Number of Events	1000+ events in a specified population	2
(Numerator)	100-999 events	3
	11-99 events	5
	<11 events	7
Sex	Male or Female	1
Age Range	>10-year age range	2
	6-10 year age range	3
	3-5 year age range	5
	1-2 year age range	7
Race Group	White, Asian, Black or African American	2
	White, Asian, Black or African American, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, Mixed	3
	Detailed Race	4
Ethnicity	Hispanic or Latino - yes or no	2
	Detailed ethnicity	4
Race/Ethnicity Combined	This applies when race and ethnicity are collected in a single data field	
	White, Asian, Black or African American, Hispanic or Latino	2
	White, Asian, Black or African American, Hispanic or Latino, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, Mixed	3
	Detailed Race/Ethnicity	4
Language Spoken	English, Spanish, Other Language	2
	Detailed Language	4
Time – Reporting Period	5 years aggregated	-5
	2-4 years aggregated	-3
	1 year (e.g., 2001)	0
	Bi-Annual	3
	Quarterly	4
	Monthly	5
Residence Geography*	State or geography with population >2,000,000	-5
	Population 1,000,001 - 2,000,000	-3
	Population 560,001 - 1,000,000	-1
	Population 250,000 - 560,000	0
	Population 100,000 - 250,000	1
	Population 50,001 - 100,000	3
	Population 20,001 - 50,000	4
	Population ≤ 20,000	5
Service Geography*	State or geography with population >2,000,000	-5
	Population 1,000,001 - 2,000,000	-4
	Population 560,001 - 1,000,000	-3

<b>Variable</b>	<b>Characteristics</b>	<b>Score</b>
	Population 250,000 - 560,000	-1
	Population of reporting region 20,001 - 250,000	0
	Population of reporting region ≤20,000	1
	Address (Street and ZIP)	3
<b>Variable Interactions</b>	Only Events (minimum of 5), Time, and Geography (Residence or Service)	-5
	Only Events (minimum of 3), Time, and Geography (Residence or Service)	-3
	Only Events (no minimum), Time, and Geography (Residence or Service)	0
	Events, Time, and Geography (Residence or Service) + 1 variable	1
	Events, Time, and Geography (Residence or Service) + 2 variable	2
	Events, Time, and Geography (Residence or Service) + 3 variable	<b>4</b>
	<b>Economic Roundtable Data Set Score</b>	<b>9</b>

## Economic Roundtable De-Identified and Anonymized Data Set

Variable	Label	Type	Public/		Data Type	Masking	Minimum Cell Size (Overlaid Demographic & Sensitive Variables)
			Private	Sensitive			
pid3	Random person identifier	Anonymous ID	Private	Yes	Random ID		
YOB	Year of Birth - rolled up to have 84,000+ per time period	Demographic	Public	No	Categorical (55)	Aggregation	
Gender_r	Sex: binary variable, male or female, B = blank	Demographic	Public	No	Categorical (2)		
Ethnicity_r	Ethnicity roll-up - 5 categories	Demographic	Public	No	Categorical (5)	Aggregation	
Hmls_Ever	Homeless in any month based on LEADER records	Clinical	Private	Yes	Binary		Suppress cells <11
Yr_Hmls	Year of first homeless stint based on LEADER, Jan 2002 - Aug 2015	Clinical	Private	Yes	Date (Year)	Aggregation	Suppress cells <11
Age_H_r	Roll-up of age when first homeless	Clinical	Private	Yes	Categorical (8)	Top coding	Suppress cells <11
hm_sum_all	Total homeless months, based upon DPSS LEADER records	Clinical	Private	Yes	Categorical (9)	Aggregation	Suppress cells <11
HmlsYrTot	Total years with homeless months in LEADER records, beginning with HStart month	Clinical	Private	Yes	Categorical (5)	Aggregation	Suppress cells <11
Y1HmlsMth	Months homeless in year 1, beginning with HStart	Clinical	Private	Yes	Categorical (5)	Aggregation	Suppress cells <11
Y2HmlsMth	Months homeless in year 2, beginning with HStart	Clinical	Private	Yes	Categorical (5)	Aggregation	Suppress cells <11
Y3HmlsMth	Months homeless in year 3, beginning with HStart	Clinical	Private	Yes	Categorical (5)	Aggregation	Suppress cells <11
Y4HmlsMth	Months homeless in year 4, beginning with HStart	Clinical	Private	Yes	Categorical (5)	Aggregation	Suppress cells <11
Y5HmlsMth	Months homeless in year 5, beginning with HStart	Clinical	Private	Yes	Categorical (5)	Aggregation	Suppress cells <11
Y6HmlsMth	Months homeless in year 6, beginning with HStart	Clinical	Private	Yes	Categorical (5)	Aggregation	Suppress cells <11
