Office of the Secretary Patient-Centered Outcome Research Trust Fund Project HP-17-002 (IAA #750117PE090019)

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

Enhancing Data Resources for Studying Patterns and Correlates of Mortality in Patient-Centered Outcomes Research: Project 1 - Adding Cause-Specific Mortality to NCHS's National Hospital Care Survey by Linking to the National Death Index

FINAL REPORT

February 2021

Table of Contents

Team Members

Team Member Organizations	Team Members
Division of Health Care Statistics, National Center for Health Statistics, Centers for Disease Control and Prevention	Carol DeFrances- Project Co-Lead Geoff Jackson Donielle White
Division of Analysis and Epidemiology, National Center for Health Statistics, Centers for Disease Control and Prevention	Lisa Mirel- Project Co-Lead Cordell Golden Jennifer Sayers
NORC	Ed Mulrow- Project Manager Dean Resnick Scott Campbell Shawn Linman Chris Cox

Enhancing Data Resources for Studying Patterns and Correlates of Mortality in Patient-Centered Outcomes Research: Project 1 - Adding Cause-Specific Mortality to NCHS's National Hospital Care Survey by Linking to the National death index

FINAL REPORT

1. Executive Summary

The National Center for Health Statistics (NCHS) houses unique national data resources and expertise, including (1) the National Hospital Care Survey (NHCS), which provides statistics on health and health care utilization based on patient hospital inpatient and emergency department (ED) visits, through the collection of administrative claims records and electronic health records (EHR), as well as information on patient characteristics; (2) the National Death Index (NDI), a centralized database of death record information on file in state vital statistics offices; and (3) the NCHS Data Linkage Program, which has substantial statistical and methodological expertise in creating, managing and analyzing linked files.

This project linked NHCS to NDI data, which was a unique and important research endeavor. While other hospital data can assess inpatient or 30-day mortality, NHCS is the only federal health care establishment data system, and therefore the only hospital data, designed to include personally identifiable information (PII) (e.g., patient name, address, Social Security Number), which is necessary for linking hospital encounters to the NDI to obtain cause-specific mortality. Cause-specific mortality is a key patient-centered outcome because it allows deaths unrelated to care to be excluded. This study created data resources that allowed for new capabilities to generate national benchmarks of clinical outcomes such as death due to heart disease 30, 60, 90 days, and up to a year following hospital admission for heart attack, and even 1-year survival post discharge. Linkage of the NHCS and NDI allowed for comparison of inpatient to ED discharge outcomes (such as mortality) for specific causes of death. In addition, the NHCS data were linked to the Centers for Medicare & Medicaid Services (CMS) Master Beneficiary Summary File (MBSF) data to provide greater insight on post hospitalization care for Medicare beneficiaries.

2. Background

2.a Goal

This project aimed to link Uniform Bill (UB-04) administrative claims and EHR data from the NCHS' National Hospital Care Survey to death record information from the NDI and to administrative data from the CMS MBSF.

2.b Data Sources

National Hospital Care Survey
NHCS is an establishment survey that collects inpatient and ED, and outpatient department (OPD) visit level data from sampled hospitals. NHCS is one of the National Health Care

Surveys, a family of surveys covering a wide spectrum of health care delivery settings from ambulatory and outpatient to hospital and long-term care providers (https://www.cdc.gov/nchs/dhcs/index.htm). The goal of NHCS, when fully implemented, will be 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-use involved ED visits.

NHCS collects patient PII such as name, address, and Social Security Number (SSN), which allows for the linkage of episodes of care within a surveyed hospital as well as to other external data sources, such as the NDI. The linkage described in this report includes only patients with inpatient and ED visits – patients with only non-emergency department OPD visits were excluded from the linkages.

Eligible hospitals for NHCS are non-institutional, non-federal hospitals with six or more staffed inpatient beds. There are 6,622 hospitals which met these criteria as of 2013 to form the survey sampling frame. The hospital sample size for the 2014 and 2016 NHCS data collection (which re-use the 2013 sample) was 581 hospitals.

Data from all inpatient, ED and outpatient visits occurring during the calendar year are collected from NHCS participating hospitals. In 2014, 95 NHCS participating hospitals provided these data in the form of UB-04 administrative claims records.

