

**Office of the Secretary Patient-Centered Outcome Research Trust Fund Project**

**HP-21-008**

**National Center for Health Statistics  
Centers for Disease Control and Prevention**

***Building infrastructure and evidence for COVID-19 related research, using  
integrated data from NCHS Data Linkage Program***

**FINAL REPORT**

**April 2026**

## Table of Contents

Team Members .....	3
List of Acronyms .....	4
1. Executive Summary .....	5
2. Project Overview .....	5
2.1 Objective .....	5
2.2 Data Sources .....	6
2.2.1 Synthetic Linked 2018 NHIS-HUD-CMS Data .....	6
2.2.2 Synthetic Linked 2016 NHCS-NDI Data .....	7
2.3 Verification Process .....	11
2.4 Tasks, Objectives, and Deliverables .....	11
3. Major Accomplishments .....	13
3.1 Synthetic Data Generation Methodology .....	13
3.2 Verification Process .....	13
4. Lessons Learned .....	13
5. Presentations and Publications .....	14
5.1 Presentations .....	14
5.2 Publications .....	14
6. Future Considerations .....	15
7. Summary .....	15
8. Accessing the NCHS Synthetic Linked Data and Requesting Verification Metrics .....	15
9. References .....	16

## Team Members

<b>Team Member Organizations</b>	<b>Team Members</b>
Division of Analysis and Epidemiology, National Center for Health Statistics, Centers for Disease Control and Prevention	Cordell Golden – Project Lead Crescent Martin Frances McCarty
Georgia Tech Research Institute	Richard Boyd Austin Himschoot Jordan Chandler
Duke University	Jerome P. Reiter

## List of Acronyms

AHRQ, Agency for Healthcare Research and Quality  
CCD, Continuity of Care Document  
CLH, Community-Level Health  
CMS, Center for Medicare & Medicaid Services  
ED, emergency department  
EHR, electronic health records  
HCV, Housing Choice Vouchers program  
HUD, Department of Housing and Urban Development  
IP, inpatient  
MF, Multifamily program  
NCHS, National Center for Health Statistics  
NDI, National Death Index  
NHCS, National Hospital Care Survey  
NHIS National Health Interview Survey  
NLP, Natural Language Processing  
PCOR, patient-centered outcomes research  
PH, Public Housing program  
RDC, Research Data Center  
UB-04, uniform billing form  
ZCTA, ZIP Code Tabulation Area

## **1. Executive Summary**

The overall goal of the project is to expand the utility of National Center for Health Statistics (NCHS) linked data resources to support patient-centered outcomes research (PCOR) through the development of publicly available synthetic linked data files. To protect NCHS survey participants' confidentiality, most of the linked datasets produced by the NCHS Data Linkage Program are only available as restricted-use files accessible through the NCHS or Federal Statistical Research Data Centers (RDC). Requirements to access restricted-use linked data through the RDCs have limited the utility of linked data.

To expand the utility of NCHS linked data resources in patient-centered outcomes research, NCHS has developed two synthetic linked data files; the first linking NCHS National Health Interview Survey (NHIS) data with Medicare enrollment and utilization data from the Centers for Medicare & Medicaid Services (CMS) and Department of Housing and Urban Development (HUD) federal housing program participation data; and the second linking patient encounter records from the National Hospital Care Survey (NHCS) with death certificate information from the National Death Index (NDI). The synthetic linked NCHS-HUD-CMS data file is available for public-use and the synthetic linked NHCS-NDI data file is expected to be released in 2026. The synthetic linked data were created using statistical techniques to maintain properties of the original restricted data while protecting survey participants' privacy.

To enhance the utility of the synthetic linked data, NCHS developed a verification process that allows data analysts to verify results from selected logistic regression models without having to access the original restricted-use linked data.

## **2. Project Overview**

### **2.1 Objective**

The main objective of this project was to identify and implement a statistically valid methodology for creating publicly available synthetic linked data files that would reduce the barriers associated with accessing National Center for Health Statistics (NCHS) restricted use linked data files. Data access for NCHS restricted-use data requires researchers to submit a detailed research application to be granted data access through the NCHS Research Data Center (RDC) Network. The synthetic linked data files developed under this project are accessible from the NCHS website and can be downloaded and analyzed by researchers on their own computer.

