FY17 PCORTF 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

Carol DeFrances, Ph.D.
Deputy Director
Division of Health Care Statistics

and

Lisa Mirel, M.S.
Director, Data Linkage Program
Division of Analysis and Epidemiology

NCHS Board of Scientific Counselors Meeting

September 6, 2019

Project Overview

- Advance Patient-Centered Outcomes Research (PCOR) by linking and leveraging multisource clinical and administrative public health data.

- Leverage data from the National Center for Health Statistics (NCHS) and Centers for Medicare & Medicaid Services (CMS) to create new data infrastructures for PCOR.

- Advance understanding on mortality following hospital care, an important patient-centered outcome.

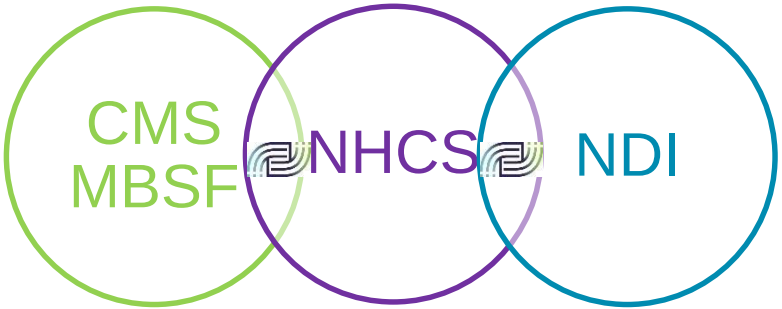
Problem

- Mortality (and its reciprocal, survival) is the only health outcome relevant to all patients, settings, and disorders.

- The major population-based data platforms used for PCOR are generally not linked to data on cause and manner of death, nor in many cases even to data on fact of death.

Project Purpose

- Link health care administrative claims and electronic health record (EHR) data from the National Hospital Care Survey (NHCS) to death record information from the National Death Index (NDI); and separately to administrative data from the Centers for Medicare & Medicaid Services (CMS) Master Beneficiary Summary File (MBSF).



Data Sources

- **NHCS** is an establishment survey, when fully implemented will provides nationally representative statistics, from claims and EHRs, on health and health care utilization from hospital inpatient stays and outpatient and emergency department visits;

- **NDI** is a centralized database of death record information on file in state vital statistics offices; and

- **MBSF** includes information on beneficiary demographic characteristics, reason for Medicare entitlement, and program enrollment type (Original Medicare vs. Medicare Advantage).

NCHS Data Linkage Program

- Cross-cutting program housed in Division of Analysis and Epidemiology
  - Supports the Center and its mission as a statistical agency
  - Links health survey data with data collected from vital and administrative records

- Designed to maximize the scientific value of the NCHS population-based surveys
  - Allow for analyses that would not be possible with each data source alone

Why Link?

- Augment available information for major diseases, risk factors, and health service utilization
  - Link exposures to outcomes
  - Provide longitudinal component
- Expand detail of information collected
  - Benefit periods
  - Health care utilization
- Reduce cost burden: follow-up surveys can be expensive

Data Linkage Methods for NHCS to NDI and MBSF

- **Identifiers used in the NHCS-NDI linkage were:** SSN, first name, last name, middle initial, month of birth, day of birth, year of birth, date of death (according to the discharge information on the hospital record), state of residence, and sex.

- **Identifiers used in the NHCS-CMS linkage were:** SSN, Medicare Health Insurance Claim Number (HICN), first name, last name, middle initial, month of birth, day of birth, year of birth, zip code of residence, state of residence, and sex.

- The linkages between the NHCS records and NDI records and CMS records were based on both deterministic and probabilistic approaches.
  - The probabilistic approach performs weighting and link adjudication following the Fellegi-Sunter method, which is the foundational methodology used for record linkage. It estimates the likeliness that each pair is a match before selecting the most probable match above a defined threshold between the two sources.

2014 NHCS linked to 2014/2015 CMS Data

Table 1. Linked 2014 NHCS – 2014/2015 CMS Medicare MBSF - Sample Sizes and Percent Linked, by Age

	Sample Size			Percent Linked	
	Total Sample	Eligible for Linkage[2]	Linked to 2014-2015 Medicare Administrative Data[3]	Total Sample[4]	Eligible Sample[5]
Age[1]					
<65	2,946,281	2,685,538	234,527	8.0%	8.7%
>=65	610,784	550,221	538,451	88.2%	97.9%
Total	3,557,065	3,235,759	772,978	21.7%	23.9%

NOTES: Data are presented at patient level. Patients were chosen by selecting the last chronological record within the survey timeframe. Age could not be determined for 1,221 patients based on available data and they are not included in this table.

[1] Age is based on the survey participant's assumed age at final encounter (date of last known contact).