For 2016, in an effort to reduce the burden of transferring their hospital records, participating hospitals were given the option to submit their patient records using the UB-04 claims data format or to submit extracts from their EHR systems. Hospitals submitting EHR records did so in the form of a custom extract or Consolidated Clinical Document Architecture (C-CDA), which are a set of Health Level Seven International (HL7) clinical document architecture specifications and include templates such as Continuity of Care Documents (CCDs) (1). In 2016, 158 hospitals participated in the NHCS. Of these hospitals, 89 hospitals provided UB-04 administrative claim records and 47 provided EHR records – 16 as custom extract and 31 as CCD. The remaining 22 hospitals provided records submitted in the format to Vizient, a healthcare performance company, which are similar to UB-04 data but contain no patient PII so were not used in linkage.

National Death Index (NDI)

The NDI is a centralized database of United States 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 (2). The NDI became operational in 1981 and includes death record information for persons dying in the U.S. or a U.S. territory from 1979 onward. The records, which are compiled annually into the longitudinal register, include detailed information on the underlying and multiple causes of death.

Centers for Medicare and Medicaid Services, Medicare Data

The CMS Medicare MBSF is an annual file containing demographic and enrollment information about beneficiaries enrolled in Medicare during each calendar year. The CMS Medicare MBSF includes information on beneficiary demographic characteristics, reason for Medicare entitlement, and program enrollment type (Original Medicare vs. Medicare Advantage (MA)).

The **Base (A/B) segment** includes beneficiary characteristics, monthly entitlement indicators, reasons for entitlement (initial and current), and monthly MA indicators. The **Part D segment** includes variables specific to Medicare Part D Prescription Drug plan enrollment. The **Cost & Utilization segment** includes summarized information about the service utilization and Medicare payment amounts by type of claim, including prescription drugs. The **Chronic Conditions segment** includes variables that indicate a Medicare beneficiary has received a service or treatment for selected chronic health conditions.

2.c Tasks, Objectives, and Deliverables

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

Table 1. Tasks, Objectives, and Deliverables

Task	Objective	Deliverables
Task 1	• Link 2014 NHCS inpatient and ED claims data to the 2014 and 2015 NDI • Create a written report on NHCS-NDI data linkage methodology and analytic considerations to be published on the web	• New data file containing linked 2014 NHCS inpatient and ED claims to the 2014 and 2015 NDI available to researchers through the NCHS and Federal Research Data Centers (RDC) • Report on NHCS-NDI data linkage methodology and analytic considerations published on the NCHS website. https://www.cdc.gov/nchs/data/datalinkage/NHCS14_NDI14_15_Methodology_Analytic_Consider.pdf

Task	Objective	Deliverables
Task 2	• Link 2014 NHCS inpatient and ED claims data to 2014 and 2015 CMS MBSF data. • Create a written report on NHCS-CMS data linkage methodology and analytic considerations to be published on the web.	• New data file containing linked 2014 NHCS inpatient and ED claims to the 2014 and 2015 CMS MBSF available to researchers through the NCHS and Federal RDCs. • Report on NHCS-CMS MBSF data linkage methodology and analytic considerations published on the NCHS website. https://www.cdc.gov/nchs/data/datalinkage/NHCS-CMS-Medicare-Llinkage-Methods-and-Analytic-Considerations.pdf
Task 3	• Conduct EHR methodological work. • Create a written report on EHR linkage process and evaluation of linkage algorithms.	• This is part of the Task 4 published report.
Task 4	• Link 2016 NHCS inpatient, ED claims, and EHR data to the 2016 and 2017 NDI • Write report on NHCS-NDI data linkage methodology containing summary statistics for the quality and quantity of final matches and detailing the algorithms for determining acceptable matches. • Write report on NHCS-NDI linkage.	• New data file containing linked 2016 NHCS inpatient and ED data to the 2016 and 2017 NDI available to researchers through the NCHS and Federal RDCs. • Report on NHCS-NDI data linkage methodology and analytic considerations published on the NCHS website. https://www.cdc.gov/nchs/data/datalinkage/NHCS16_NDI16_17_Methodology_Analytic_Consider.pdf
Task 5	• Conduct outreach and dissemination activities.	• Announcements of availability of linked data (i.e., website, conferences). • Citation list of research using linked NHCS-NDI and NHCS-CMS files. • Quarterly reports to ASPE. • Submitted final report to ASPE.

