Using Uniform Bill (UB)-04 Administrative Claims Data to Describe Hospital-Based Care

Sonja Williams[†], Naga Shanmugam[†], Amy M. Brown[†], Carol DeFrances[†]

† Division of Health Care Statistics, National Center for Health Statistics, 3311 Toledo Road, Hyattsville, MD 20782

Abstract

The National Hospital Care Survey (NHCS), conducted by the National Center for Health Statistics, is designed to describe national patterns of health care encounters in hospital-based settings. The goal of NHCS is to collect reliable and timely health care utilization statistics. Hospitals can participate in NHCS by transmitting either Uniform Bill (UB)-04 administrative claims or electronic health record (EHR) data. Since EHR adoption is still being implemented at many hospitals, most hospitals currently participating in NHCS send UB-04 claims data. This proceedings paper discusses the benefits and challenges of UB-04 data. Benefits include the collection of personally identifiable information (PII) that allows for data linkages across hospital settings and to outside data sources, such as the National Death Index. It also allows the inclusion of both services that were previously excluded from outpatient settings and information on intensive care use and revisits. The collection of the UB-04 data has also posed some challenges. Many hospitals could not send the UB-04 data in the desired file format, so the data submission specifications had to be expanded. Since more than one claim is often generated for a single episode of care, methods had to be developed to consolidate multiple claims into one record. The identification of substance-involved emergency department (ED) visits from the UB-04 claims data has proved to be difficult. Algorithms are being developed that incorporate diagnosis, procedure, and symptom information to more accurately identify substance-involved ED visits.

Introduction

The National Hospital Care Survey (NHCS) is one of the National Health Care Surveys conducted by the Division of Health Care Statistics (DHCS) at the National Center for Health Statistics (NCHS). NHCS began in 2011 and integrates three long-standing surveys -- the National Hospital Discharge Survey (NHDS), the National Hospital Ambulatory Medical Care Survey (NHAMCS), and the Drug Abuse Warning Network (DAWN). NHDS, conducted by NCHS from 1965-2010, was the longest continuously fielded survey of inpatient care. NHDS collected data from a sample of hospitals through abstraction of medical records and through electronic data files. NHAMCS, conducted by NCHS and currently still in the field, collects data from a sample of hospitals on visits to emergency departments (EDs), and outpatient departments (OPDs), including ambulatory surgery. DAWN, was established in 1972 by the Drug Enforcement Administration, then conducted by the National Institute on Drug Abuse in 1980, and finally conducted by the Substance Abuse and Mental Health Services Administration (SAMHSA) from 1992-2011[1]. It collected data on substance-involved ED visits through medical record abstraction.

The goal of NHCS is to provide accurate and timely healthcare utilization statistics. The survey has three main objectives. The first objective of NHCS is to move toward electronic data collection utilizing electronic health record (EHR) data over time. A second objective is to permit special in-depth studies, such as examining substance-involved ED cases. The final objective is to link episodes of care across hospital settings as well as link to other data sources such as the National Death Index (NDI) and Medicare data.

The NHCS's target universe is inpatient discharges and in-person visits made to EDs and OPDs, including ambulatory surgery. Eligible hospitals are non-institutional, non-federal hospitals with six or more staffed beds. A base sample of 500 hospitals and a reserve sample of 500 additional hospitals was drawn. In 2013, to help with the estimates for substance-involved ED visits, 81 hospitals with 500 staffed beds or more were added from the reserve sample to the base sample of 500 hospitals, making the current sample equal to 581 hospitals. The 81 additional 500 or more staffed bed hospitals is intended to increase the number of ED visits collected and subsequently increase the number of substance-involved ED visits, as large hospitals usually have more of these types of visits. Currently,

recruitment efforts are focused on hospitals with 300 or more staffed beds. To date, there are approximately 100 hospitals actively participating in the survey and sending Uniform Bill (UB)-04 administrative claims data.

Use of the Uniform Bill (UB)-04 for NHCS Data Collection

Initially, NCHS requested the submission of UB-04 claims from participating hospitals to collect the information needed for the survey. The UB-04 is a data specification that is used by hospitals to transmit patient encounter information to the Centers for Medicare and Medicaid Services (CMS) and to insurance payers for payment. This standard was created by the National Uniform Billing Committee (NUBC) [2] who is tasked with maintaining and regulating its use. It is transmitted electronically via the American National Standards Institute (ANSI) X12 837 transmittal format which is a messaging standard used to transmit data between two entities under the Health Insurance Portability and Accountability Act of 1996 (HIPAA)[3]. This standard has several different versions, of which two are of interest to NHCS. The 837 Research version (837R) includes additional variables intended for research use, such as race, ethnicity, and marital status. However, this version of 837 is not required for use by hospitals. If hospitals cannot provide data using 837R, the 837 Institutional version (837I) is accepted – it is used by most hospitals since it is required by CMS.