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

[3] This group includes linkage-eligible patients who linked to Medicare MBSF administrative records at any time during the linkage interval (2014 - 2015).

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

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

2016 NHCS linked to 2016/2017 NDI

Table 1. 2016 NHCS - Sample Sizes and Unweighted Percentages of Patients Who Were Identified as Deceased in the Interval (2016-2017), by Age

	Sample Size			Percent Deceased	
Age[1]	Total Sample	Eligible for Linkage[2]	Identified as Deceased in 2016 - 2017[3]	Total Sample[4]	Eligible Sample[5]
<18	1,291,571	1,203,905	3,429	0.3%	0.3%
18-44	1,476,724	1,386,108	14,598	1.0%	1.1%
45-64	919,554	864,444	49,531	5.4%	5.7%
>=65	767,846	722,593	144,597	18.8%	20.0%
Total	4,455,695	4,177,050	212,155	4.8%	5.1%

SOURCE: 2016 NHCS linked to 2016/2017 NDI data

NOTES: Data are presented at the patient level. Patients may have multiple encounter records submitted during the survey data collection period but only the last chronological encounter was used in this table.

[1] Age is calculated by subtracting date of birth from the date of discharge from the last chronological encounter record submitted during the survey collection period. Age could not be calculated for 1,367,470 patients due to missing date of birth and are not included in this table. This included 767 patients who matched to the NDI and were identified as deceased.

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

[3] This group includes any patient who was identified as deceased through linkage to the NDI at any time during the linkage interval (2016 - 2017).

[4] This percentage is calculated by dividing the number of deceased patients by the number of patients in the total sample.

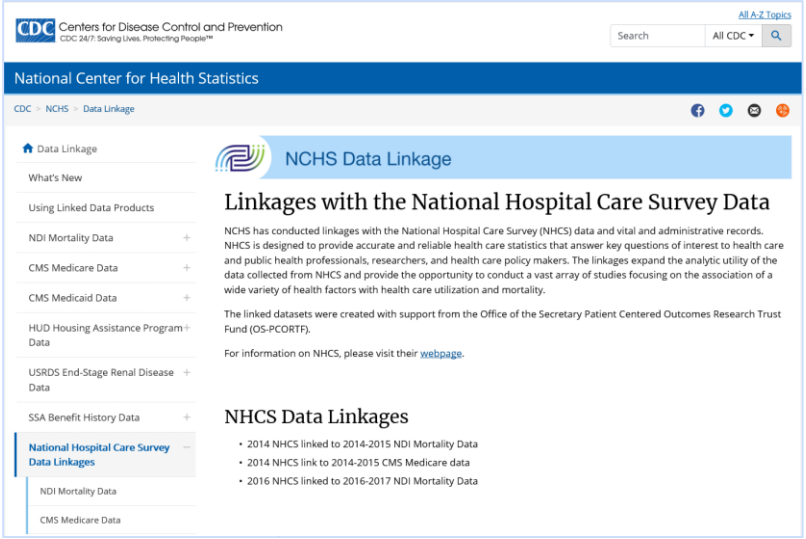
[5] This percentage is calculated by dividing the number of deceased patients by the total number of linkage-eligible patients.

Project Products – New linked datasets

Available through the

NCHS Research Data Center

- **2014 NHCS data linked to the 2014/2015 NDI**
- **2014 NHCS data linked to the 2014/2015 CMS MBSF**
- **2016 NHCS data linked to the 2016/2017 NDI**



Project Products: Analytic Guidelines, Reports, and Presentations

- **NHCS-NDI data linkage methodology**
  https://www.cdc.gov/nchs/data/datalinkage/NHCS14_NDI14_15Methodology_Analytic_Consider.pdf
  https://www.cdc.gov/nchs/data/datalinkage/NHCS16_NDI16_17_Methodology_Analytic_Consider.pdf

- **NHCS-CMS MBSF data linkage** https://www.cdc.gov/nchs/data/datalinkage/NHCS-CMS-Medicare-Llinkage-Methods-and-Analytic-Considerations.pdf

- **NHCS Alzheimer demonstration National Health Statistics Report (NHSR)**
  https://www.cdc.gov/nchs/data/nhsr/nhsr121-508.pdf

- **NHCS Pneumonia demonstration NHSR**
  https://www.cdc.gov/nchs/data/nhsr/nhsr116.pdf

- **Presentations at:**
  - 2018 Federal Committee on Statistical Methodology
  - 2018 Academy Health Annual Research Meeting
  - 2019 American College of Preventive Medicine Annual Meeting

Enhancements

- Allows for tracking mortality 30, 60, and 90 days following hospital visits for patient conditions and even 1-year survival rates.