3. Major Accomplishments

3.a Enhanced Linkage Algorithm

The linkage methodology utilized in this project was based on both deterministic and probabilistic approaches. The probabilistic approach performs weighting and link adjudication following the Fellegi-Sunter method, which is an accepted standard approach used for record linkage (3). The three main steps taken to link NHCS data to the NDI are outlined below. A more detailed explanation can be found in the published methods and analytic guidelines reports (6,7,8).

1. Deterministic linkage performs joins on exact SSN– - and is validated by comparison of other identifying fields: if these criteria are met, these records are assigned a probability of being a valid match (match probability) of **1.00**.
2. Probabilistic linkage identifies likely matches, or links, between all records. Records are linked and scored as follows (note: SSN is excluded from the analysis for this step):
   a. Identify possible matched pairs
   b. Score potential match pairs—matches are scored based on the concurrence of these variables: First Name, Middle Initial, Last Name (or Father's Surname), State of Residence, Year of Birth, Month of Birth, Day of Birth, Date of Death (if available on hospital record), Sex
   c. Probability modeling – estimate match probability.
3. Determine a match probability cutoff. For each patient record, keep the linked NDI record having the highest estimated match probability as long as it is above the linkage cutoff.

The linkage algorithms developed for this project were created with customized code (using SAS 9.4) and was tailored to perform these specific linkages, to produce high-quality matches with a low degree of linkage error.

3.b Linkage Error Estimation: NDI

Given that the deterministic links were considered truth, it was possible to estimate the Type I and Type II errors of the linkage process for all of these projects. To do this, Type I error (false-positive links) rates were computed based on SSN agreement rates among linked records with valid SSNs in both sources being combined.

To measure Type II error (false-negative links), the truth deck that was developed in the deterministic linkage was used. It was expected that this truth deck had only a few exceptional pairs that were not true matches. For the probabilistic records, Type II error was estimated as the percentage of the truth deck records that were not returned as links by the probabilistic method.

Tables 2 and 3 show the Type I and II errors for each of the linkages in this project.

Table 1. Estimated Linkage Error for the 2014 NHCS linkage to the 2014-2015 NDI

Record Source	Cutoff	Eligible NHCS Patients	Total Links	Deterministic Matches	Non-Deterministic Links	Est Incorrect (Type I)	Est Not Found (Type II)
UB-04 Claims	See Notes	3,251,186	167,678 (5.2%)	109,141 (65.1%)	58,537 (34.9%)	0.26%	1.1%

NOTES: Data are presented at the patient level. Data in the table represent the combined error for inpatient and ED records. The probability cutoff used for inpatient records was 0.9 and 0.925 for ED records. Est, estimated.
SOURCE: 2014 NHCS – 2014-2015 NDI

Table 2. Estimated Linkage Error for the 2016 NHCS linkage to the 2016-2017 NDI

Record Source	Cutoff	Eligible NHCS Patients	Total Links	Deterministic Matches	Non-Deterministic Links	Est Incorrect (Type I)	Est Not Found (Type II)
All Sources	0.925	4,181,406	213,698 (5.1%)	125,437 (58.7%)	88,261 (41.3%)	0.24%	1.14%

*Unable to estimate Type II linkage error for CCD records due to no SSN information on CCD records.
NOTES: Data are presented at the patient level. The estimated Basal Error Adjustment made to the type I and II error rates was 0.0035. Est, estimated.
SOURCE: 2016 NHCS – 2016-2017 NDI

3.c Linkage Error Estimation: CMS

The estimation of Type I and Type II errors for the CMS linkage followed the methods described above and are noted in Table 4.

Table 4. Linked 2014 NHCS – 2014-2015 CMS Medicare MBSF

Record Source	Cutoff	Eligible NHCS Patients	Total Links	Deterministic Matches	Non-Deterministic Links	Est Incorrect (Type I)	Est Not Found (Type II)
UB-04 Claims	0.85	3,235,759	865, 125 (26.7%)	630,779 (72.9%)	234,346 (27.1%)	0.23%	0.32%

SOURCE: 2014 NHCS – 2014-2015 CMS Medicare MBSF. Est, estimated.

