

**Office of the Secretary Patient-Centered Outcome Research Trust Fund Project HP-19-014
(IAA #750119PE090038)**

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

***Augmenting the National Hospital Care Survey (NHCS) Data through Linkages
with Administrative Records: Linking the National Hospital Care Survey to Centers for
Medicare & Medicaid Services and Housing and Urban Development Data***

FINAL REPORT

April 2022

Table of Contents

Team Members	3
1. Executive Summary	4
2. Background	4
3. Major Accomplishments	9
4. Lessons Learned	13
5. Publications and Presentations	13
6. Future Considerations	14
7. Summary	15
8. How to Request Linked NHCS Data	15
9. References	15

Team Members

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

***Augmenting the National Hospital Care Survey (NHCS) Data through Linkages
with Administrative Records: Linking the National Hospital Care Survey to Centers for
Medicare & Medicaid Services and Housing and Urban Development Data***

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; and (2) the NCHS Data Linkage Program, which has substantial statistical and methodological expertise in creating, managing, and analyzing linked files. Furthermore, NCHS has agreements with the Centers for Medicare & Medicaid Services (CMS) and the Department of Housing and Urban Development (HUD) to link survey data with administrative records from these two organizations.

This project linked data from the NHCS to CMS and HUD administrative data, which was a unique and important research endeavor. Although various organizations compile and disseminate data collected by hospitals, few integrate subsequent health care coverage information or contextual variables, like receipt of federal housing assistance. Through data linkage, the NCHS Data Linkage Program creates rich new data resources that can be used to study the relationship between health insurance and health outcomes, outpatient care and inpatient care, and participation in federal housing programs and management of chronic health conditions. The enhanced data linkage algorithms that were adopted in this project are critical additions to the portfolio of methods used by the NCHS Data Linkage Program.

2. Background

2.a Goal

This project aimed to link inpatient and ED claims and EHR data from the NCHS' National Hospital Care Survey (NHCS) to administrative data from CMS and housing data from HUD.

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 the 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 personally identifiable information (PII) such as name, date of birth, and Social Security Number (SSN), which allows for the linkage of episodes of care within a surveyed hospital as well as linkages to other external data sources, such as the CMS and HUD administrative data. The linkage described in this report includes inpatient and ED claims, as well as EHR data.

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.

Centers for Medicare and Medicaid Services, Medicare Data

Medicare is the primary federal health insurance program for people aged 65 or older, people under age 65 with qualifying disabilities, and people of all ages with End Stage Renal Disease (ESRD). Nearly two-thirds of persons enrolled in Medicare, known as Medicare beneficiaries, are enrolled in traditional Medicare, also known as Medicare fee-for-service (FFS). Nearly all Medicare FFS beneficiaries receive Part A hospital insurance benefits, which help cover IP hospital care, Skilled Nursing Facility (SNF) stays (not custodial or long-term care), home health care, and hospice care. Most FFS beneficiaries also enroll in Medicare Part B medical insurance benefits, which help to cover physician services, OP care, durable medical equipment (DME), and some home health care services. Approximately one-third of Medicare beneficiaries receive Medicare benefits through a Medicare Advantage (MA) plan, also known as Medicare Part C. MA plans are administered by approved Medicare Advantage Organizations (MAOs). MAOs sponsor privately managed care plans such as Health Maintenance Organization (HMOs), Preferred Provider Organization (PPOs), and Special Needs Plans (SNPs) which provide, at a

minimum, the same covered services provided in Medicare Parts A and B. MAOs may also elect to provide additional services not covered by Medicare Parts A and B such as dental and vision care. MAOs are responsible for providing Medicare benefits directly to enrollees through prior arrangements with providers or by paying for the benefits on behalf of enrollees. In 2006, Medicare beneficiaries could begin to elect optional prescription drug coverage, known as Medicare Part D. Part D coverage can be obtained through Medicare approved Part D private plans, known as Prescription Drug Plans (PDPs) or through Medicare Advantage Prescription Drug Plans (MA-PDs). MA-PDs provide prescription drug coverage that is integrated with the health care coverage provided to Medicare beneficiaries enrolled in MA plans.