Benefits of UB-04 Data

Collection of all inpatient and ambulatory encounters - Historically, NHDS and NHAMCS have been limited to the collection of survey data from a sample of manually abstracted health records from participating hospitals that were subsequently weighted to create estimates of national hospital care. The direct submission of electronic UB-04 claims enabled NCHS to collect all hospital inpatient and ambulatory encounters for the first time. By collecting all encounters, it is possible to study rare diseases that previously did not yield enough cases to create reliable estimates in NHDS or NHAMCS. For example, in the 2013 NHCS data collection, there were over 11,000 inpatient cases with a first-listed diagnosis of traumatic brain injury (TBI). Previously, five years of NHDS needed to be combined to yield almost the same number TBI inpatient discharges.

Access to new variables and more detailed information - The UB-04 data allow collection of new variables and more detailed categories for previously collected variables. In NHDS and NHAMCS, abstraction was limited by the time field staff were allowed to collect data in participating hospitals. To keep time and cost burden as low as possible, these surveys restricted data collection to a relatively small number of core data elements on a patient record form that could be efficiently abstracted by field staff. With the UB-04 data, there are no restrictions to what can be collected unless it is not a part of the UB-04 structure. NHCS collects up to twenty-five diagnoses and procedures where previously NHDS only collected seven diagnoses and four procedures.

In the UB-04, there are over thirty categories for discharge status, so now NHCS has the ability to study detailed discharge outcomes related to a specific diagnosis or procedure. The same is true for other key variables such as point of origin and reason for ambulatory visit. Additionally, since the UB-04 is a billing source, it contains information on actual services rendered during a hospital stay. These services rendered, called revenue codes, allow the collection of new data on topics such as intensive care use, dialysis, and radiology. Previously, these items could not be collected because of the burden it would have incurred to the hospital.

Enhanced insurance payer coding - The UB-04 provides a coded variable for payer and also the actual name of expected payers. NHCS uses the payer name and the coded variable, when necessary, to create detailed categories for payer that were not available in previous surveys. The payer or expected source of payment typology is based on the Public Health Data Standards Consortium's Source of Payment Typology version 5.0[4] and was enhanced by additional categories created from payer names found in the National Health Interview Survey's[5] insurance codes list and additional payer names found in the NHCS data. The NHCS data collection contractor uses a supervised learning algorithm called a support vector machine to estimate a detailed payer category. This algorithm uses several variables for the estimation, including expected payer name, patient age group, sex, marital status, and patient address.

As an example, while UB-04 provides just one code for Medicaid, NHCS has several detailed payer categories for Medicaid (Table 1).

Table 1. Example of payer codes available in UB-04 and NHCS

Code provided by UB-04	Codes available in NHCS
Medicaid	State SCHIP programs
	Medicaid (Managed Care)
	Medicaid (Non-managed care)
	Long Term Care Medicaid

Greater in-scope ambulatory visits - Another benefit of the UB-04 data is that it allows for all outpatient visits to be in-scope for data collection. In NHAMCS, there is a restriction on the types of in-scope clinics because of cost and burden. With the collection of UB-04 data, NHCS has been able to broaden its scope and collect all OPD encounters within sampled hospitals. This expansion of the definition includes previously excluded ambulatory visits such as radiology, infusion therapy, dental, and rehabilitation.

Collection of personally identifiable information (PII) - Finally, the inclusion of personally identifiable information (PII) in the UB-04, including name, address, and social security number, allows for the identification of unique patients, repeat visits to the same hospital, data linkage within and across each hospital setting and data collection year, and data linkage to outside sources. The collection of PII enables the creation of a patient identification variable (Patient ID) that can track a person's journey through the health care system within sampled hospitals. For example, NHCS can track a person who comes to the ED presenting with severe symptoms, leading to a diagnosis of diabetic ketoacidosis. That person is admitted and spends time in the hospital's intensive care unit. That same patient's entire hospital stay is included through to discharge, with follow-up outpatient care at one of the hospital's outpatient clinics also captured in the data.