- Allows for leveraging information on comorbidities (e.g., CMS' Chronic Conditions Data Warehouse) and Medicare coverage that may be associated with hospital-related service outcomes and health outcomes.

- Creates the infrastructure to sustain the data linkage for future years of NHCS data collection.

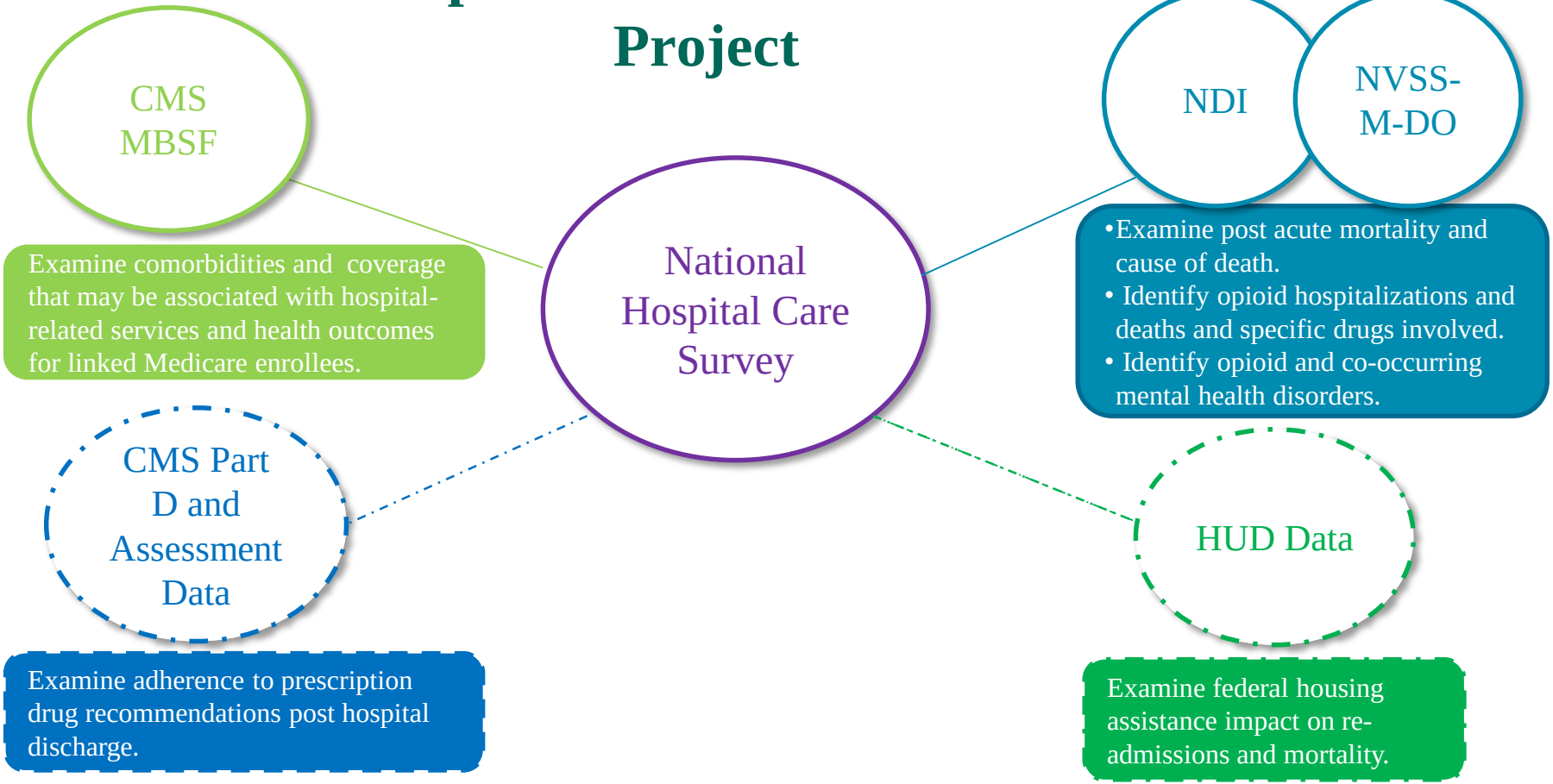
Lessons Learned

- Linkage is an inference with different methods producing different results. Transparency is very important so detailed documentation is crucial.

- Good quality PII data are critical to the linkage process.

- The successful link to the NDI and CMS MBSF indicates that the NHCS data (claims and EHR) potentially can be linked to other external data sources in the future.

Impact of FY17 PCORTF Project



Acknowledgement to the project team

Cordell Golden, DAE
Jennifer Sayers, DAE
Geoffrey Jackson, DHCS
Scott Campbell, NORC
Chris Cox, NORC
Shawn Linman, NORC
Dean Resnick, NORC

Thank you!