3.d Machine Learning as Part of the Linkage Algorithm

Blocking is a key step in record linkage. It identifies potential candidate pairs without comparing every single pair in the Cartesian product. Blocking or indexing, "splits each database into smaller blocks according to some blocking criteria (generally known as a blocking key)" (4). Rules can be used to define the blocking criteria however, for the CMS linkage, instead of rules, machine learning techniques were used to help create a set of blocks that would efficiently join the datasets together. By using the data to create the efficient block set, the number of false positive links were reduced while retaining a high percentage of true positive links. For the purpose of this linkage, the 'truth deck' was used as the training dataset. When the data are used in this manner, it is commonly referred to as a machine learning algorithm. For more detailed information on the method that was used please refer to "Learning Blocking Schemes for Record Linkage" (5) and the published methods and analytic guidelines (5, 8).

4. Lesson Learned

Several critical lessons were learned from this project. This project highlighted the importance of linking data sources to answer key policy and patient-centered research questions looking at 30, 60, 90-day post-acute hospital mortality for all causes and cause specific analyses. The linkage from this project highlighted the importance of collecting patient PII data in the NHCS and the resulting data files have been featured as important recruitment tool to encourage hospital participation. It also supported enhancements to linkage algorithms that will improve efficiency and accuracy for future linkage projects. This work aligns with the Evidence Act of 2018 and HHS Federal Data Strategy. The 2014 and 2016 NHCS UB-04 administrative claims data, and EHR data were successfully linked to the NDI and CMS MBSF to produce linked files that are available to researchers through the NCHS and Federal RDCs. NCHS is accepting requests to access the linked files, and the data have been analyzed by internal NCHS staff.

5. Publications and Presentations

5.a Presentations:

- **March 9, 2018** - Jennifer Sayers presented at the Federal Committee on Statistical Methodology Research Conference in Washington, DC titled "Data Linkage with an Establishment Survey". Co-authors: Scott Campbell (NORC), Clint Thompson, Geoff Jackson
- **June 6, 2018 -** NCHS staff presented at the HHS Evaluation and Administrative Data Day in Washington, DC:
    - Carol DeFrances - "An Overview of the National Hospital Care Survey (NHCS) and NHCS-based Patient-Centered Outcomes Research Trust Fund (PCORTF) Projects"
    - Lisa Mirel - "Linking the National Hospital Care Survey Data to the National Death Index and CMS Medicare Administrative Data"
    - Ernest Moy - "Health Outcomes Research Using Linked Hospital and Death Data"

- **June 19, 2018** – Carol DeFrances and Margaret Noonan presented "Using the National Hospital Care Survey (NHCS) to identify opioid-related hospital visits" at the National Drug Early Warning System Network in College Park, MD. https://www.cdc.gov/nchs/data/nhcs/NDEWS_webinar.pdf
- **June 25, 2018 -** NCHS staff presented at the Academy Health Annual Research Meeting in Seattle, WA:
  - Irma Arispe - "Meeting Data Needs for Patient-Centered Outcomes Research through Data Linkage"
  - Lisa Mirel – "Overview of the National Center for Health Statistics' Data Linkage Program"
  - Jennifer Sayers – "National Center for Health Statistics Linked Data Files: Accessing, Analyzing, and Presenting Results"
- **September 13, 2018 –** Lisa Mirel presented "Quality of Linked Data: Linking the National Hospital Care Survey Data to the National Death Index" at the International Population Data Linkage Network Conference in Banff, Alberta, Canada. Co-authors Dean Resnick, Scott Campbell, and Cordell Golden
- **October 19, 2018 –** Jennifer Sayers presented "Augmenting Hospital Claims Records through Data Linkage" at the Women in Statistics and Data Science Conference in Cincinnati, OH. Co-authors: Kathleen Taylor, Anita Bercovitz, Lisa Mirel
- **November 8, 2018 –** Dean Resnick presented "Evaluation of the Accuracy of Links or Candidate Pairs within Record Linkage Analyses" at the Statistics Canada 2018 International Methodology Symposium in Ottawa, Ontario, Canada. Co-author Lisa Mirel
- **November 13, 2018** – NCHS staff presented at the Administrative Data Research Facilities Network Conference in Washington, DC
  - Marc Roemer presented "An Assessment of Using Frequency Weights for Record Linkage." Co-author Scott Campbell
  - Dean Resnick presented "Simulation Approach to Assess the Precision of Estimates Derived from Linking Survey and Administrative Records." Co-author Lisa Mirel
- **November 28, 2018** – Lisa Mirel presented "Linking New Data Sources: NCHS Data Linkage Program" at the Statistics Canada and NCHS Interchange meeting in Ottawa, Ontario, Canada.
- **December 12, 2018** - Dean Resnick (NORC) presented on the "Evaluation of the Accuracy of Links or Candidate Pairs within Record Linkage Analyses" at the Data Analysis Group (DAG) seminar. Co-author: Lisa Mirel.
- **December 17, 2018** – NCHS staff were invited speakers at ASPE's Addressing the Opioid Epidemic: Harnessing the Power of Data for Patient–Centered Research workshop at HHS in Washington, DC:
  - Lisa Mirel presented "Augmenting the National Hospital Care Survey Data through Linkages with Administrative Records"
  - Carol DeFrances presented "Enhancing Identification of Opioid-Involved Health Outcomes Using Linked Hospital Care and Mortality Data"