The Patient ID allows patients to be tracked if they come back to the same hospital within the data year and even across data years. The PII also allow the data to be linked to the NDI to report mortality rates, for example 30- and 60-day mortality. Currently, the 2012 inpatient and emergency department NHCS data have been linked to the NDI with promising results, and plans to link the 2013 NHCS data are underway.

Challenges of UB-04 Data

Although there are many benefits in using the UB-04 as a data source, there are challenges as well. The UB-04 was created for billing rather than research purposes. As such, the collection of race and ethnicity are not required on the UB-04, but these are important variables for health care research. Additionally, no clinical information such as vital signs, medications, and lab results are collected. These challenges posed unique opportunities for learning new strategies to work with administrative data.

Difficulty in data transmission - Initially, hospitals were only allowed to submit via the 837R or 837I transmittal vehicle. It was soon discovered that many hospitals only had the capability to send data directly to CMS and had no way to send to another source. Other hospitals had no knowledge on how to export the data to other sources. Finally, some larger hospitals had archival systems that inhibited access to the data on the quarterly timetable initially requested.

To combat the issue of data transmission, the best strategy was to be flexible. Once the difficulties were identified, efforts were made to decrease burden on the hospitals. Several formats were accepted, such as fixed column, CSV, and even Excel. In some cases, this flexibility helped get data that was otherwise inaccessible via the 837 format such as race, marital status, and charity and self-pay records. For example, some states require the reporting of variables such as race and marital status. If a hospital submitted their state file to us, we received those additional variables. The submission schedule was also relaxed to include daily feeds, monthly or quarterly submissions, and if it was easier for the hospital, even a yearly transmission was accepted.

Generating a single encounter from multiple claims - NHCS uses single encounters for its unit of analysis but UB-04 claims are not necessarily submitted in that fashion. The single encounter is defined by the date the patient enters

into the health care facility to the date the patient leaves the facility. UB-04 claims data are usually summaries of all care occurring in the hospital for a particular time period, not necessarily just one encounter; alternatively several claims could all be iterations of the same hospital stay. For instance, hospitals may bill the insurers for patients with very long stays in the hospital on a monthly basis resulting in many claims all related to one hospital stay. Each claim may have slightly different information that needs to be pieced together to create a complete hospital stay. Also, hospitals often create multiple claims for the same hospital stay when a patient has multiple insurance providers. This is problematic because it is difficult to decipher what claim belongs where and if all the information for one encounter is included in the submitted claims. Finally, in the outpatient setting, several visits to the same department are often bundled together into one claim, such as weekly or monthly cancer treatments or multiple physical therapy sessions.

Through consulting with experts in claims clearinghouses and researching other methods of deciphering claims data, two methods were established to address this issue. The first method, called "de-duplication," identifies groupings of claims for the same patient and identifies the information that would create a complete discharge for an inpatient or ambulatory visit. This technique uses different methods to group encounters by specific variables. Once the encounters are grouped, the de-duplication process reviews the grouped encounters and duplicate groups that identify a set of claims for the same encounter are combined to create a single encounter. The second method, called "claims splitting", is used exclusively with ambulatory data. Before the standard "de-duplication" process begins, the ambulatory data go through this claims splitting process that splits bundled claims into individual visits.

To verify the accuracy of the de-duplication and claims splitting methods, information is gathered during the annual hospital interview about how hospitals report discharges and visits. The hospital's reported totals, along with other published data, are used to compare the final encounter totals created from the two methods described above.

Missing some types of visits - Unfortunately, because UB-04 claims are only universally required when there is a government or commercial payer to be billed, some hospitals do not generate claims for patients who are self-pay, charity, research, or prisoners. Although some states do require all hospital encounters to have a claim generated, this lack of a universal requirement leads to undercounts of all hospital encounters.

To mitigate this problem, hospitals are asked during the intake process if their UB-04 data include all inpatient and ambulatory visits and if not, what type of visits are excluded. For the future weighting of the data, this information will be used to make adjustments to the weighting procedures to account for those missing records. If the state requires universal reporting of all visits, hospitals are asked to submit their state reporting file. Hospitals are also asked if there is another format that includes the missing encounters, particularly an EHR system. All encounters should be included in an EHR system, since EHRs are not dependent upon a specific payment source.

