

Data Resource Profile: State Inpatient Databases

David Metcalfe^{1*}, Cheryl K. Zogg², Elliott R. Haut³, Timothy M. Pawlik⁴, Adil H. Haider⁵, Daniel C. Perry¹

1. Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences (NDORMS), University of Oxford, Oxford, UK.
2. Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA.
3. Wexner Medical Center, The Ohio State University, Columbus, OH, USA.
4. Center for Surgery and Public Health, Harvard Medical School, Boston, MA, USA.
5. Yale School of Medicine, New Haven, CT, USA.

Profile in a nutshell

- The SIDs contain data on almost all inpatient discharges from hospitals in participating states. They were established from a range of data sources for the specific purpose of undertaking health-related research.
- The first SIDs are available starting in 1990.
- SID data elements vary by year and state, but typically include demographic characteristics, admission and discharge details, diagnoses and procedures, charge/cost and payment details, and outcomes such as in-hospital death. They are linkable to information contained in AHA.
- Some SIDs permit longitudinal linkage so that individual patients can be tracked between hospital admissions. Patients can often also be tracked through Emergency Department visits through linkage to other state-level HCUP databases.
- Access to data is swift and at reasonable reduced cost, particularly for non-profit and student researchers.

Data Resource Basics

Scope

The State Inpatient Databases (SID) comprise a collection of state-specific encounter-level administrative claims from 32/50 states in the United States (US). These databases are maintained by the US Agency for Healthcare Research and Quality (AHRQ) as a part of the organization's Healthcare Cost and Utilization Project (HCUP). As part of a family of six other HCUP databases (Table 1), the SIDs represent the largest collection of longitudinal hospital care data in the US inclusive of all insurance payers (*e.g.* Medicare, Medicaid, private insurance, and uninsured) and all patient ages. Taken together, the SIDs capture more than 97% of all eligible hospital discharges within each state.

Purpose of data collection

Originally created as the "Agency for Health Care Policy and Research," AHRQ was established by the US Omnibus Budget Reconciliation Act of 1989 with the purpose of enhancing the "quality, appropriateness, and effectiveness of health care services [...] through the establishment of a broad base of scientific research."¹ In particular, the Act required that databases be created to support healthcare research. The HCUP-family of databases were, therefore, established as a federal-state-industry partnership sponsored by AHRQ in order to "enable research on a broad range of health policy issues, including cost and quality of health services, medical practice patterns, access to health care programs, and outcomes of treatment at the national, state, and local market levels."² The HCUP SIDs maintain data expressly for research purposes dating back to 1990.

Structure

The SID share a common structure, although the precise list of available data elements vary slightly by state and calendar year. Differences in key variables used to link SID data are presented in Table 2. The SIDs are organised by calendar year based on the date of hospital discharge. The data for each SID

year (e.g. Florida 2014) are presented in up to five separate files: core, charges, American Hospital Association (AHA) linkage, Diagnosis Related Groups (DRGs), and disease severity. The core file contains the discharge-level data elements that make up the majority of each SID. Some core data elements are universal (e.g. AGE, DISPUNIFORM [discharge disposition], DX1-DXn [primary and secondary diagnosis codes], PR1-PRn [primary and secondary procedure codes] and LOS [length of stay]). Others vary by state (e.g. ZIP [patient residential zip code], RACE [patient race/ethnicity], and visitLink/DaysToEvent [variables needed for longitudinal patient tracking]). The charges file contains information needed to convert total hospital charges (core: TOTCHG) into estimated costs incurred. The AHA linkage file permits linkage to data from the AHA Annual Survey Database, which contains additional hospital-level information on institution-specific parameters such as the number of hospital beds, the number of full-time residents, or cancer hospital/trauma centre designations.³ The DRGs file represents an attempt by the US Centers for Medicare and Medicaid Services to classify ICD-9-CM/ICD-10-CM diagnosis codes into meaningful diagnostic groups at the level of an inpatient stay record.⁴ For example, two ICD-9-CM codes (401.1 “hypertension, benign” and 401.9 “hypertension, unspecified”) are combined to form a single Clinical Classifications Software category: 98 “essential hypertension.”⁵ Finally, the disease severity file provides data from AHRQ Co-morbidity Software using individual inpatient stays as the unit of observation to provide information on the presence or absence of pre-specified comorbidities and overall categorical rankings of a patient’s risk of mortality and disease severity.