The synthetic linked data were created using statistical techniques to maintain properties of the original restricted data while protecting survey participants' privacy. Each NCHS synthetic linked data file consists of multiple synthetic data files, called implicates, which are intended to be analyzed together. Multiple implicates increase the utility of the data by providing more

appropriate variance estimates that better reflect the additional uncertainty added to the data by the synthesis process.

The project resulted in the creation of two synthetic linked data files:

- 1) Synthetic Linked 2018 NHIS-HUD-CMS data
- 2) Synthetic Linked 2016 NHCS-NDI data

[Table 1](#) provides an overview of the two synthetic linked files developed through this project.

In addition, the project created a verification process to enable synthetic linked data users to verify selected regression model results against the original restricted-use linked data. While it is not possible to support every analytic model with the NCHS synthetic linked data, the verification process provides data users with a resource for identifying instances where the restricted-use linked data must be used to support specific research objectives.

## 2.2 Data Sources

### 2.2.1 Synthetic Linked 2018 NHIS-HUD-CMS Data

The Synthetic Linked 2018 NHIS-HUD-CMS data file integrates person-level micro-data from the following data sources:

- Public-use 2018 NHIS data files
- Restricted-use 2018 NHIS-HUD linked data files
- Restricted-use 2018 NHIS-CMS Medicare linked data files
- Public-use ZCTA-level variables from the AHRQ CLH database, 2018 release

The population represented in the Synthetic Linked 2018 NHIS-HUD-CMS data is limited to 2018 NHIS survey participants 18 years or older who were linkage eligible for the NHIS-HUD data linkage (i.e., provided consent to link and necessary personally identifiable information for linkage).

#### Public-use 2018 National Health Interview Survey

The National Health Interview Survey (NHIS) is a nationally representative, cross-sectional household interview survey that serves as an important source of information on the health of the civilian, noninstitutionalized population of the United States. The sample design is a probability design that permits the representative sampling of households and noninstitutional group quarters [1]. For the 2018 survey design, from each family within a sampled household, one “sample adult” aged 18 years or older was randomly selected to complete a detailed questionnaire. Only sample adults were included in the Synthetic Linked 2018 NHIS-HUD-CMS data.

### Restricted-use 2018 NHIS-HUD linked data files

The Department of Housing and Urban Development (HUD) is the federal agency responsible for overseeing domestic housing programs and policies. While HUD administers various housing and community development programs, the linkage with 2018 NHIS focuses on HUD's three largest housing assistance programs: Housing Choice Vouchers (HCV), Public Housing (PH), and Multifamily (MF) programs. People receiving housing assistance from HUD are represented in HUD administrative data because they receive a rental subsidy or pay a below-market rent. HUD uses data about household characteristics, income, and expenses to determine the amount of the rental subsidy under federal law [2]. More information on HUD programs, their administration, and HUD data systems is provided in [A Primer on HUD Programs and Associated Administrative Data](#).

NCHS previously linked eligible survey participant information collected as part of the 2018 NHIS to administrative records from HUD [3]. More detailed information describing the methods used to conduct the NHIS-HUD linkage is available in [The Linkage of the National Center for Health Statistics \(NCHS\) Survey Data to U.S. Department of Housing and Urban Development \(HUD\) Administrative Data: Linkage Methodology and Analytic Considerations](#).

### Public-use AHRQ CLH Database, 2018 Release

The Agency for Healthcare Research and Quality (AHRQ) Community-Level Health (CLH) Database was developed to make available CLH-focused contextual data without the need to access multiple sources [4]. The CLH Database includes variables derived from the Decennial Census and American Community Survey for ZIP Code Tabulation Areas (ZCTAs). ZCTAs are geographic features made by the U.S. Census Bureau to approximate ZIP Codes. Although most ZIP Codes have a matched ZCTA Code (which have the same five digits as the ZIP Codes they represent), ZCTA boundaries are not exact representations of the United States Postal Service's ZIP Code delivery areas.

### **2.2.2 Synthetic Linked 2016 NHCS-NDI Data**

The Synthetic Linked 2016 NHCS-NDI data file integrates patient-level micro-data from the following data sources:

- Restricted-use 2016 NHCS data files
- Restricted-use linked 2016 NHCS-2016-2017 NDI data files
- Restricted-use 2016 NHCS Enhanced Opioid Identification file

The population represented in the Synthetic Linked 2016 NHCS-NDI data is limited to 2016 NHCS patient records for adults aged 18 and older who were linkage eligible for the NHCS-NDI data linkage (i.e., participating hospitals provided necessary personally identifiable information for linkage).

