



## The National Hospital Care Survey (NHCS) is a Unique Source of Data on Rare Diseases

Alex Strashny, Ph.D.,

Josephine Alford, M.S.,

Catherine Rappole, M.P.H.,

Loredana Santo, M.D., M.P.H.

National Center for Health Statistics, Hyattsville, Maryland, United States

### Abstract

**Objectives.**—To demonstrate the usefulness of National Hospital Care Survey (NHCS) for studying rare diseases.

**Methods.**—NHCS contains data on millions of hospital patients from participating US hospitals, including diagnoses coded using 10th revision of International Classification of Diseases, Clinical Modification (ICD-10-CM), making it likely that some of the patients have a diagnosed rare disease. The data for 2016 are unweighted and are not nationally representative. The Orphanet Nomenclature Pack (ONP) lists 877 ICD-10 codes that correspond to 536 rare diseases. Using ONP, we identified NHCS patients with diagnosed rare diseases. We demonstrate the usefulness of NHCS for studying rare diseases by reporting, for each rare disease, the number of patients in NHCS with the disease, the average number of hospital encounters per patient, the average length of hospital stay, and the percent of patients who died either in-hospital or within 90 days post-discharge.

**Results.**—In just one year of NHCS, we identified hundreds of rare diseases with 30 or more patients each (313 rare diseases in the inpatient [IP] setting and 273 in the emergency department

---

**Contact information for corresponding author:** Alex Strashny, Ph.D., DHCS, National Center for Health Statistics, 3311 Toledo Road, Hyattsville, Maryland 20782-2064, AStrashny@cdc.gov, Phone: 301.458.4564.

Author contributions.

- Concept and design: Strashny, Alford
- Acquisition of data: Alford, Rappole
- Analysis and interpretation of data: Strashny, Alford, Rappole, Santo
- Drafting of the manuscript: Strashny
- Critical revision of the paper for important intellectual content: Strashny, Alford, Santo
- Statistical analysis: Strashny, Alford, Rappole
- Provision of study materials or patients:
- Obtaining funding:
- Administrative, technical, or logistic support: Strashny, Santo
- Supervision: Strashny
- Other:

Conflict of Interest Disclosures.

- Dr. Strashny has nothing to disclose.
- Dr. Alford has nothing to disclose.
- Dr. Rappole has nothing to disclose.
- Dr. Santo has nothing to disclose.

**Disclaimer.** All views expressed in this manuscript are those of the authors, and do not necessarily reflect those of the Centers for Disease Control and Prevention or the National Center for Health Statistics.

[ED] setting). Although ICD-10-CM codes identify a small percentage of known rare diseases, 12.9% of IP patients and 3.4% of ED patients had a diagnosed rare disease.

**Conclusions.**—NHCS is a rich source of administrative and EHR data on hospital patients with rare diseases, providing unique variables and observations on many patients. Although the percentage of patients with each rare disease is low, a large percentage of hospital patients has a rare disease.

### Précis.

NHCS has administrative and EHR data on hospital patients, providing records and unique variables on many patients with each one of hundreds of rare diseases.

---

### Introduction.

Worldwide, over a hundred different definitions of the term “rare disease” are used.<sup>1</sup> According to Orphanet, “rare diseases are diseases which affect a small number of people compared to the general population and specific issues are raised in relation to their rarity.”<sup>2</sup> Although each disease is rare, because the number of diseases is large, a large number of people could have a rare disease. While estimates vary, according to one estimate, 6.2% of the general population has one of 798 rare diseases.<sup>3</sup>

Data systems consisting of information from electronic health records (EHRs) are useful for clinical research on patients with rare diseases.<sup>4</sup> However, as many public health data systems have a limited number of observations, they have few if any observations on patients with each rare disease. This makes it difficult to study administrative or EHR data on patients in the health care setting with a specific diagnosed rare disease.