- **May 31, 2019** – Scott Campbell (NORC) presented on the "Overcoming Big Data: Linking the 2014 National Hospital Care Survey to the 2014/2015 Medicare CMS Master Beneficiary Summary File" at the 2019 Symposium on Data Science and Statistics. Co-author: Lisa Mirel.
- **July 8, 2019** – Carol DeFrances presented "Enhancing Data Resources for Studying Patterns and Correlates of Mortality in Patient-Centered Outcomes Research: Project 1 – Adding Cause-Specific Mortality to NCHS' National Hospital Care Survey by Linking to the National Death Index (IAA #750117PE090019)" at the Patient-Centered Outcome Research Summer Webinar Series.
- **July 18, 2019** – Lisa Mirel and Carol DeFrances presented "Future of NHCS Data Linkages" at the NCHS Board of Scientific Counselors, Drug Work Group meeting in Hyattsville, MD.
- **September 6, 2019** – NCHS staff presented an overview of the project, "Enhancing Data Resources for Studying Patterns and Correlates of Mortality in Patient-Centered Outcomes Research: Project 1 – Adding Cause-Specific Mortality to NCHS' National Hospital Care Survey by Linking to the National Death Index", at the National Center for Health Statistics Board of Scientific Counselors meeting in Hyattsville, MD.
- **October 24, 2019** – Carol DeFrances, Amy Brown, and Geoff Jackson conducted a seminar at the University of Kentucky Public Health program on the NHCS data linked to the NDI data available in the NCHS and Federal RDC.
  https://www.cdc.gov/nchs/data/nhcs/RDC_Request-508.pdf
- **January 6-8, 2020** – NCHS staff presented at the International Conference on Health Policy Statistics in San Diego, California.
  - Mirel, Lisa, "Leveraging Linked Data for Evidence Based Policymaking",
  - DeFrances, Carol, "Identification of Opioid Involved Health Outcomes Using Linked Hospital Care and Mortality Data" https://ww2.amstat.org/meetings/ichps/2020/onlineprogram/AbstractDetails.cfm?AbstractID=306607
  - Jackson, Geoff, "Mortality for Women within One Year after Delivery in the National Hospital Care Survey, 2016" https://www.cdc.gov/nchs/data/nhcs/mortality_women-508.pdf
  - Resnick, Dean (NORC), Mirel, Lisa, "Using Synthetic Data to Replace Linkage Derived Elements: A Case Study"
- **August 4, 2020** – NCHS staff presented at the 2020 Joint Statistical Meeting.
  - Golden, Cordell, "Using Surveys and Administrative Data to Assess the Quality of Linked Data: National Center for Health Statistics' Data Linkage Program" https://ww2.amstat.org/meetings/jsm/2020/onlineprogram/AbstractDetails.cfm?abstractid=312744
  - Jackson, Geoff, "Assessing National Hospital Care Survey and National Ambulatory Medical Care Survey Data: A Comparison of Opioid and Respiratory Disease Encounters"
  - Resnick, Dean (NORC), "Adjusting Record Linkage Match Weights to Partial Levels of String Agreement"