### Restricted-use 2016 NHCS Data

The National Health Hospital Survey (NHCS) is an establishment survey that collects inpatient (IP) and emergency department (ED) episode-level data from sampled hospitals. It is one of the National Health Care Surveys, a family of surveys covering a wide spectrum of healthcare delivery settings including ambulatory, hospital, and post-acute and long-term care providers. The goal of the NHCS is to provide reliable and timely healthcare utilization data for hospital-based settings, including prevalence of conditions, health status of patients, health services utilization, and substance-involved ED visits.

From participating hospitals, NHCS collects data on all IP and ambulatory care visits occurring during the calendar year. In previous years of the survey, hospitals were required to provide data from claims records, but to reduce the burden of reporting on participating hospitals, for the 2016 data collection hospitals were given the option of providing their data in the form of electronic health records (EHRs) or as claims records. Thus, participating hospitals provided data in the form of Uniform Bill (UB-04) administrative claim records or EHR data, where the EHR data are an amalgamation of custom extracts and Consolidated Clinical Documents (CCDs). NHCS collects patient personally identifiable information (e.g., name, date of birth, and social security number), which allows for the linkage of episodes of care across hospital units as well as to other data sources, such as the NDI [5].

### Restricted-use Linked 2016 NHCS – 2016-2017 NDI Data

Through its data linkage program, NCHS conducted a linkage of the 2016 NHCS with 2016-2017 NDI data. The NDI is a centralized database of U.S. death record information on file in state vital statistics offices. Working with these state offices, NCHS established the NDI as a resource to aid epidemiologists and other health and medical investigators with their mortality ascertainment activities. The records, which are compiled annually, include detailed information on the underlying and multiple causes of death [6]. More detailed information describing the methods used to conduct the NHCS-NDI linkage is available in [The Linkage of the 2016 National Hospital Care Survey to the 2016-2017 National Death Index: Methodology Overview and Analytic Considerations](#).

### Restricted-use 2016 NHCS Enhanced Opioid Identification File

The NCHS Division of Health Care Statistics created an Enhanced Opioid Identification Algorithm to identify opioid-involved hospital encounters in ED and IP settings [7]. The algorithm was designed to capture specific codes reported in hospital encounter records and included natural language processing (NLP) machine learning techniques for scanning physician's clinical notes to identify opioid involvement. An encounter-level file, the 2016 NHCS Enhanced Opioid Identification file was created to identify hospital IP or ED encounter records that met the criteria for opioid involvement.

The restricted-use 2016 NHCS Enhanced Opioid Identification file is available for research use via the NCHS RDC network; however, this data source does not include patient-level variables. Patient-level variables were constructed from this data source for use in the synthetic linked NHCS-NDI pilot project.

**Table 1. Overview of NCHS Synthetic Linked Data Files**

	<b>Synthetic Linked 2018 NHIS-HUD-CMS file</b>	<b>Synthetic Linked 2016 NHCS-NDI file</b>
<b>Data sources used</b>	<ol style="list-style-type: none"> <li>1. Public-use 2018 NHIS</li> <li>2. Restricted-use linked 2018 NHIS-HUD data</li> <li>3. Restricted-use linked 2018 NHIS-CMS Medicare data</li> <li>4. AHRQ Community-Level Health Database (2018)</li> </ol>	<ol style="list-style-type: none"> <li>1. Restricted-use 2016 NHCS</li> <li>2. Restricted-use 2016-2017 NDI</li> <li>3. Restricted-use 2016 NHCS Enhanced Opioid Identification file</li> </ol>
<b>Representativeness</b>	2018 NHIS participants aged 18 years and older who were eligible for HUD linkage	2016 NHCS patients aged 18 years and older who eligible for NDI linkage
<b># of synthetic data implicates / records per implicate</b>	<ul style="list-style-type: none"> <li>- 25 implicates (intended to be analyzed together)</li> <li>- 22,426 records per implicate</li> </ul>	<ul style="list-style-type: none"> <li>- 16 implicates (intended to be analyzed together)</li> <li>- 2,916,833 records per implicate</li> </ul>
<b># of variables</b>	51 variables including: <ul style="list-style-type: none"> <li>- Original synthesized variables</li> <li>- Selected recodes of original variables to support logistic regression analyses</li> <li>- Universe/subpopulation variables to subset selected analytic cohorts (e.g., ages 65 and older, income to poverty ratio &lt; 2.0)</li> </ul>	47 variables including: <ul style="list-style-type: none"> <li>- Original synthesized variables</li> <li>- Selected recodes of original variables to support logistic regression analyses</li> <li>- Universe/subpopulation variable to subset based on vital status</li> </ul>
<b>File status</b>	Approved by NCHS DRB and available on NCHS website <a href="https://www.cdc.gov/nchs/linked-data/synthetic/index.html">https://www.cdc.gov/nchs/linked-data/synthetic/index.html</a>	Approved by NCHS DRB. Anticipated release in 2026.