The National Hospital Care Survey (NHCS) collects administrative claims and EHR data from sampled US hospitals. These data include diagnoses coded using the 10th revision of the International Classification of Diseases, Clinical Modification (ICD-10-CM) codes, which is the US modification of ICD-10. Because NHCS has data on millions of patients, it is likely that some of the patients have a diagnosed rare disease, such as sickle cell anemia, multiple myeloma, primitive portal vein thrombosis, extrapelvic endometriosis, idiopathic intracranial hypertension, Down syndrome, or Lyme disease. However, NHCS has not been previously used in the literature as a source of data on rare diseases.<sup>5,6</sup>

In the US, the Genetic and Rare Diseases Information Center, which is maintained by the Office of Rare Diseases Research in the National Institutes of Health, refers to Orphanet as the resource for ICD coding of rare diseases.<sup>7</sup> The Orphanet Nomenclature Pack (ONP) is a list of all rare diseases known to Orphanet, along with their corresponding ICD-10 codes. While codes in ICD-10-CM typically have the same meaning in ICD-10, that is not always the case. We reviewed the ICD-10-CM codes corresponding to the ICD-10 codes from ONP, selected only the ones with the same or narrower meaning, and thus identified patients with diagnosed rare diseases in NHCS data.

We examine how many observations of patients with diagnosed rare diseases are available in NHCS for the year 2016 and illustrate the usefulness of NHCS for studying patients with each of these diseases in the hospital setting.

## Methods.

NHCS collects inpatient (IP) and emergency department (ED) data from sampled US hospitals in the form of either Uniform Billing (UB)-04 administrative claims or EHR data. These data include information captured during hospital encounters, such as diagnoses and discharge status. During data collection, diagnoses were provided by hospitals using either ICD-10-CM codes or SNOMED Clinical Terms. The SNOMED terms have been converted to ICD-10-CM codes in NHCS. Data that come from administrative claims have up to 25 diagnosis codes per encounter, while data that come from EHRs can have any number of diagnosis codes per encounter. The 2016 NHCS data are unweighted and are not nationally representative. Details about NHCS methodology are published elsewhere.<sup>8</sup> NHCS can identify multiple encounters associated with each patient within and across participating hospitals. NHCS for the year 2016 contains observations on 2.0 million IP patients and 4.5 million ED patients from 141 hospitals. Patients who were transferred from ED to IP are counted in each setting.

The 2016 NHCS has been linked with the 2016–2017 National Death Index (NDI) through the National Center for Health Statistics (NCHS) Data Linkage Program.<sup>9</sup> NDI is a centralized database of death record information compiled from state vital statistics offices and includes data on all deaths occurring within the United States. Thus, the NHCS discharge status along with the linked NDI data allow the identification of patients who have either died at the hospital or within 90 days after their latest 2016 discharge.<sup>10</sup> NHCS and NDI are available to researchers through the NCHS Research Data Center.<sup>11</sup> Data collection for both NHCS and NDI was approved by the NCHS Research Ethics Review Board. Analysis of de-identified NHCS and NDI data is exempt from federal regulations for the protection of human research participants.<sup>12</sup>

ONP, July 2020 version, lists 7,138 disorders or disorder subtypes that are “active” rare diseases according to European statistics, along with their four-character ICD-10 codes if they are available.<sup>13</sup> For this analysis, we only considered ICD-10 codes whose relationships to the rare diseases have been validated; where the term of the Orphanet nomenclature had a specific code in ICD-10; and where the ICD-10 code was either exactly equivalent to the rare disease or described a narrower clinical entity than the rare disease. In this way, the ICD-10 codes that we considered always correspond to rare diseases. ONP contains 877 four-character ICD-10 codes that identify 561 rare diseases. Some of these rare diseases are grouped together because they are identified by the same ICD-10 codes. Therefore, there are 536 groups of rare diseases identifiable by ICD-10 codes – for simplicity, we refer to these groups as rare diseases in this paper.