The CMS Medicare Data Files are comprised of Standard Analytic Files, or SAFs, containing standard format extracts of research-oriented Medicare program data. The CMS Medicare Data Files contain information on the enrollment status, health care utilization, and expenditures of Medicare-enrolled beneficiaries. The SAFs for Medicare beneficiaries enrolled in FFS Medicare contain final action health care claims submitted for payment by both institutional and noninstitutional health care providers. A final action claim contains all payment adjustments between Medicare and providers and represents Medicare's final payment action for a given health care claim. Medicare FFS SAFs are organized by seven health care settings: IP, SNF, institutional outpatient (OP), practitioner/provider services (Carrier), home health agency (HHA), DME, and hospice care. The SAFs for MA-enrolled beneficiaries contain all health care encounter records submitted by MAOs for the given calendar year for each enrolled Medicare beneficiary. MA SAFs are organized by six health care settings: IP, SNFs, OP, Carrier, HHA, and DME. Hospice care services provided to Medicare beneficiaries enrolled in MA are paid under Medicare FFS rather than as part of the managed care plan.

The Medicare Part D Prescription Drug Event (PDE) File contains a summary of prescription drug costs and payment data used by CMS to administer benefits for all Medicare Part D enrollees including beneficiaries enrolled in both Medicare PDPs and MA-PDs. In addition to the SAFs and the PDE Files, two assessments are also included in the linked dataset – the Home Health Outcome and Assessment Information Set (OASIS) and the Long-Term Care Minimum Data Set (MDS). The OASIS assessment contains data pertaining to patient outcomes and home health care. The OASIS assessments are required of all HHAs certified to accept Medicare and Medicaid payments. The MDS is a health status screening and assessment tool used for all residents of long-term care nursing facilities certified to participate in Medicare or Medicaid, regardless of payer. The MDS assessment is also required for Medicare payment of SNF stays.

U.S. Department of Housing and Urban Development, Housing Assistance Data

The U.S. Department of Housing and Urban Development (HUD) is the primary federal agency responsible for overseeing domestic housing programs and policies. While HUD is responsible for administering various housing and community development programs, the linkage with the 2014 and 2016 NHCS focuses on HUD's three largest housing assistance programs: Housing Choice Vouchers (HCV), Public Housing (PH), and Multifamily programs (MF). Persons and households participating in these program types are "HUD-assisted." People living in HUD-assisted households 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. Generally, rental subsidies seek to reduce gross housing costs for the tenant to approximately 30% of household income, although program rules may allow for variations in that ratio. A HUD subsidy pays the remaining amount up to a specified limit that varies by program.

The HUD Housing Choice Voucher (HCV) program is the federal government's largest housing assistance program, allowing low-income families, elderly persons, and persons with disabilities to choose and lease safe and affordable housing. In the HCV program, housing assistance is tenant-based, meaning that participants find their own housing in the private market. Participants are free to choose any housing that meets program requirements and are not limited to units located in subsidized housing projects. In the NHCS-HUD linked data, the HCV program also includes the Homeownership Voucher, Project-Based Voucher, Section 8 Moderate Rehabilitation, and Section 8 Rental Certificate programs. Among 2014 and 2016 NHCS patients that linked to HUD administrative data, just over 50% were participating in an HCV program.