### 2.3 Verification Process

To enhance the utility of the NCHS Synthetic Linked Data, users will be able to verify selected results without having to access the original restricted-use linked data. The verification process is currently limited to binomial logistic regression models which will be run in R.

The verification process will provide the user with the following four metrics for each estimated model coefficient:

1. Does the sign of the coefficient estimate match for the original and synthetic estimates? (true or false)
2. Is the p-value in the same direction, above or below, with reference to 0.05 for both coefficient estimates? (true or false)
3. Is the synthetic coefficient point estimate contained within the confidence interval of the original coefficient estimate? (true or false)
4. What is the overlap of the two confidence intervals (original and synthetic)?[8] This is a calculated value between 0-1 and presented to the user in ranges: 0, (0-0.5), [0.5-0.6), [0.6-0.7), [0.7-0.8), [0.8-0.9), [0.9-1.0].

Users can submit a request for verification to the NCHS Data Linkage Program. Users may select an outcome (dependent) variable and one or more predictor (independent) variables from a prepopulated list. Additionally, the user can specify the reference levels for their predictor/independent variables and a universe or subpopulation for the model.

### 2.4 Tasks, Objectives, and Deliverables

This section outlines the tasks, the objectives, and the deliverables.

**Table 2. Tasks, Objectives, and Deliverables.**

Task	Objective	Deliverables
Task 1	<ul style="list-style-type: none"> <li>• Conduct outreach through stakeholder meetings to gather information that will identify greatest data needs and to prioritize data items for public-use release.</li> </ul>	<ul style="list-style-type: none"> <li>• An internal summary of the findings of the stakeholder’s meetings, including statistical aspects of the generation of synthetic data files, identification and definition of priority variables to include on synthetic record-level linked data files and key summary statistics to include on a public facing dashboard.</li> </ul>

Task	Objective	Deliverables
Task 2	<ul style="list-style-type: none"> <li>Establish a synthetic data creation methodology that can be applied to linked data resources. Once the methodology is established, create publicly available synthetic linked data files and a data enclave validation process to confirm analytic results with restricted-use data.</li> </ul>	<ul style="list-style-type: none"> <li>Up to four linked synthetic data files that can be accessed on the NCHS data linkage website and used for modeling and statistical inference. <ul style="list-style-type: none"> <li>Synthetic Linked 2018 NHIS-CMS-HUD Data and technical documentation describing the methodology used to create the synthetic is available on the NCHS website: <a href="https://www.cdc.gov/nchs/linked-data/synthetic/index.html">https://www.cdc.gov/nchs/linked-data/synthetic/index.html</a></li> <li>Synthetic Linked 2016 NHCS-NDI Data and technical documentation describing the methodology used to create the synthetic expected to be available on the NCHS website in 2026</li> <li>A verification process has been developed for synthetic linked data users to verify selected analytic results against analytic results with restricted-use data</li> <li>Verification metrics are returned for selected analyses</li> </ul> </li> </ul>
Task 3	<ul style="list-style-type: none"> <li>Create a publicly available analytic dashboard to broaden data accessibility and utility</li> </ul>	<ul style="list-style-type: none"> <li>The verification process includes a dashboard that is deployed within the NCHS computing environment. A comparison of univariate distributions for selected variables and verification metrics are returned to data users who submit requests to the NCHS Data Linkage Program.</li> </ul>
Task 4	<ul style="list-style-type: none"> <li>Engage in knowledge dissemination to inform greater research community on how to incorporate best practices and improve accessibility of linked data for evidence building</li> </ul>	<ul style="list-style-type: none"> <li>A journal manuscript describing the methodology for developing the synthetic linked data files and verification process has been submitted for agency clearance</li> <li>Conference presentations that included an overview or discussion of the project are listed in <a href="#">Presentations</a>.</li> </ul>

### **3. Major Accomplishments**

#### **3.1 Synthetic Data Generation Methodology**

This project resulted in two public-use synthetic linked data files that can be used to support public health and patient-centered outcomes research. Also, a scalable and replicable methodology for generating statistically valid synthetic linked data files was implemented. The NCHS Data Linkage Program continues to link NCHS survey data to a wide variety of PCOR-related administrative data sources including Medicare, Medicaid, HUD federal housing program data, and death certificate records. This provides the potential for new synthetic linked data files to be generated, using the methodology developed under this project.