Research uses

The SIDs are frequently used because of their universal coverage, ability to track patients longitudinally between hospital admissions, ability to link to AHA data, inclusion of geographic specific information such as patient residential zip codes, and ability to identify individual US states. As the SIDs are state-specific, these data can be used to evaluate health policy interventions that take place at the state level or to compare practice patterns/outcomes that vary between states by using data

from neighbouring or otherwise comparable states as regional controls. Applications in published literature have ranged from geographic mapping⁶ of access to care and variations in quality metrics to assessment of changes in practice patterns and outcomes over time⁷ to more traditional analyses of health policy utility and predictors of adverse outcomes/disparities and quality ranking.⁸⁻¹⁰

Processing cycle and frequency of data collection

The AHRQ partners with a range of organizations in participating states, which include hospital associations and private organisations, as well as the federal government to collect data. Datasets are made available through the HCUP Central Distributor when they have been adequately processed and approved by state partner organizations. This practice introduces a delay (typically 18-24 months) before data becomes available for analysis, although such intervals vary by state.

Linkage within the SID

A number of SIDs (Figure 1; Table 2) include a re-visit variable (“visitLink”), which is a unique patient identifier that permits individuals to be tracked between discharge records from a randomly generated time origin point encoded in the variable “DaysToEvent.” Starting in 2003, revisit data were available for a total of 8 states. As of 2015, that number increased to 14 states, with a total of 16 states contributing re-visit data between 2009-2016 (Table 3). Although patients can be tracked between years within the same state, they cannot be tracked between admissions to hospitals in different states.

Linkage to other datasets

Individual patients can be tracked over time between inpatient admissions (above) using SIDs in isolation. However, linkage of SIDs to State Emergency Department Databases (SEDD) and State Ambulatory Surgery and Services Databases (Table 1) can also facilitate tracking of patients between emergency department attendances, hospital admissions, and outpatient visits. Complete assessment

of all Emergency Department visits within a state requires a combination of SID and SEDD. Many SIDs provide a variable that permits linkage to the AHA Annual Survey Database, which contains a comprehensive range of hospital characteristics that can be used to enrich individual SID. The AHA Annual Survey Database includes more than 1,000 institution-level data elements about more than 4,600 hospitals within the US.³

Data Collected

As described in the “Structure” section above, the availability of data elements within a given SID depends on the state and calendar year. Some data elements (e.g. discharge destination) are available in both a “raw form” as provided by individual states (e.g. “disp_X”) and as a second variable that has been standardised by HCUP (e.g. “dispuniform”). Key linkage variables are shown in Table 2. Information on other important parameters is provided below.

Demographic data

Key SID demographic data elements include information on age, sex, race, marital status, zip code (first 3 digits or all 5 depending on the SID), primary payer/insurance status (e.g. Medicare, Medicaid, private insurance, or self-pay), income quartile (median household income based on residential zip code), and state of residence. Age is available in days for patients under 1 year and months for patients under 11 years. There are also data elements specific to new-borns, including birth month and year, as well as birth weight.

Admission and discharge data

Key data elements on admission and discharge data include admission time (hour, month) as well as type (e.g. emergency, urgent, elective) and whether or not the admission was during a weekend, which is defined as an admission on a Saturday or Sunday. There is also a flag to identify patients transferred in from another hospital. Discharge time, month, and quarter as well as destination are

reported (e.g. transfer to a skilled nursing facility, transfer to an inpatient rehabilitation facility, died, transfer to hospice, etc.).

Diagnoses and procedures

Diagnoses are recorded using ICD-9-CM diagnostic and procedure codes from the beginning of data collection until 31st September 2015 when the US switched to ICD-10. Since 1st October 2015, the SID have used ICD-10-CM diagnostic and ICD-10-PCS procedure codes.¹¹ For each discharge record, up to 15 diagnostic and 15 procedure codes are reported as are up to 4 E-codes for designated trauma patients. By convention, the primary diagnosis and procedure codes occupy the first-named variable, i.e. “DX1” and “PR1,” respectively. A “present on admission” variable can be used to distinguish between co-morbidities present at the time of admission and those that developed in hospital. However, the present on arrival variables are not always consistently reported. Researchers have often relied more heavily on clinical experience as to which diagnoses are likely to constitute a pre-existing medical condition when incorporating information from secondary diagnosis codes. For example, chronic diseases (e.g. diabetes) might be assumed to represent a pre-existing condition whereas this is less likely to be true for acute diagnoses such as Acute Respiratory Distress Syndrome. Researchers have previously used user-generated software such as CHARLSON¹² and ICDPIC,¹³ which can translate ICD-9-CM codes to standard disease/injury scores such as the Charlson Comorbidity Score, Elixhauser Co-morbidity Index, and Injury Severity Score.^{8, 10, 14, 15} The SIDs also include Diagnosis Related Groups (DRGs) for each disease diagnosis that are defined using the Medicare DRG Grouper¹⁶ algorithm during HCUP processing.