Y7HmlsMth	Months homeless in year 7, beginning with HStart	Clinical	Private	Yes	Categorical (5)	Aggregation	Suppress cells <11
Y8HmlsMth	Months homeless in year 8, beginning with HStart	Clinical	Private	Yes	Categorical (5)	Aggregation	Suppress cells <11
hhld_type	Summary of household type history	Clinical	Private	No	Categorical (3)	Aggregation	
fam_max	Ever in Family hshld? [MAX(fam_1 to fam_164)]	Demographic	Public	No	Binary	Aggregation	
sng_max	Ever in Single hshld? [MAX(fam_1 to fam_164)]	Demographic	Public	No	Binary	Aggregation	
Hsld_St_02	Primary household structure in 2002	Demographic	Public	No	Binary Annual	Aggregation	
Hsld_St_03	Primary household structure in 2003	Demographic	Public	No	Binary Annual	Aggregation	
Hsld_St_04	Primary household structure in 2004	Demographic	Public	No	Binary Annual	Aggregation	
Hsld_St_05	Primary household structure in 2005	Demographic	Public	No	Binary Annual	Aggregation	
Hsld_St_06	Primary household structure in 2006	Demographic	Public	No	Binary Annual	Aggregation	
Hsld_St_07	Primary household structure in 2007	Demographic	Public	No	Binary Annual	Aggregation	
Hsld_St_08	Primary household structure in 2008	Demographic	Public	No	Binary Annual	Aggregation	
Hsld_St_09	Primary household structure in 2009	Demographic	Public	No	Binary Annual	Aggregation	
Hsld_St_10	Primary household structure in 2010	Demographic	Public	No	Binary Annual	Aggregation	
Hsld_St_11	Primary household structure in 2011	Demographic	Public	No	Binary Annual	Aggregation	
Hsld_St_12	Primary household structure in 2012	Demographic	Public	No	Binary Annual	Aggregation	
Hsld_St_13	Primary household structure in 2013	Demographic	Public	No	Binary Annual	Aggregation	

Variable	Label	Type	Public/		Data Type	Masking	Minimum Cell Size (Overlaid Demographic & Sensitive Variables)
			Private	Sensitive			
Hsld_St_14	Primary household structure in 2014	Demographic	Public	No	Binary Annual	Aggregation	
Hsld_St_15	Primary household structure in 2015	Demographic	Public	No	Binary Annual	Aggregation	
Emergency	Flag for any county emergency room visit	Clinical	Private	No	Binary	Aggregation	
Inpatient	Flag for any county hospital inpatient admission	Clinical	Private	Yes	Binary	Aggregation	Suppress cells <11
Med_Diag	Flag for medical diagnosis by DHS or DMH	Clinical	Private	Yes	Binary	Aggregation	Suppress cells <11
Infection	Infectious and Parasitic Diseases (001-139) DHS & DMH diagnoses	Clinical	Private	Yes	Binary	Aggregation	Suppress cells <11
Endocrine	Endocrine, Nutritional and Metabolic Diseases, and Immunity Disorders (ICD-9-CM 240-279) DHS & DMH diagnoses	Clinical	Private	Yes	Binary	Aggregation	Suppress cells <11
Mental	Mental Disorders (ICD-9-CM 290-319) DHS & DMH diagnoses	Clinical	Private	Yes	Binary	Aggregation	Suppress cells <11
Nervous	Diseases of the Nervous System and Sense Organs (ICD-9-CM 320-389) DHS & DMH diagnoses	Clinical	Private	Yes	Binary	Aggregation	Suppress cells <11
Circulatory	Diseases of the Circulatory System (ICD-9-CM 390-459) DHS & DMH diagnoses	Clinical	Private	Yes	Binary	Aggregation	Suppress cells <11
Respiratory	Diseases of the Respiratory System (ICD-9-CM 460-519) DHS & DMH diagnoses	Clinical	Private	Yes	Binary	Aggregation	Suppress cells <11
Digestive	Diseases of the Digestive System (ICD-9-CM 520-579) DHS & DMH diagnoses	Clinical	Private	Yes	Binary	Aggregation	Suppress cells <11
Genit_Urin	Diseases of the Genitourinary System (ICD-9-CM 580-629) DHS & DMH diagnoses	Clinical	Private	Yes	Binary	Aggregation	Suppress cells <11
Skin	Diseases of the Skin and Subcutaneous Tissue (ICD-9-CM 680-709) DHS & DMH diagnoses	Clinical	Private	Yes	Binary	Aggregation	Suppress cells <11
Musc_Skel	Diseases of the Musculoskeletal System and Connective Tissue (ICD-9-CM 710-739) DHS & DMH diagnoses	Clinical	Private	Yes	Binary	Aggregation	Suppress cells <11
Symptoms	Symptoms, Signs, and Ill-Defined Conditions (ICD-9-CM 780-799) DHS & DMH diagnoses	Clinical	Private	Yes	Binary	Aggregation	Suppress cells <11