#### **3.2 Verification Process**

The project also created a verification process to assist data users with analyses using the synthetic linked data and verify that the synthetic data are suitable for their research needs. While it isn't possible to support every type of analysis and analytic model with the NCHS synthetic linked data, the release of the verification service offers synthetic data users the opportunity to compare selected analytic results and identify instances where the restricted-use linked data better support their specific research objectives.

### **4. Lessons Learned**

Several critical lessons were learned from this project. The restricted-use linked data that is used to create synthetic files can contain a large number of records and variables. It may not be possible to synthesize all data elements and preserve the statistical properties or desired variable correlations from the original restricted data. It is important to engage with potential data users early in the process to identify a limited set of key variables and populations of interest.

There are different methods that can be used to generate synthetic data. It is important to identify a process that can account for the attributes of the original restricted data, such as the complex survey design for NHIS or the large sample size for NHCS.

The process of generating synthetic data is iterative. Analytic validity and disclosure risk must be assessed and taken into account throughout process. Results from interim utility and disclosure assessments may lead to modifications to synthetic data generation process and the need to repeat some or all of the steps. In some situations, it may be necessary to accept a reduction in analytic utility to preserve data privacy and prevent disclosure. With the generation on synthetic data, some uncertainty is introduced that can impact variance estimation. Generating and releasing multiple implicates can be effective for maintaining data utility by providing more appropriate variance estimates that reflect the additional uncertainty

added by the synthesis process. With the release of multiple implications, it is important to provide resources such as sample code to assist in conducting analyses.

## 5. Presentations and Publications

### 5.1 Presentations

- **May 19, 2021**, L. Mirel presented “Privacy Preserving Techniques: Case Studies from the Data Linkage Program” at the NCHS Board of Scientific Counselors Meeting.
- **May 26, 2022**, L. Mirel presented “Advancing the Data Modernization Initiative through Data Linkages” at the NCHS Board of Scientific Counselors Meeting.
- **January 9, 2023**, L. Mirel presented “Privacy Preserving Techniques: Case Studies from the NCHS Data Linkage Program” at the NISS IOF workshop: Advancing Demographic Equity with Privacy Preserving Methodologies.
- **August 4, 2023**, C. Golden presented “Privacy Preserving Techniques: Case Studies from the National Center for Health Statistics” at the Joint Statistical Meetings.
- **September 14, 2023**, C. Golden presented “Privacy Preserving Techniques: Synthetic Linked Data Files” at the NCHS Board of Scientific Counselors Meeting.
- **August 6, 2024**, C. Golden presented “National Center for Health Statistics Data Linkage Program: Synthetic Data to Support Tiered Access” at the Joint Statistical Meetings (JSM).
- **September 13, 2024**, C. Golden presented “National Center for Health Statistics Data Linkage Program: Generating Synthetic Data to Support Tiered Access” at the Privacy and Public Policy Conference.
- **September 17, 2024**, C. Golden presented “National Center for Health Statistics Data Linkage Program: Generating Synthetic Data to Support Tiered Access” at the International Population Data Linkage Network Conference.
- **June 9, 2025**, C. Golden participated in “Opening Panel – Success Stories from Other Domains” at the Summer Institute on Privacy Enhancing Technologies for Education Data hosted by the Massive Data Institute at Georgetown University.
- **March 12, 2026**, C. Golden presented “NCHS Data Linkage Program: Exploring Privacy Enhancing Technologies for Data Linkage” at the Federal Committee on Statistical Methodology Conference.

### 5.2 Publications

- Himschoot, A., Boyd, R., Cox C., Martin, C.B., Golden, C., Reiter, J.P. Creation and Results Verification for National Center for Health Statistics Synthetic Linked Data. *In Clearance*.