The date of each coded procedure is available with reference to the day of admission. For example, a record with “3512” under variable “PR1” and “3” under “PRDAY1” would indicate that the primary procedure was a mitral valve replacement (ICD-9-CM 35.12 “Open heart valvuloplasty of mitral valve without replacement”), which took place on the third day after admission.

Possible outcome measures

Variables that have previously been used as outcomes include documented occurrence of specific complications, in-hospital mortality, receipt of specific interventions, scores on pre-defined quality metrics, discharge destination, length of stay, total hospital charges in US dollars for the corresponding calendar year (convertible to total hospital costs), and routine versus non-routine discharges. Longitudinal linkage variables enable assessment of readmissions, prior hospitalizations, repeat operations and/or repeat need for hospitalizations under similar or changing diagnoses, and corresponding times to all of these events. Although, SID data have been used to identify readmissions, there is no single data element that can be used to determine whether a readmission was “unplanned.”¹⁷ The data also lack information on complications or fatalities that occur outside of a hospital setting. Some complications and other outcomes are particularly susceptible to surveillance bias, i.e. “the more you look, the more you find”.¹⁸ Although this problem can be addressed in some datasets by controlling for diagnostic procedures, this is difficult for researchers using the SIDs as some low cost procedures are not routinely captured in the billing driven HCUP dataset suite.¹⁹

Data Resource Use

There has not been a definitive bibliography of research papers published using the SID. However, a search of PubMed using “state inpatient database*”[tw] OR “state inpatient databases” identified n=536 publications as of December 2018 (Figure 2). Figure 2 shows that SID use has increased every year since the first publication appeared in 2000.

The ability to track patients between admissions over time in SID has been utilized by researchers interested in hospital readmissions.¹⁵ SID containing identifiers for individual physicians and hospitals have been used to determine annual case volumes and search for associations between volume and patient outcomes.^{10, 14} SID that can be linked to the AHA Annual Survey Database have been analysed

for reported associations between hospital characteristics (e.g. bed size, teaching status, rural/urban location) and outcomes.^{20, 21} SID have also been used to describe the burden of disease,²² evaluate disparities,^{8, 23} challenge measures of hospital quality,^{24, 25} describe inter-hospital variation,²⁶ and evaluate health policy initiatives,²⁷ including differences in implementation at the state level.²⁸

Strengths and Weaknesses

Very few of the papers identified in the search (above) sought to validate or critically evaluate SID, although this is likely to change as the databases are increasingly used by health services researchers. The AHRQ has undertaken a number of methodological exercises aimed at improving the quality of HCUP data.²⁹ Recognized limitations of SIDs include the potential for absent or incorrect coding of relevant information and a lack of nuanced clinical detail. Databases are not identical across states, do not track information between state lines, and do not capture information that occurs outside of a hospital setting—most notably including a lack of links to death registry data. The use of research checklists, particularly those designed to improve analysis of HCUP data sets, may help mitigate these limitations.³⁰⁻³³

Coverage

A major strength of the SID is the comprehensive coverage with approximately 97% case capture.^{14, 29} The broad coverage is particularly important in the US, where health datasets are often based on payer status and so may not be representative of the broader population. However, as there is substantial variation among individual states, the SID are less well-suited to generate national estimates. The large size of some SID (e.g. almost 4 million discharges per year in California³⁴) can help to study rare diseases or outcomes, particularly when multiple SIDs are combined.

Longitudinal and horizontal linkage

The ability to track patients over time is a valuable feature of the SIDs, particularly as linkage to other HCUP datasets can permit follow-up through future Emergency Department attendances and day-case surgical procedures, as well as hospital admissions. This linkage permits researchers to capture a range of outcomes such as future operations and healthcare utilization.