Injury	Injury and Poisoning (ICD-9-CM 800-999) DHS & DMH diagnoses	Clinical	Private	Yes	Binary	Aggregation	Suppress cells <11
Health_Srv	Supplementary Classification of Factors Influencing Health Status and Contact with Health Services (ICD-9-CM V01-V90) DHS & DMH diagnoses	Clinical	Private	Yes	Binary	Aggregation	Suppress cells <11
Psychosis	Psychosis diagnosed by DHS or DMH: ICD-9 CM 290-299	Clinical	Private	Yes	Binary	Aggregation	Suppress cells <11
Schizophrenia	Schizophrenic condition diagnosed by DHS or DMH: ICD-9 CM 295	Clinical	Private	Yes	Binary		Suppress cells <11

Variable	Label	Type	Public/		Data Type	Masking	Minimum Cell Size (Overlaid Demographic & Sensitive Variables)
			Private	Sensitive			
DHS_OP.2006	DHS Outpatient & public-private partnership clinic visits in 2006	Clinical	Private	No	Categorical (2)	Aggregation	
DHS_OP.2007	DHS Outpatient & public-private partnership clinic visits in 2007	Clinical	Private	No	Categorical (2)	Aggregation	
DHS_OP.2008	DHS Outpatient & public-private partnership clinic visits in 2008	Clinical	Private	No	Categorical (2)	Aggregation	
DHS_OP.2009	DHS Outpatient & public-private partnership clinic visits in 2009	Clinical	Private	No	Categorical (2)	Aggregation	
DHS_OP.2010	DHS Outpatient & public-private partnership clinic visits in 2010	Clinical	Private	No	Categorical (2)	Aggregation	
DHS_OP.2011	DHS Outpatient & public-private partnership clinic visits in 2011	Clinical	Private	No	Categorical (2)	Aggregation	
DHS_OP.2012	DHS Outpatient & public-private partnership clinic visits in 2012	Clinical	Private	No	Categorical (2)	Aggregation	
DHS_OP.2013	DHS Outpatient & public-private partnership clinic visits in 2013	Clinical	Private	No	Categorical (2)	Aggregation	
DHS_OP.2014	DHS Outpatient & public-private partnership clinic visits in 2014	Clinical	Private	No	Categorical (2)	Aggregation	
DHS_ER.2006	DHS Emergency room visits in 2006	Clinical	Private	No	Categorical (3)	Aggregation	
DHS_ER.2007	DHS Emergency room visits in 2007	Clinical	Private	No	Categorical (3)	Aggregation	
DHS_ER.2008	DHS Emergency room visits in 2008	Clinical	Private	No	Categorical (3)	Aggregation	
DHS_ER.2009	DHS Emergency room visits in 2009	Clinical	Private	No	Categorical (3)	Aggregation	
DHS_ER.2010	DHS Emergency room visits in 2010	Clinical	Private	No	Categorical (3)	Aggregation	
DHS_ER.2011	DHS Emergency room visits in 2011	Clinical	Private	No	Categorical (3)	Aggregation	
DHS_ER.2012	DHS Emergency room visits in 2012	Clinical	Private	No	Categorical (3)	Aggregation	
DHS_ER.2013	DHS Emergency room visits in 2013	Clinical	Private	No	Categorical (3)	Aggregation	
DHS_ER.2014	DHS Emergency room visits in 2014	Clinical	Private	No	Categorical (3)	Aggregation	
DHS_Inp.2006	DHS Inpatient days in 2006	Clinical	Private	No	Categorical (4)	Aggregation	
DHS_Inp.2007	DHS Inpatient days in 2007	Clinical	Private	No	Categorical (4)	Aggregation	
DHS_Inp.2008	DHS Inpatient days in 2008	Clinical	Private	No	Categorical (4)	Aggregation	
DHS_Inp.2009	DHS Inpatient days in 2009	Clinical	Private	No	Categorical (4)	Aggregation	
DHS_Inp.2010	DHS Inpatient days in 2010	Clinical	Private	No	Categorical (4)	Aggregation	
DHS_Inp.2011	DHS Inpatient days in 2011	Clinical	Private	No	Categorical (4)	Aggregation	
DHS_Inp.2012	DHS Inpatient days in 2012	Clinical	Private	No	Categorical (4)	Aggregation	
DHS_Inp.2013	DHS Inpatient days in 2013	Clinical	Private	No	Categorical (4)	Aggregation	
DHS_Inp.2014	DHS Inpatient days in 2014	Clinical	Private	No	Categorical (4)	Aggregation	

Variable	Label	Type	Public/		Data Type	Masking	Minimum Cell Size
			Private	Sensitive			(Overlaid Demographic & Sensitive Variables)
FC_Svc	Foster care service flag	Social Service	Private	Yes	Binary	Aggregation	Suppress cells <11
FCsv02_04	Foster care services in one or more years 2002 to 2004	Social Service	Private	Yes	Binary	Aggregation	Suppress cells <11
FCsv05	Foster care services in 2005	Social Service	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
FCsv06	Foster care services in 2006	Social Service	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