The multifamily (MF) program category in the linked NCHS-HUD data encompasses a number of separate, distinct HUD programs, including: Project-Based Section 8 (or PBS8) Voucher Assistance in Multifamily Housing (the largest MF program), Section 221(d)(3) Below Market Interest Rate, Section 236 Multifamily Housing, Rental Assistance, Section 202 Supportive Housing for the Elderly Program, Section 202/162—Project Assistance Contract, Section 811 Supportive Housing for Persons with Disabilities, and Rent Supplement. Because each of the remaining MF programs lacked sufficient sample size on an individual basis in the linked file, they were combined into a single MF program category. In all MF programs, subsidies are paid directly to private property owners who provide a certain percentage of their housing units at affordable rates for low-income persons who qualify. MF program assistance is tied to the property, unlike tenant-based rental assistance programs (e.g., HCVs), and tenants cannot take their rental housing assistance subsidy elsewhere. Approximately 25% of the 2014 and 2016 NHCS patients that linked to HUD were participating in a MF program.

The Public Housing (PH) program was established to provide safe rental housing for eligible low-income families, the elderly, and persons with disabilities. HUD provides capital subsidies and operating subsidies to local Public Housing Agencies (PHAs) that manage public housing for eligible low-income residents. HUD also provides technical assistance to help PHAs plan, develop, and manage PH developments. Approximately 25% of the 2014 and 2016 NHCS patients that linked to HUD were participating in a PH program.

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	<ul style="list-style-type: none"> • Link 2016 inpatient and ED claims and EHR data to 2016-2017 CMS data • Create a written report on NHCS-CMS data linkage methodology and analytic considerations to be published on the web 	<ul style="list-style-type: none"> • New data file containing linked 2016 NHCS inpatient and ED claims and EHR to 2016-2017 CMS data available to researchers through the NCHS and Federal Research Data Centers (RDC) https://www.cdc.gov/nchs/data-linkage/CMS-Medicare-Restricted.htm • Report on NHCS-CMS data linkage methodology and analytic considerations published on the NCHS website. https://www.cdc.gov/nchs/data/datalinkage/NHCS-CMS-Medicare-Linkage-Methods-and-Analytic-Considerations.pdf
Task 2	<ul style="list-style-type: none"> • Research and Evaluation of the CMS Linked Data • Create NCHS publications and peer-reviewed articles detailing innovative aspects of the NHCS-CMS linkage methods and assessments of linkage quality 	<ul style="list-style-type: none"> • Synthetic Data Journal Paper (https://link.springer.com/article/10.1007/s10742-021-00241-z) • Machine Learning Journal Paper (https://content.iospress.com/articles/statistical-journal-of-the-iaos/sji200779) • Citation list: https://www.cdc.gov/nchs/data/datalinkage/LinkedNHCSDataCitationList_508.pdf
Task 3	<ul style="list-style-type: none"> • Link the 2014 and 2016 inpatient & ED claims and EHR data (2016) to HUD data • Create a written report on NHCS-HUD data linkage methodology and analytic considerations to be published on the web 	<ul style="list-style-type: none"> • New data file containing linked 2014 and 2016 NHCS inpatient and ED claims and EHR to HUD data available to researchers through the NCHS and Federal Research Data Centers (RDC) https://www.cdc.gov/nchs/data-linkage/nhcs-hud.htm • Report on NHCS-HUD data linkage methodology and analytic considerations published on the NCHS website. https://www.cdc.gov/nchs/data/datalinkage/NHCS-HUD-Linkage-Methods-and-Analytic-Considerations.pdf

Task	Objective	Deliverables
Task 4	<ul style="list-style-type: none"> Research and evaluation of the linked HUD linked files 	<ul style="list-style-type: none"> HUD linkage methodology comparison (see Appendix II) https://www.cdc.gov/nchs/data/datalinkage/NCHS-HUD-Linked-Data-Methodology-and-Analytic-Considerations.pdf

3. Major Accomplishments

3a. Enhanced Linkage Algorithm

The linkage methodology utilized in this project includes technical enhancements not included in previous linkages and a robust analysis of linkage accuracy. The enhancement involved a new sequential coverage algorithm (SCA). The SCA is a machine learning technique that implements and tests a rule, evaluates its accuracy, then removes it from the linkage routine and tests a new rule. This approach was compared to a previously used linkage method that did not use the SCA. A comparison of Type I and Type II error rates determined that the SCA method was the most accurate.