Within NHCS, we only considered ICD-10-CM codes if they refer to the same or narrower clinical entities as the corresponding ICD-10 codes from ONP. Thus, the ICD-10-CM codes that we considered always correspond to a rare disease as well. In the 2016 NHCS, we

found 1,496 ICD-10-CM codes that refer to the same or narrower clinical entities as the four-character ICD-10 codes from ONP. There may be more ICD-10-CM codes that refer to rare diseases, but they were not found in the 2016 NHCS. Some ICD-10-CM codes were excluded because they refer to a different clinical entity than the corresponding four-character ICD-10 code. For example, consider acute opioid poisoning, which in ONP is identified using ICD-10 codes T40.0, T40.1, and T40.2. In ICD-10-CM, T40.OX6, a subcode of T40.0, refers to underdosing of opium, which is not a type of acute opioid poisoning. We therefore excluded this ICD-10-CM code from our analysis.

For each rare disease with 30 or more patients in the data, we present certain descriptive statistics. As the objective of this paper is to demonstrate the usefulness of NHCS for studying rare diseases, the threshold of 30 was selected because researchers might choose to avoid statistical calculations involving too few patients. The specific descriptive statistics that we present are: a) the number and percentage of patients in NHCS for whom this rare disease was one of their diagnoses; b) the average number of hospital encounters per patient; c) the average length of stay (ALOS) for these patients, measured in days; and d) the percentage of patients with the rare disease who died either at the hospital or within 90 days after hospital discharge. ALOS is only available in the IP and not the ED setting. We report ALOS at the patient level – encounter-level ALOS is equal to our reported ALOS divided by the average number of encounters per patient. Note that the diagnosed rare disease might not have been the reason for the patient's hospitalization or for their death.

## Results.

In the IP setting, there were 313 rare diseases with 30 or more patients each, such as interatrial communication, acute opioid poisoning, and cardiogenic shock (Table A1). In addition, there were 190 rare diseases with anywhere between 1 and 29 patients, such as farmer's lung disease, Q fever, and didelphys uterus. Using ICD-10-CM codes, 252,394 patients (12.9% of all IP patients) were identified as having a diagnosed rare disease. Interatrial communication was the rare disease with the most patients (N = 18,235; 0.93% of all IP patients); patients with hepatoblastoma had the highest average number of IP encounters per patient during 2016 (4.3); patients with bronchopulmonary dysplasia had the longest ALOS (45.1 days); patients with cardiogenic shock had the highest mortality in-hospital or within 90 days post-discharge (58.6%).

In the ED setting, there were 273 rare diseases with 30 or more patients each, such as acute opioid poisoning, systemic lupus erythematosus, and sickle cell anemia (Table A2). In addition, there were 218 rare diseases with anywhere between 1 and 29 patients, such as Asherman syndrome, anaplastic large cell lymphoma, and mucopolysaccharidosis type 2. Using ICD-10-CM codes, 156,178 patients (3.4% of all ED patients) were identified as having a diagnosed rare disease. Acute opioid poisoning was the rare disease with the most patients (N = 14,365; 0.32% of all ED patients); patients with sickle cell anemia had the highest average number of ED encounters per patient during 2016 (3.0); patients with cardiogenic shock had the highest mortality in-hospital or within 90 days post-discharge (49.2%).

## Discussion.

NHCS presents a unique opportunity for researchers to examine the medical records of large numbers of patients with specific rare diseases from hundreds of US hospitals. As one example, consider nocardiosis. Just one year of NHCS has records on 135 IP patients and 48 ED patients with nocardiosis. In comparison, one recent study used the medical records of 60 patients with the disease, which were all the patients that the study could find who were hospitalized over a 12-year period at an Israeli hospital.<sup>14</sup> Another recent study used the medical records of 67 patients with nocardiosis, which were all the patients that the study could find who had the disease sometime within a 7-year period at an Australian hospital.<sup>15</sup> NHCS has data on appreciably more patients with nocardiosis in just one year due to collecting data from hundreds of hospitals. Since several years of NHCS data are available,<sup>11</sup> by combining years of data, even more patient records for each rare disease could be identified.