FCsv07	Foster care services in 2007	Social Service	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
FCsv08	Foster care services in 2008	Social Service	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
FCsv09	Foster care services in 2009	Social Service	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
FCsv10	Foster care services in 2010	Social Service	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
FCsv11	Foster care services in 2011	Social Service	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
FCsv12	Foster care services in 2012	Social Service	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
FCsv13	Foster care services in 2013	Social Service	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
FCsv14	Foster care services in 2014	Social Service	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
FCsv15	Foster care services in 2015	Social Service	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
LEADER_02	2002 any public benefits - cash aid, food stamps, Medi-Cal	Public Assistance	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
LEADER_03	2003 any public benefits - cash aid, food stamps, Medi-Cal	Public Assistance	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
LEADER_04	2004 any public benefits - cash aid, food stamps, Medi-Cal	Public Assistance	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
LEADER_05	2005 any public benefits - cash aid, food stamps, Medi-Cal	Public Assistance	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
LEADER_06	2006 any public benefits - cash aid, food stamps, Medi-Cal	Public Assistance	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
LEADER_07	2007 any public benefits - cash aid, food stamps, Medi-Cal	Public Assistance	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
LEADER_08	2008 any public benefits - cash aid, food stamps, Medi-Cal	Public Assistance	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
LEADER_09	2009 any public benefits - cash aid, food stamps, Medi-Cal	Public Assistance	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
LEADER_10	2010 any public benefits - cash aid, food stamps, Medi-Cal	Public Assistance	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
LEADER_11	2011 any public benefits - cash aid, food stamps, Medi-Cal	Public Assistance	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
LEADER_12	2012 any public benefits - cash aid, food stamps, Medi-Cal	Public Assistance	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
LEADER_13	2013 any public benefits - cash aid, food stamps, Medi-Cal	Public Assistance	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
LEADER_14	2014 any public benefits - cash aid, food stamps, Medi-Cal	Public Assistance	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
LEADER_15	2015 any public benefits - cash aid, food stamps, Medi-Cal	Public Assistance	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
CashMth02	DPSS CalWORKs or General Relief cash aid in 2002	Public Assistance	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
CashMth03	DPSS CalWORKs or General Relief cash aid in 2003	Public Assistance	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
CashMth04	DPSS CalWORKs or General Relief cash aid in 2004	Public Assistance	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
CashMth05	DPSS CalWORKs or General Relief cash aid in 2005	Public Assistance	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
CashMth06	DPSS CalWORKs or General Relief cash aid in 2006	Public Assistance	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
CashMth07	DPSS CalWORKs or General Relief cash aid in 2007	Public Assistance	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
CashMth08	DPSS CalWORKs or General Relief cash aid in 2008	Public Assistance	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
CashMth09	DPSS CalWORKs or General Relief cash aid in 2009	Public Assistance	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
CashMth10	DPSS CalWORKs or General Relief cash aid in 2010	Public Assistance	Private	Yes	Binary Annual	Aggregation	Suppress cells <11