Identifying substance-involved ED visits - One of the objectives of the survey is to produce national estimates of substance-involved ED visits, similar to those previously produced by DAWN. In DAWN, field staff were placed in sampled hospitals on an ongoing basis and all ED visits occurring on designated dates were reviewed for evidence of substance involvement[6]. Initially in NHCS, likely substance-involved cases needed to be identified prior to going into the hospital to abstract records to collect the clinical information needed to confirm substance involvement. These submitted claims were used to target a sample of visits for abstraction. A single list of ICD-9-CM codes was created to identify visits involving a wide range of drug, medicinal, nonmedicinal, and biological substances. This list was refined many times to try to identify the same types of visits previously captured by DAWN. Visits with at least one of the substance-involved ICD-9-CM code in any of the 25 collected diagnosis fields were flagged, and a simple random sample of those visits were abstracted by field staff to verify suspected substance involvement.

Unfortunately, each version of the single ICD-9-CM code list yielded a high number of false positives. SAMHSA used a very specific definition of a substance-involved visit for DAWN with two main criteria: (1) the substance of interest must have been used recently and (2) the substance use must have been directly related to the reason for the ED visit. It was discovered that having a diagnosis of substance use was not necessarily indicative of either criteria listed above. In many instances, reported diagnoses on the claims merely represented a patient's history of past substance use that continued to be listed on subsequent claims. Or, a visit met the first condition—a patient did use a substance recently, but the second criteria was not met—the reason for the ED visit was not attributed to substance use.

Rather than try to produce national estimates for all of the substances previously included in DAWN, refinements for a small number of high priority substances are being made to the methodology. This smaller list contains illicit drugs such as heroin and cocaine, and prescription drugs such as benzodiazepines and antidepressants. Once visits involving these substances can be reliably identified, additional substances will be gradually added to the list.

Using one broad algorithm that includes a wide range of substances does not reliably predict visits for a particular substance. Therefore, multiple substance-specific algorithms, each with a unique ICD-9-CM code list, are currently being developed. Each unique code list is being used in combination with a code list that identifies common substance-related symptoms, such as palpitations and altered mental state as well as procedures and tests such as drug screenings and drug antagonist injections, to increase the likelihood that visits involving recent substance use as the reason of the visit are identified. All available data with this information, including reason for visit, diagnosis, E-code, and procedures are examined so that substance involved cases can be reliably identified.

A stratified sampling design is under development to sample an array of visits with varying degrees of substance involvement. This design will allow the oversampling of visits with the strongest evidence of substance involvement as well as the inclusion of some visits with weak or no evidence of substance involvement that are actually true substance-involved cases, that would have otherwise been missed. While these new substance-involved visit identification methods are still being tested and refined, preliminary results indicate these new methods may help reduce false positives and yield large enough case counts to make national estimates for a select number of high priority substances.

Future Direction

The collection of UB-04 claims data for NHCS has increased access to data previously not available through NHDS or NHAMCS. Despite its challenges, the UB-04 is a rich source of data. Along with the strategies being implemented to address the challenges posed by UB-04, there have also been efforts to look for other ways to collect data. Looking toward the future, NCHS staff are asking hospitals to submit EHR data as a first option, which will minimize the need for additional abstraction used in collecting substance-involved visit information. To standardize this effort, NCHS staff created a Health Level-7 (HL7) clinical document architecture (CDA) implementation guide for hospitals to use in programming the needed variables for NHCS. Also, NHCS was named in the interim and the final Stage 3 Meaningful Use (MU) Rule. This means that hospitals can now register and participate in NHCS and gain MU credit, possibly making participation in NHCS more attractive to hospitals. Finally, NCHS has partnered with UHC (formerly University HealthSystem Consortium) to obtain data from their members as an interim step, until more hospitals are able to transmit EHR data to NHCS.

References
1. About Emergency Department Data. Substance Abuse and Mental Health Services Administration. Available from: http://www.samhsa.gov/data/emergency-department-data-dawn/about.
2. About the NUBC. NUBC National Uniform Billing Committee. Available from: http://www. nubc. org/aboutus/index. dhtml.
3. 837 101 Tutorial. Available from: http://www. phdsc. org/standards/x12/tutorial. asp.
4. Users guide for source of payment typology Version 5. Public Health Data Standards Consortium Payer Type Subcommittee. October 2011. Available from: http://www. phdsc. org/standards/pdfs/SourceofPaymentTypologyUsersGuideVersion5.0_final. pdf
5. National Health Interview Survey. Available from: http://www.cdc.gov/nchs/nhis.htm
6. Center for Behavioral Health Statistics and Quality (2013). *Drug Abuse Warning Network Methodology Report, 2011 Update*. Rockville, MD: Substance Abuse and Mental Health Services Administration.