Not only does NHCS provide records on a lot of patients with each rare disease, it also contains features that may not be present in other data systems. For each patient, NHCS tracks encounters within and across participating hospitals. Thus, a unique feature of NHCS data is that it's available at both the encounter level and the patient level. The distinction is especially important for the diseases the patients with which tend to have multiple hospital encounters every year. For example, in 2016, the patients with hepatoblastoma had 4.3 IP encounters per patient. The patient-level ALOS of 26.1 days, as reported in this paper, is more reflective of the full disease burden than the encounter-level ALOS of 6.1 days (= 26.1 / 4.3).

NHCS is also capable of providing more thorough information about patient mortality. Hospital records only indicate whether a patient has died at the hospital. However, depending on the disease, an important percentage of patients might die relatively soon after hospital discharge. This would not be reflected in hospital records or in studies based on those records. By contrast, because NHCS has been linked with NDI, NHCS has data not just on in-hospital mortality, but also on post-discharge mortality. In this paper, as an illustration, we presented mortality as whether a patient died either in-hospital or within 90 days post-discharge. Other time horizons can be used as well. Previous NHCS studies have used the following time horizons for mortality: in-hospital, 1–30 days post-discharge, 31–60 days post-discharge, 61–90 days post-discharge,<sup>16</sup> and even within 1 year post-discharge<sup>17</sup>.

As seen from NHCS's data dictionary,<sup>11</sup> in addition to the variables discussed in this paper, NHCS has many other variables that researchers can analyze. These include variables about the hospital, procedure codes, revenue codes, and so on. While NHCS does have a number of variables that researchers can analyze, including unique variables, depending on the specific investigation, actual patient charts do provide greater patient details. Returning to the nocardiosis example, patient charts used in one of the studies had information on the exact species of the bacteria that caused the disease,<sup>15</sup> whereas NHCS does not have such information.

Other limitations of NHCS and this study include:

- Hospital care received before or after 2016 is not examined in this paper.
- The 2016 data are not nationally representative.
- It's possible that for some rare diseases, the ICD codes identify a narrower clinical entity than the rare disease itself, which means that some of the patients with that rare disease were not identified.
- Most rare diseases cannot be identified using ICD codes.
- Among IP patients with rare diseases, 32.8% of encounters had missing length of stay information. Mortality information was missing for 34.6% of rare disease patients. The mortality figures presented in this paper are based on patients with known mortality status.

## Conclusions.

For each one of hundreds of rare diseases, NHCS has observations on many hospital patients with that disease. In addition to having many observations, NHCS has unique variables that may not be available in other data systems.

In NHCS, the patients with rare diseases are identified using ICD-10-CM codes. It is well-known that ICD-10 codes identify only a small number of rare diseases<sup>18</sup> – they identify less than 8% of the rare diseases in the current version of ONP. Even though we are only able to identify a relatively small proportion of all rare diseases, we can see that a large percentage of hospital patients has a diagnosed rare disease.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements.

We thank Nadia Bougacha for her assistance in mapping ICD-10 codes to OrphaCodes.

## Funding/Support.

This work was performed under employment of the U.S. federal government; the authors did not receive any outside funding.

## Role of the Funder/Sponsor:

The funder had no role in the design and conduct of the study; analysis and interpretation of the data; preparation of the manuscript; and decision to submit the manuscript for publication. The funder collects and manages the data; and reviewed and approved the manuscript.

## References.

1. Richter T, Nestler-Parr S, Babela R, et al. Rare Disease Terminology and Definitions-A Systematic Global Review: Report of the ISPOR Rare Disease Special Interest Group. *Value Health*. Sep 2015;18(6):906–14. doi:10.1016/j.jval.2015.05.008 [PubMed: 26409619]
2. Orphanet. About Rare Diseases. 2021. [https://www.orpha.net/consor/cgi-bin/Education\\_AboutRareDiseases.php?lng=EN](https://www.orpha.net/consor/cgi-bin/Education_AboutRareDiseases.php?lng=EN)