Variable	Label	Type	Public/		Data Type	Masking	Minimum Cell Size
			Private	Sensitive			(Overlaid Demographic & Sensitive Variables)
CashMth11	DPSS CalWORKs or General Relief cash aid in 2011	Public Assistance	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
CashMth12	DPSS CalWORKs or General Relief cash aid in 2012	Public Assistance	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
CashMth13	DPSS CalWORKs or General Relief cash aid in 2013	Public Assistance	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
CashMth14	DPSS CalWORKs or General Relief cash aid in 2014	Public Assistance	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
CashMth15	DPSS CalWORKs or General Relief cash aid in 2015	Public Assistance	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
JailDays05	Jail Days in 2005	Justice System	Public	Yes	Binary Annual	Aggregation	Suppress cells <11
JailDays06	Jail Days in 2006	Justice System	Public	Yes	Binary Annual	Aggregation	Suppress cells <11
JailDays07	Jail Days in 2007	Justice System	Public	Yes	Binary Annual	Aggregation	Suppress cells <11
JailDays08	Jail Days in 2008	Justice System	Public	Yes	Binary Annual	Aggregation	Suppress cells <11
JailDays09	Jail Days in 2009	Justice System	Public	Yes	Binary Annual	Aggregation	Suppress cells <11
JailDays10	Jail Days in 2010	Justice System	Public	Yes	Binary Annual	Aggregation	Suppress cells <11
JailDays11	Jail Days in 2011	Justice System	Public	Yes	Binary Annual	Aggregation	Suppress cells <11
JailDays12	Jail Days in 2012	Justice System	Public	Yes	Binary Annual	Aggregation	Suppress cells <11
JailDays13	Jail Days in 2013	Justice System	Public	Yes	Binary Annual	Aggregation	Suppress cells <11
JailDays14	Jail Days in 2014	Justice System	Public	Yes	Binary Annual	Aggregation	Suppress cells <11
JailDays15	Jail Days in 2015 - data for January to August	Justice System	Public	Yes	Binary Annual	Aggregation	Suppress cells <11
Jail_Med	Jail medical incarceration facility any year	Justice System	Private	Yes	Binary	Aggregation	Suppress cells <11
Jail_MH	Jail mental health incarceration facility any year	Justice System	Private	Yes	Binary	Aggregation	Suppress cells <11
Prob.2005	Adult Probation services in 2005	Justice System	Public	Yes	Binary Annual	Aggregation	Suppress cells <11
Prob.2006	Adult Probation services in 2006	Justice System	Public	Yes	Binary Annual	Aggregation	Suppress cells <11
Prob.2007	Adult Probation services in 2007	Justice System	Public	Yes	Binary Annual	Aggregation	Suppress cells <11
Prob.2008	Adult Probation services in 2008	Justice System	Public	Yes	Binary Annual	Aggregation	Suppress cells <11
Prob.2009	Adult Probation services in 2009	Justice System	Public	Yes	Binary Annual	Aggregation	Suppress cells <11
Prob.2010	Adult Probation services in 2010	Justice System	Public	Yes	Binary Annual	Aggregation	Suppress cells <11
Prob.2011	Adult Probation services in 2011	Justice System	Public	Yes	Binary Annual	Aggregation	Suppress cells <11
Prob.2012	Adult Probation services in 2012	Justice System	Public	Yes	Binary Annual	Aggregation	Suppress cells <11
Prob.2013	Adult Probation services in 2013	Justice System	Public	Yes	Binary Annual	Aggregation	Suppress cells <11
DMHsrv06	DMH services in 2006	Clinical	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
DMHsrv07	DMH services in 2007	Clinical	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
DMHsrv08	DMH services in 2008	Clinical	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
DMHsrv09	DMH services in 2009	Clinical	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
DMHsrv10	DMH services in 2010	Clinical	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
DMHsrv11	DMH services in 2011	Clinical	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
DMHsrv12	DMH services in 2012	Clinical	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
DMHsrv13	DMH services in 2013	Clinical	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
DMHsrv14	DMH services in 2014	Clinical	Private	Yes	Binary Annual	Aggregation	Suppress cells <11