- **October 20, 2020** – Lisa Mirel presented a NCHS Webinar titled "Overview of NCHS' Data Linkage Program: Leveraging National Data for Evidence-based Decision Making." (https://www.cdc.gov/nchs/data-linkage/datalinkage-webinar.htm)
- **January 12, 2021** – Carol DeFrances and Geoff Jackson presented "Overview of the National Hospital Care Survey and FY18-FY19 Patient-Centered Outcomes Research Trust Fund Projects" at the ASPE Provisional Drug Overdose Monthly Briefing.

5.b Written publications:

- Ashman JJ, Cairns C, DeFrances CJ, Schwartzman A. Respiratory illness emergency department visits in the National Hospital Care Survey and the National Hospital Ambulatory Medical Care Survey. National Health Statistics Reports; no 151. Hyattsville, MD: National Center for Health Statistics. 2021. https://www.cdc.gov/nchs/data/nhsr/nhsr151-508.pdf
- Bercovitz A, Jamoom E, Lau DT. National Hospital Care Survey demonstration projects: Characteristics of inpatient and emergency department encounters among patients with any listed diagnosis of Alzheimer disease. National Health Statistics Reports; no 121. Hyattsville, MD: National Center for Health Statistics. 2018. https://www.cdc.gov/nchs/data/nhsr/nhsr121-508.pdf
- Hoyert DL, Miniño AM. Maternal mortality in the United States: Changes in coding, publication, and data release, 2018. National Vital Statistics Reports; vol 69 no 2. Hyattsville, MD: National Center for Health Statistics. 2020. https://www.cdc.gov/nchs/data/nvsr/nvsr69/nvsr69-02-508.pdf
- Jackson G, Chari K. National Hospital Care Survey demonstration projects: Stroke inpatient hospitalizations. National Health Statistics Reports; no 132. Hyattsville, MD: National Center for Health Statistics. 2019. https://www.cdc.gov/nchs/data/nhsr/NHSR132-508.pdf
- Jackson G, Brown AM, DeFrances C. Opioid-involved emergency department visits in the National Hospital Care Survey and the National Hospital Ambulatory Medical Care Survey. National Health Statistics Reports; no 149. Hyattsville, MD: National Center for Health Statistics. 2020. https://www.cdc.gov/nchs/data/nhsr/nhsr149-508.pdf
- Prell, Mark, Chapman C, Adeshiyan S, Fixler D, Garin T, Mirel L, and Phipps P. 2019. Transparent Reporting for Integrated Data Quality: Practices of Seven Federal Statistical Agencies. FCSM 19-01. Federal Committee on Statistical Methodology. September 2019. https://nces.ed.gov/FCSM/pdf/Transparent_Reporting_FCSM_19.01.pdf
- Resnick DM, Mirel LB, Roemer MI, Campbell SR. 2020. Adjusting Match Weights to Partial Levels of String Agreement in Data Linkage. In JSM Proceedings Government Statistics Section. Alexandria, VA: American Statistical Association. 994-1003.
- Spencer MR, Flagg LA, Jackson G, DeFrances C, Hedegaard H. National Hospital Care Survey demonstration projects: Opioid-involved emergency department visits, hospitalizations, and deaths. National Health Statistics Reports; no 141. Hyattsville, MD:

National Center for Health Statistics. 2020. https://www.cdc.gov/nchs/data/nhsr/nhsr141-508.pdf

- Williams S, Gousen S, DeFrances C. National Hospital Care Survey demonstration projects: Pneumonia inpatient hospitalizations and emergency department visits. National Health Statistics Reports; no 116. Hyattsville, MD: National Center for Health Statistics. 2018. https://www.cdc.gov/nchs/data/nhsr/nhsr116.pdf

6. Future Considerations

This project opened the door for many other opportunities, including linking the NHCS to additional administrative data sources, including the 2016/2017 CMS Medicare MBSF and Medicare fee-for service claims and encounter data, Housing and Urban Development administrative data, and Medicaid T-MSIS standard analytic files. The work from these FY 17 funded projects has also offered the opportunity to compare linkage results obtained using a commercial product that enables privacy preserving record linkage (PPRL) with the results obtained from these PCORTF supported NHCS-NDI data linkages, The results of the PPRL project may allow for linkages of new sources of data across the health care spectrum.