## 6. Future Considerations

The synthetic data generation methodology used for the project and new data resources developed support current Federal open data initiatives and aligns with tiered access data model as described in the Foundations for Evidence-Based Policymaking Act of 2018. NCHS will continue to compile user feedback to inform the potential development of additional synthetic linked data resources in the future and enhancements to the verification process.

## 7. Summary

This project provides two new sources of data for the PCOR community. The methodology used is scalable for different linked data sources and can potentially be applied to create additional synthetic linked data files. NCHS will continue to monitor and promote the use of the synthetic linked data and verification process. NCHS will apply the user feedback received to inform decisions on applying the methodology to potentially create new synthetic linked data resources in the future.

## 8. Accessing the NCHS Synthetic Linked Data and Requesting Verification Metrics

Information about NCHS Synthetic Linked Data is available on the NCHS Data Linkage website:

<https://www.cdc.gov/nchs/linked-data/synthetic/index.html>

Link to download the Synthetic Linked 2018 NHIS-HUD-CMS Data:

[https://ftp.cdc.gov/pub/Health\\_Statistics/NCHS/datalinkage/synthetic\\_data/NHIS\\_HUD\\_CMS/](https://ftp.cdc.gov/pub/Health_Statistics/NCHS/datalinkage/synthetic_data/NHIS_HUD_CMS/)

Data Dictionary for the Synthetic Linked 2018 NHIS-HUD-CMS Data:

<https://www.cdc.gov/nchs/data/datalinkage/synthetic-linked-2018-nhis-hud-cms-data-codebook.PDF>

Additional information about the Synthetic Linked 2018 NHIS-HUD-CMS Data, including an overview of the dataset and a description of the synthetic data generation process, and the synthetic linked data results verification process are available in the [Synthetic Linked 2018 NHIS-HUD-CMS Data User Guide](#).

To request that NCHS verify your synthetic data results, email a completed [Template for Requesting Verification Metrics for the Synthetic Linked NHIS-HUD-CMS Data](#) to NCHS Data Linkage Team ([datalinkage@cdc.gov](mailto:datalinkage@cdc.gov)).

## 9. References

1. National Center for Health Statistics. 2018 National Health Interview Survey (NHIS) Public Use Data Release Survey Description. Hyattsville, MD. 2019. Available from: [https://ftp.cdc.gov/pub/Health\\_Statistics/NCHS/Dataset\\_Documentation/NHIS/2018/srvydesc.pdf](https://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2018/srvydesc.pdf)
2. National Center for Health Statistics. A Primer on HUD Programs and Associated Administrative Data. Hyattsville, Maryland. 2019. Available from: <https://www.cdc.gov/nchs/data/datalinkage/primer-on-hud-programs.pdf>
3. The Linkage of the National Center for Health Statistics (NCHS) Survey Data to U.S. Department of Housing and Urban Development (HUD) Administrative Data: Linkage Methodology and Analytic Considerations. Hyattsville, Maryland. 2022 Available from: <https://www.cdc.gov/nchs/linked-data/hud/index.html>
4. Agency for Healthcare Research and Quality. Community-Level Health Database. Rockville, MD. 2023. Available from: <https://www.ahrq.gov/data/innovations/clh-data.html>
5. National Center for Health Statistics. 2016 National Hospital Care Survey (NHCS) Restricted Data. Hyattsville, MD. <https://www.cdc.gov/nchs/nhcs/data/index.html>
6. The Linkage of the 2016 National Hospital Care Survey to the 2016/2017 National Death Index: Methodology Overview and Analytic Considerations. Hyattsville, Maryland. 2019. Available from: <https://www.cdc.gov/nchs/data/datalinkage/NHCS16-NDI16-17-Methodology-Analytic-Consider.pdf>
7. White D.G., Adams N.B., Brown A.M., O’Jiaku-Okorie A., Badwe R., Shaikh S., Adegboye A. Enhancing identification of opioid-involved health outcomes using National Hospital Care Survey data. National Center for Health Statistics. Vital Health Stat 2(188). 2021. DOI: <https://dx.doi.org/10.15620/cdc:108998>
8. Karr, A.F., Kohnen, C.N., Oganian, A. Reiter, J.P., and Sanil, A.P. A Framework for Evaluating the Utility of Data Altered to Protect Confidentiality. The American Statistician. 2006. 60(3), 224-232. Available from: <https://www2.stat.duke.edu/~jerry/Papers/tas2006.pdf>