Out of hospital deaths

Although longitudinal and horizontal linkage through healthcare episodes are particular strengths of SIDs, an important weakness is the inability to capture deaths outside of the hospital setting. The re-visit identifiers are encrypted by HCUP and data recipients must agree not to link visits to other datasets. As SID records cannot be linked to death registration data, researchers are limited to using in-hospital death, which is known to be affected by early discharge.³⁵ For example, in-hospital hip fracture mortality in the California SID (2007-2011) is 1.8% with a corresponding reported median length of stay of 5 days with a US care specific high rate of discharge to post-operative rehabilitation.¹⁴ In the United Kingdom, where index hospitalization and rehabilitation are usually combined, in-hospital mortality is much higher at 8.1% with a reported median length of stay of 15 days.³⁶ Lacking a constant follow-up point (e.g. 30-days post-injury) such comparisons can be fraught with difficulty.^{17,}

³⁵ There is, however, evidence to suggest that in-hospital mortality can be used as a surrogate for 30-day mortality,³⁷ but this assumption needs to be explicitly considered by researchers undertaking comparative analyses. The SIDs can, however, account for patients who survive the initial hospitalization but who die during a subsequent readmission within a set number of days or who are discharged but then return to hospital as “dead on arrival.”

Patients re-admitted within other states

Although a number of states permit longitudinal tracking of patients between hospital admissions (Figure 1; Table 3), the unique identifiers do not permit tracking *between* SIDs. It is therefore possible that patients living near state borders might be re-admitted to a hospital in a different state and not

be captured within the SID where they initially presented for care. At one extreme, the National Association of Health Data Organizations (NAHDO) reported that 4.8% of patients in the Maryland SID were residents of neighbouring states and that 1.4% of Maryland residents discharged from Maryland hospitals were initially treated in a different state.³⁸ It should be noted that Maryland is the ninth smallest state by total area in the US and that it shares close geographic borders with Virginia and Washington DC. As such, it is not surprising to see a relatively high level of cross-state exchange within the Maryland SID. By contrast, the NAHDO found that only 0.4% of Florida residents were discharged from a hospital in Georgia, which is one of the southern state's only two immediate neighbours.

Standardized coding

The pre-processing of state datasets ensures consistency of variables across the SIDs. All SIDs used ICD-9-CM diagnostic and procedure codes from their inception until 31st September 2015. When the US transitioned to ICD-10-CM diagnostic and ICD-10-PCS procedure codes on 1st October 2015, all subsequent datasets SIDs used ICD-10. The 2015 SID files are split into two separate coding profiles representing quarters 1-3 (January-September) as ICD-9 and quarter 4 (October-December) as ICD-10¹¹.

Sensitivity to admission thresholds and other changes over time

As only admitted patients are included, the SIDs are likely to be sensitive to changes in admission thresholds.³⁹ For example, a trend towards treating haemodynamically stable patients with pulmonary embolism (PE) on an ambulatory pathway⁴⁰ would manifest as fewer PE cases in the SID over time. Concurrent use of SEDD is one means of quantifying such changes. Other disorders (such as hip fracture) are always likely to result in hospitalization and so should be unaffected by admission thresholds. Changing practice patterns have also been known to yield similar effects, as has been reported for perforated appendicitis admissions.²⁸ Similarly, changes in surgical care have moved a

number of common operations (e.g. appendectomy, cholecystectomy) from the inpatient to the outpatient setting.⁴¹

Missing data

The levels of missing data in the SIDs are usually much lower than for other large administrative databases; although, some variables are poorly coded. One report using the Michigan SID noted that missingness was <1% for most data elements. However, a small number of variables had high levels of missingness (e.g. total charges 19.8% missing, race 17.2%, and physician ID 4.5%).⁴² The HCUP website and instructional literature for each SID often indicate why values are missing in these cases (e.g. some hospitals intentionally do not report race) and can help to determine how missing data should be handled in analyses.

Data Resource Access

Data releases are available on application to the HCUP Central Distributor (<https://www.distributor.hcup-us.ahrq.gov>). Applicants are required to complete online “data use agreement training,” which takes approximately 15 minutes. Applicants must then complete a number of agreements online and provide a “Statement of Intended Use.” Once the application has been approved and payment made, the data files are made available for download using a separately emailed protected password. Historically, some state data including SID had to be sent by post. The cost of data files varies by state and year with specific information available online from the HCUP Central Distributor. Prices typically vary depending on whether or not the applicant is a student, has an educational or not-for-profit affiliation, or represents a commercial interest. For example, the Arizona SID costs \$35-50 per year for students, while the Utah SID can cost upward of \$3,200 per year for private organizations. The process from application to download of data can be completed within a few days but could require slightly longer if individual states require additional proposal review.

There are not any geographical limitations on the data and so it is not necessary for applicants to be working in a specific state for which they request data or even to be working in the US.