3. Ferreira CR. The burden of rare diseases. *Am J Med Genet A*. Jun 2019;179(6):885–892. doi:10.1002/ajmg.a.61124 [PubMed: 30883013]
4. Bremond-Gignac D, Lewandowski E, Copin H. Contribution of Electronic Medical Records to the Management of Rare Diseases. *Biomed Res Int*. 2015;2015:954283. doi:10.1155/2015/954283 [PubMed: 26539543]
5. Lyu HG, Haider AH, Landman AB, Raut CP. The opportunities and shortcomings of using big data and national databases for sarcoma research. *Cancer*. Sep 1 2019;125(17):2926–2934. doi:10.1002/cncr.32118 [PubMed: 31090929]
6. *Rare Diseases Epidemiology: Update and Overview*. 2nd ed. Springer; 2017.
7. Genetic and Rare Diseases Information Center (GARD). ICD Coding for Rare Diseases. 2021. <https://rarediseases.info.nih.gov/guides/pages/123/icd-coding-for-rare-diseases>
8. Levant S, Chari K, DeFrances C. National Hospital Care Survey Demonstration Projects: Traumatic Brain Injury. *Natl Health Stat Report*. Jul 2016;(97):1–16.
9. The Linkage of the 2016 National Hospital Care Survey to the 2016/2017 National Death Index: Methodology Overview and Analytic Considerations (2019).
10. Linked Data on Hospitalizations, Mortality, and Drugs: Data from the National Hospital Care Survey 2016, National Death Index 2016–2017, and the Drug-Involved Mortality 2016–2017 (2020).
11. National Center for Health Statistics. National Hospital Care Survey (NHCS). 2021. <https://www.cdc.gov/rdc/b1datatype/Dt1224h.htm>
12. National Center for Health Statistics. Publishing Guidelines. 2021. <https://www.cdc.gov/rdc/b6pubeyond/pub600.htm>
13. Orphanet. Orphanet Nomenclature Pack (July 2020). 2021. <http://www.orphadata.org/cgi-bin/ORPHAnomenclature.html>
14. Margalit I, Goldberg E, Ben Ari Y, et al. Clinical correlates of nocardiosis. *Sci Rep* Aug 31 2020;10(1):14272. doi:10.1038/s41598-020-71214-4 [PubMed: 32868850]
15. Paige EK, Spelman D. Nocardiosis: 7-year experience at an Australian tertiary hospital. *Intern Med J* Mar 2019;49(3):373–379. doi:10.1111/imj.14068 [PubMed: 30091232]
16. Jackson G, Chari K. National Hospital Care Survey Demonstration Projects: Stroke Inpatient Hospitalizations. *Natl Health Stat Report*. Nov 2019;(132):1–11.
17. Spencer MR, Flagg LA, Jackson G, DeFrances C, Hedegaard H. National Hospital Care Survey Demonstration Projects: Opioid-involved Emergency Department Visits, Hospitalizations, and Deaths. *Natl Health Stat Report*. Jun 2020;(141):1–19.
18. Nestler-Parr S, Korchagina D, Toumi M, et al. Challenges in Research and Health Technology Assessment of Rare Disease Technologies: Report of the ISPOR Rare Disease Special Interest Group. *Value Health*. May 2018;21(5):493–500. doi:10.1016/j.jval.2018.03.004 [PubMed: 29753344]



### Highlights

- Each year of the National Hospital Care Survey (NHCS) has administrative and electronic health record (EHR) data on millions of hospital patients. These data include ICD-10-CM diagnosis codes, which can identify patients with certain rare diseases.
- Each year of NHCS has data on hundreds of rare diseases, such as sickle cell anemia, multiple myeloma, primitive portal vein thrombosis, extrapelvic endometriosis, idiopathic intracranial hypertension, Down syndrome, and Lyme disease. Patient-level variables include number of hospital encounters, length of stay, whether the patient died in-hospital or within a certain time post-discharge.
- Although the percentage of hospital patients with each rare disease is small, a large percentage of hospital patients has a diagnosed rare disease.