Variable	Label	Type	Public/		Data Type	Masking	Minimum Cell Size (Overlaid Demographic & Sensitive Variables)
			Private	Sensitive			
DMHsrv15	DMH services in 2015 - data for January to August	Clinical	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
DMH_AC_Imp	DMH acute inpatient care 2009 to 2015	Clinical	Private	Yes	Binary	Aggregation	Suppress cells <11
DPH_ever	DPH substance abuse services sometime from 2002 to 2015	Clinical	Private	Yes	Binary	Aggregation	Suppress cells <11
SA_Flag	Substance abuse flag (recode of DPH Dx, DHS Dx, DPSS flag) 2002-2015	Clinical	Private	Yes	Binary	Aggregation	Suppress cells <11
Disability02	Disabling condition identified by DPSS in 2002	Clinical	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
Disability03	Disabling condition identified by DPSS or DHS in 2003	Clinical	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
Disability04	Disabling condition identified by DPSS or DHS in 2004	Clinical	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
Disability05	Disabling condition identified by DPSS or DHS in 2005	Clinical	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
Disability06	Disabling condition identified by DPSS, DHS or DMH in 2006	Clinical	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
Disability07	Disabling condition identified by DPSS, DHS or DMH in 2007	Clinical	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
Disability08	Disabling condition identified by DPSS, DHS or DMH in 2008	Clinical	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
Disability09	Disabling condition identified by DPSS, DHS or DMH in 2009	Clinical	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
Disability10	Disabling condition identified by DPSS, DHS or DMH in 2010	Clinical	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
Disability11	Disabling condition identified by DPSS, DHS or DMH in 2011	Clinical	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
Disability12	Disabling condition identified by DPSS, DHS or DMH in 2012	Clinical	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
Disability13	Disabling condition identified by DPSS, DHS or DMH in 2013	Clinical	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
Disability14	Disabling condition identified by DPSS, DHS or DMH in 2014	Clinical	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
Disability15	Disabling condition identified by DPSS, DHS or DMH in 2015	Clinical	Private	Yes	Binary Annual	Aggregation	Suppress cells <11
Cost2002	Average annual monthly cost2002, 2017\$, rounded to \$10	Summary	Private	No	Numeric, rounded	Aggregation	
Cost2003	Average annual monthly cost2003, 2017\$, rounded to \$10	Summary	Private	No	Numeric, rounded	Aggregation	
Cost2004	Average annual monthly cost2004, 2017\$, rounded to \$10	Summary	Private	No	Numeric, rounded	Aggregation	
Cost2005	Average annual monthly cost2005, 2017\$, rounded to \$10	Summary	Private	No	Numeric, rounded	Aggregation	
Cost2006	Average annual monthly cost2006, 2017\$, rounded to \$10	Summary	Private	No	Numeric, rounded	Aggregation	
Cost2007	Average annual monthly cost2007, 2017\$, rounded to \$10	Summary	Private	No	Numeric, rounded	Aggregation	
Cost2008	Average annual monthly cost2008, 2017\$, rounded to \$10	Summary	Private	No	Numeric, rounded	Aggregation	
Cost2009	Average annual monthly cost2009, 2017\$, rounded to \$10	Summary	Private	No	Numeric, rounded	Aggregation	
Cost2010	Average annual monthly cost2010, 2017\$, rounded to \$10	Summary	Private	No	Numeric, rounded	Aggregation	
Cost2011	Average annual monthly cost2011, 2017\$, rounded to \$10	Summary	Private	No	Numeric, rounded	Aggregation	
Cost2012	Average annual monthly cost2012, 2017\$, rounded to \$10	Summary	Private	No	Numeric, rounded	Aggregation	
Cost2013	Average annual monthly cost2013, 2017\$, rounded to \$10	Summary	Private	No	Numeric, rounded	Aggregation	
Cost2014	Average annual monthly cost2014, 2017\$, rounded to \$10	Summary	Private	No	Numeric, rounded	Aggregation	
Cost2015	Average annual monthly cost2015, 2017\$, rounded to \$10	Summary	Private	No	Numeric, rounded	Aggregation	