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” (2) and the published methods and analytic guidelines (2, 3).

The new linkage also included an SSN odds adjustment to further improve the accuracy of the probabilistic matching within the linkage. Furthermore, the logistic regression step in the linkage algorithm was replaced with a partial expected-maximization (EM) model, which further improved the accuracy of the linkage. The linkage algorithms developed for this project were created and customized using SAS statistical software and saved for future use by the NCHS Data Linkage Program.

3b. Final Linkage Rate Tables

The tables below illustrate the linkage rates by demographic variables for the linked 2016 NHCS to 2016/2017 CMS Medicare administrative records and the linked 2014 and 2016 NHCS to HUD administrative records.

Table 2. Linked 2016 NHCS – 2016/2017 CMS Medicare Administrative Records - Sample Sizes and Percent Linked, by Age

	Sample Size			Percent Linked	
	Total Sample	Eligible for Linkage ²	Linked to 2016-2017 Medicare Administrative Data ³	Total Sample ⁴	Eligible Sample ⁵
Age¹					
<65	3,692,926	3,459,122	294,388	8.0	8.5
>=65	762,766	717,624	699,734	91.7	97.5
Total	4,455,692	4,176,746	994,122	22.3	23.8

NOTES: Data are presented at patient level. It is possible that NHCS patients had more than one date of birth. When more than one date of birth was present, the minimum of the non-missing DOB was selected for the patient.

¹Age is as of final encounter (date of last known contact). Age could not be determined for 24,121 patients based on availability of date of birth and age could not be determined for an additional 1,343,352 patients due to patient records missing PII.

²Eligibility for linkage is based upon having sufficient PII in at least two of three data element groups: SSN/HICN, name, and date of birth.

³This group includes linkage-eligible patients who linked to Medicare administrative records at any time during the linkage interval (2016 - 2017).

⁴This percentage is calculated by dividing the number of linked patients by the number of patients in the total sample.

⁵This percentage is calculated by dividing the number of linked patients by the total number of linkage-eligible patients.

Table 3. Linked NHCS – HUD Administrative Records: Sample Sizes and Percent Linked, by Age and Sex

	Sample Size		Percent Linked		
	Total Sample	Eligible for Linkage ³	Linked to HUD Data ⁴	Total Sample ⁵	Eligible Sample ⁶
2014 NHCS					
Age¹					
0-17	1,063,289	961,790	100,939	9.5	10.5
18-44	1,155,989	1,050,841	86,242	7.5	8.2
45-61	630,731	574,740	40,557	6.4	7.1
62 and over	707,187	640,714	41,448	5.9	6.5
Total	3,557,196	3,228,085	269,186	7.6	8.3
Sex²					
Male	1,577,255	1,434,577	92,104	5.8	6.4
Female	1,978,165	1,791,197	176,941	8.9	9.9
Total	3,555,420	3,225,774	269,045	7.6	8.3
2016 NHCS					
Age¹					
0-17	1,293,458	1,205,473	122,502	9.5	10.2
18-44	1,477,611	1,386,926	112,874	7.6	8.1
45-61	796,022	748,333	54,309	6.8	7.3
62 and over	888,601	836,014	56,013	6.3	6.7
Total	4,455,692	4,176,746	345,698	7.8	8.3
Sex²					
Male	2,597,453	1,851,201	116,174	4.5	6.3
Female	3,157,461	2,278,263	225,141	7.1	9.9
Total	5,754,914	4,129,464	341,315	5.9	8.3

NOTES: Data are presented at patient level.

¹ Age is as of final IP or ED encounter (date of last known contact). Age could not be determined for 1,090 patients in the 2014 NHCS and for 1,367,473 patients in the 2016 NHCS due to missing data. Age is calculated by subtracting patient date of birth (DOB) from the final encounter date. When more than one DOB was present, the minimum of the non-missing DOB was selected.