Figure legends

Figure 1: Map of the United States showing states with State Inpatient Databases and those with 2016 re-visit variables that permit longitudinal linkage of patients between hospital admissions. *The California SID has paused after the 2011 data release, although additional years may become available in future.* Created with <https://mapchart.net>.

Figure 2: PubMed-indexed peer-reviewed research publications using data from HCUP State Inpatient Databases from January 2000-December 2018.

Conflicts of Interest

Dr Haut reports support by contracts CE-12-11-4489, DI-1603-34596, and PCS-1511-32745 from the Patient-Centered Outcomes Research Institute (PCORI), by grant 1R01HS024547 from the Agency for Healthcare Research and Quality (AHRQ) and grant R21HL129028 from the National Heart, Lung and Blood Institute of the National Institutes of Health (NIH). Dr Haut reports receiving support by contract W81XWH-16-R-BAA1 from the Department of Defense. Dr. Haut receives book royalties from Lippincott, Williams & Wilkins for *Avoiding Common ICU Errors*. Dr. Haut received compensation for authoring a paper commissioned by the National Academies of Medicine titled “Military Trauma Care’s Learning Health System: The Importance of Data Driven Decision Making.”. The corresponding author confirms on behalf of all other co-authors that there have been no involvements that might raise the question of bias in the work reported or in the conclusions, implications, or opinions stated.

References

1. Field M, Lohr K. Appendix A: The Omnibus Budget Reconciliation Act of 1989. *Institute of Medicine (US) Committee to Advise the Public Health Service on Clinical Practice Guidelines*. Washington DC, USA: National Academies Press; 1990.
2. Healthcare Cost and Utilization Project (HCUP). *SID overview*. 2018 [cited 2018 27th November]; Available from: <https://www.hcup-us.ahrq.gov/sidoverview.jsp>
3. American Hospital Association (AHA). *AHA Annual Survey Database*. 2018 [cited; Available from: <https://www.ahadataviewer.com/additional-data-products/AHA-Survey/>
4. Gibson T, Casto A, Young J, Karnell L, Coenen N. *Impact of ICD-10-CM/PCS on research using administrative databases*; 2016 25th July 2016.
5. Elixhauser A, Steiner C, Palmer L. *Clinical Classifications Software (CCS) 2014*; 2014.
6. Lee CC, Kimmell KT, Lalonde A, et al. Geographic variation in cost of care for pituitary tumor surgery. *Pituitary* 2016; **19**: 515-21.
7. Goto T, Faridi MK, Gibo K, et al. Trends in 30-day readmission rates after COPD hospitalization, 2006-2012. *Respir Med* 2017; **130**: 92-7.
8. Metcalfe D, Davis WA, Olufajo OA, et al. Access to post-discharge inpatient care after lower limb trauma. *J Surg Res* 2016; **203**: 140-4.
9. Metcalfe D, Olufajo O, Rios-Diaz AJ, et al. Are appendectomy outcomes in level I trauma centers as good as we think? *J Surg Res* 2016; **202**: 239-45.
10. Olufajo OA, Metcalfe D, Rios-Diaz A, et al. Does Hospital Experience Rather than Volume Improve Outcomes in Geriatric Trauma Patients? *J Am Coll Surg* 2016; **223**: 32-40 e1.
11. Barrett M. *Healthcare Cost and Utilization Project (HCUP) State Databases: change in structure and data elements caused by transition to ICD-10-CM/PCS*; 2015.
12. Stagg V. CHARLSON: Stata module to calculate Charlson index of comorbidity. *Statistical Software Components S456719*. Boston, MA, USA: Boston College Department of Economics; 2006.
13. Clark DE, Osler TM, Hahn DR. ICDPIC: Stata module to provide methods for translating International Classification of Diseases (Ninth Revision) diagnosis codes into standard injury categories and/or scores. *Statistical Software Components S457028*. Boston, MA, USA: Boston College Department of Economics; 2010.
14. Metcalfe D, Salim A, Olufajo O, et al. Hospital case volume and outcomes for proximal femoral fractures in the USA: an observational study. *BMJ Open* 2016; **6**: e010743.
15. Olufajo OA, Cooper Z, Yorkgitis BK, et al. The truth about trauma readmissions. *Am J Surg* 2016; **211**: 649-55.
16. Centers for Medicare & Medicaid Services. *MS-DRG Classifications and Software*. 2018 [cited 2018 10th October]; Available from: <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/MS-DRG-Classifications-and-Software.html>
17. Zogg CK, Scott JW, Bhulani N, et al. Impact of Affordable Care Act Insurance Expansion on Pre-Hospital Access to Care: Changes in Adult Perforated Appendix Admission Rates after Medicaid Expansion and the Dependent Coverage Provision. *J Am Coll Surg* 2018.
18. Haut ER, Pronovost PJ. Surveillance bias in outcomes reporting. *JAMA* 2011; **305**: 2462-3.
19. Haut ER, Pronovost PJ, Schneider EB. Limitations of administrative databases. *JAMA* 2012; **307**: 2589; author reply -90.