The linked NHCS-NDI files created from this project were the foundation for additional PCORTF projects, (1) FY18 "Enhancing Identification of Opioid-Involved Health Outcomes Using Linked Hospital Care and Mortality Data" and (2) FY19 "Identifying Co-Occurring Disorders among Opioid Users Using Linked Hospital Care and Mortality Data: Capstone to an Existing FY18 PCORTF Project". The FY18 PCORTF project provides greater context on overdose deaths by adding information on specific drugs mentions from the NCHS Drug-Involved Mortality (DIM) file to the linked NHCS-NDI records. Additionally, the FY18 PCORTF project developed an enhanced method for identifying opioid-involved and opioid overdose information in hospital visit and mortality data. The FY19 PCORTF project supplemented the linked 2016 NHCS – 2016-2017 NDI files with information on opioid-involved hospital visits with co-occurring substance use and mental disorders.

7. Summary

This project accomplished comprehensive data linkages to provide new sources of data for the PCOR community. The methods developed and subsequent improvements in linked data quality through this project will be utilized in future NCHS data linkage projects. NCHS will continue to monitor and promote the use of the linked NHCS data. As more years of the NHCS become available, the algorithms developed through this project will continued to be used to link new sources of data and create new resources.

8. How to request linked NHCS data

The 2014 NHCS – 2014/2015 NDI, 2014 NHCS – 2014/2015 CMS MBSF, and 2016 NHCS – 2016/2017 NDI linked files are made available through the NCHS Research Data Center (RDC) or Federal Statistical Research Data Centers (FSRDC). Researchers must submit a written proposal that will be reviewed by NCHS staff. For more information on RDC access, please see this link: https://www.cdc.gov/rdc/. Questions related to the linked files can be directed to the NCHS Data Linkage Team (datalinkage@cdc.gov).

9. References

1. National Center for Health Statistics. National Hospital Care Survey Research Data Center Documentation. Hyattsville, MD. 2019. Available from: https://www.cdc.gov/rdc/data/b1/NHCS_RDC_USERS_GUIDE.pdf
2. National Center for Health Statistics. National Death Index. Hyattsville, MD. 2017. Available from: https://www.cdc.gov/nchs/ndi/index.htm.
3. Fellegi IP, Sunter AB. A Theory for Record Linkage. 1969. JASA 64(328):1183–210.
4. Christen, Peter. Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection. Data-Centric Systems and Applications. Berlin Heidelberg: Springer-Verlag, 2012.
5. Michelson, Matthew, and Craig A. Knoblock. "Learning Blocking Schemes for Record Linkage." In Proceedings of the 21st National Conference on Artificial Intelligence - Volume 1, 440–445. AAAI'06. Boston, Massachusetts: AAAI Press, 2006
6. National Center for Health Statistics. Division of Analysis and Epidemiology. *The Linkage of the 2014 National Hospital Care Survey to the 2014/2015 National Death Index: Methodology Overview and Analytic Considerations*, August 2019. Hyattsville, Maryland. Available at the following address: https://www.cdc.gov/nchs/data-linkage/index.htm
7. National Center for Health Statistics. Division of Analysis and Epidemiology. *The Linkage of the 2016 National Hospital Care Survey to the 2016/2017 National Death Index: Methodology Overview and Analytic Considerations*, August 2019. Hyattsville, Maryland. Available at the following address: https://www.cdc.gov/nchs/data-linkage/index.htm
8. National Center for Health Statistics. Division of Analysis and Epidemiology. *The Linkage of the 2014 National Hospital Care Survey to the 2014/2015 Centers for Medicare & Medicaid Services Master Beneficiary Summary File: Methodology Overview and Analytic Considerations*, August 2019. Hyattsville, Maryland. Available at the following address: https://www.cdc.gov/nchs/data-linkage/index.htm