² Sex could not be determined for 2,866 patients in the 2014 NHCS and for 68,251 in the 2016 NHCS due to missing data.

³ Eligibility for linkage is based upon having sufficient PII in at least two of three data element groups: SSN, name, and date of birth. 330,104 patient records in the 2014 NHCS and 1,642,060 in the 2016 NHCS were missing all PII and were also considered ineligible for linkage.

⁴ This group includes linkage-eligible patients who linked to HUD enrollment database at any time during the linkage interval (2014 NHCS: 2013 – 2015 HUD, 2016 NHCS: 2015 – 2017 HUD).

⁵ This percentage is calculated by dividing the number of linked patients by the number of patients in the total sample.

⁶ This percentage is calculated by dividing the number of linked patients by the total number of linkage-eligible patients.

3c. Linkage Error Estimation: CMS and HUD

The estimation of Type I and Type II errors for the CMS linkage followed the methods described in the appendix of the methods and analytic guidelines report (<https://www.cdc.gov/nchs/data/datalinkage/2016-nhcs-cms-linkage-methodology.pdf>) and are noted in Table 4.

Table 4. Linkage Algorithm Results for CMS linkage by 2016 NHCS Record Source

Record Source	Cutoff	Eligible NHCS Patients	Total Links	Deterministic Matches	Non-Deterministic Links	Est Incorrect (Type I) ¹	Est Not Found (Type II) ²
UB-04 Claims	0.85	3,294,026	771,231 (23.4%)	595,413 (77.2%)	175,818 (22.8%)	0.01%	0.17%
EHR Custom Extract	0.85	491,373	124,413 (25.3%)	111,041 (89.3%)	13,372 (10.7%)	<0.01%	0.05%
CCD	0.85	395,706	102,687 (26%)	0 (0%)	102,687 (100%)	0.12%	*

NOTES: Data are presented at patient level.

*Unable to estimate Type II linkage error for CCD records due to no SSN/HICN information on CCD records.

¹The estimated percentage of linked pairs that were not true matches

²The estimated percentage of true matches that were not linked

The estimation of Type I and Type II errors for the HUD linkage followed the methods described in the appendix of the methods and analytic guidelines report (<https://www.cdc.gov/nchs/data/datalinkage/NHCS-HUD-Linkage-Methods-and-Analytic-Considerations.pdf>) and are noted in Table 5.

Table 5. Linkage Algorithm Results for HUD Linkage by Year of NHCS

NHCS Year	Cutoff	Total Selected Links	Deterministic Matches	Probabilistic Links	Est Incorrect (Type I) ¹	Est Not Found (Type II) ²
2014 NHCS	0.9225	336,354	98,034	238,320	0.1%	2.2%
2016 NHCS	0.9225	422,920	124,328	298,592	0.1%	1.9%

¹The estimated percentage of linked pairs that were not true matches

²The estimated percentage of true matches that were not linked

4. Lessons Learned

Several critical lessons were learned from this project. This project highlighted the importance of creating high-quality linkage algorithms to combine data sources that can answer key policy and patient-centered research questions. The linkage from this project highlighted the importance of collecting patient PII data as part of the NHCS to enable such linkages 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 CMS and HUD administrative data 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

5a. Presentations:

- **January 6, 2020** Geoff Jackson, Donielle White and Sonja Williams presented “Mortality for Women within One Year after Delivery in the National Hospital Care Survey, 2016” at the International Conference on Health Policy Statistics.
<https://ww2.amstat.org/meetings/ichps/2020/onlineprogram/AbstractDetails.cfm?AbstractID=306772>
- **January 7, 2020** Lisa Mirel and Dean Resnick presented “Using Synthetic Data to Replace Linkage Derived Elements, a Case Study” at the International Conference on Health Policy and Statistics.
<https://ww2.amstat.org/meetings/ichps/2020/onlineprogram/AbstractDetails.cfm?AbstractID=306675>
- **January 8, 2020** Lisa Mirel presented “Leveraging Linked Data for Evidence Based Policymaking” at the International Conference on Health Policy Statistics.
<https://ww2.amstat.org/meetings/ichps/2020/onlineprogram/AbstractDetails.cfm?AbstractID=306597>
- **April 14, 2020** Amy Brown and Merianne Spencer presented “Using Linked Hospital Care and Mortality Data to Enhance Identification of Opioid-Involved Health Outcomes” at the Rx Drug Abuse & Heroin Summit.
<https://www.eventscribe.com/2020/rxsummit2020/agenda.asp?pfp=FullSchedule/>
- **August 4, 2020** Merianne Spencer presented “Linked Health Data from the National Center for Health Statistics: A Study of Opioid-Involved Emergency Department Visits, Hospitalization and Mortality” at the Academy Health Annual Research Meeting.
<https://academyhealth.confex.com/academyhealth/2020arm/meetingapp.cgi/Paper/41953>

- **August 4, 2020** Geoffrey Jackson, Carol DeFrances and Jill Ashman presented “Using Multiple Sources of Data to Assess and Improve Data Quality” at the American Statistical Association’s 2020 Joint Statistical Meetings.
<https://ww2.amstat.org/meetings/jsm/2020/onlineprogram/AbstractDetails.cfm?abstractid=312915>
- **August 5, 2020** Dean Resnick, Lisa Mirel, Marc Roemer and Scott Campbell presented “Adjusting Records Linkage Match Weights to Partial Levels of String Agreement” at the American Statistical Association’s 2020 Joint Statistical Meetings.
<https://ww2.amstat.org/meetings/jsm/2020/onlineprogram/AbstractDetails.cfm?abstractid=312203>
- **June 17, 2021** Karishma Chari and Donielle White presented “Demonstration of the national Hospital Care Survey: Inpatient and Emergency Department Encounters for Congestive Heart Failure, 2016” at the International Conference on Establishment Statistics.
<https://ww2.amstat.org/meetings/ices/2021/onlineprogram/AbstractDetails.cfm?AbstractID=308142>

5b. Publications:

- Jackson G, Brown A, 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>
- Ashman J, Cairns C, DeFrances C, 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>
- Resnick D, Cox C, Mirel L. Using Synthetic Data to Replicate Linkage Derived Elements: A Case Study. Health Services and Outcomes Research Methodology; 21, 389-406. 2021.
<https://link.springer.com/article/10.1007/s10742-021-00241-z>
- Campbell S, Resnick D, Cox C. Using Supervised Machine Learning to Identify Efficient Blocking Schemes for Record Linkage. Statistical Journal of the IAOS, vol 37, no. 2, pp673-680. 2021. <https://content.iospress.com/articles/statistical-journal-of-the-iaos/sji200779>

6. Future Considerations

This project opened the door for many other opportunities, including PCORTF FY20 projects that assess privacy preserving record linkage methodology and linkage of the National Hospital

Care Survey to the Transformed Medicaid Statistical Information System (T-MSIS) administrative data.

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 2016 NHCS – 2016/2017 CMS, 2014 and 2016 NHCS – 2016 HUD 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. Michelson M, Knoblock CA. “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
3. 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>
4. Christen P. Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection. Data-Centric Systems and Applications. Berlin Heidelberg: Springer-Verlag, 2012. <http://www.springer.com/us/book/9783642311635> (Accessed August 6, 2019).

5. Resnick D, Cox C, Mirel L. Using Synthetic Data to Replicate Linkage Derived Elements: A Case Study. *Health Services and Outcomes Research Methodology*; 21, 389-406. 2021. <https://link.springer.com/article/10.1007/s10742-021-00241-z>
6. Campbell S, Resnick D, Cox C. Using Supervised Machine Learning to Identify Efficient Blocking Schemes for Record Linkage. *Statistical Journal of the IAOS*, vol 37, no. 2, 673-680. 2021. <https://content.iospress.com/articles/statistical-journal-of-the-iaos/sji200779>