20. Guglielminotti J, Deneux-Tharaux C, Wong CA, Li G. Hospital-Level Factors Associated with Anesthesia-Related Adverse Events in Cesarean Deliveries, New York State, 2009-2011. *Anesth Analg* 2016; **122**: 1947-56.
21. Hechenbleikner EM, Zheng C, Lawrence S, et al. Do hospital factors impact readmissions and mortality after colorectal resections at minority-serving hospitals? *Surgery* 2017; **161**: 846-54.
22. Won RP, Friedlander S, Lee SL. Outcomes and Costs of Managing Appendicitis at Safety-Net Hospitals. *JAMA Surg* 2017; **152**: 1001-6.
23. Blanco BA, Kothari AN, Blackwell RH, et al. "Take the Volume Pledge" may result in disparity in access to care. *Surgery* 2017; **161**: 837-45.
24. Cobb AN, Erickson TR, Kothari AN, et al. Commercial quality "awards" are not a strong indicator of quality surgical care. *Surgery* 2018; **164**: 379-86.
25. Thompson MP, Kaplan CM, Cao Y, Bazzoli GJ, Waters TM. Reliability of 30-Day Readmission Measures Used in the Hospital Readmission Reduction Program. *Health Serv Res* 2016; **51**: 2095-114.
26. Udyavar R, Cornwell EE, Havens JM, et al. Surgeon-driven variability in emergency general surgery outcomes: Does it matter who is on call? *Surgery* 2018.
27. Thirukumaran CP, Glance LG, Rosenthal MB, et al. Impact of Medicare's Nonpayment Program on Venous Thromboembolism Following Hip and Knee Replacements. *Health Serv Res* 2018.
28. Zogg CK, Pawlik TM, Haider AH. Three Common Methodological Issues in Studies of Surgical Readmission Rates: The Trouble With Readmissions. *JAMA Surg* 2018.
29. Healthcare Cost and Utilization Project (HCUP). *HCUP Methods Series*. 2018 [cited 2018 10th October]; Available from: <https://www.hcup-us.ahrq.gov/reports/methods/methods.jsp>
30. Haider AH, Bilimoria KY, Kibbe MR. A Checklist to Elevate the Science of Surgical Database Research. *JAMA Surg* 2018; **153**: 505-7.
31. Khera R, Angraal S, Couch T, et al. Adherence to Methodological Standards in Research Using the National Inpatient Sample. *JAMA* 2017; **318**: 2011-8.
32. Khera R, Krumholz HM. With Great Power Comes Great Responsibility: Big Data Research From the National Inpatient Sample. *Circ Cardiovasc Qual Outcomes* 2017; **10**.
33. Stulberg JJ, Haut ER. Practical Guide to Surgical Data Sets: Healthcare Cost and Utilization Project National Inpatient Sample (NIS). *JAMA Surg* 2018; **153**: 586-7.
34. Healthcare Cost and Utilization Project (HCUP). *HCUP Central Distributor SID Number of Discharges by Year*. 2018 [cited 2018 9th October]; Available from: https://www.hcup-us.ahrq.gov/db/state/siddist/siddist_discharge.jsp
35. Pouw ME, Peelen LM, Moons KG, Kalkman CJ, Lingsma HF. Including post-discharge mortality in calculation of hospital standardised mortality ratios: retrospective analysis of hospital episode statistics. *BMJ* 2013; **347**: f5913.
36. Metcalfe D, Gabbe BJ, Perry DC, et al. Quality of care for patients with a fracture of the hip in major trauma centres: a national observational study. *Bone Joint J* 2016; **98-B**: 414-9.
37. Borzecki AM, Christiansen CL, Chew P, Loveland S, Rosen AK. Comparison of in-hospital versus 30-day mortality assessments for selected medical conditions. *Med Care* 2010; **48**: 1117-21.

38. National Association of Health Data Organizations (NAHDO). *Next steps for the inter-state exchange of non-resident data between state health data organizations*. Salt Lake City, UT, USA; 2009.
39. Herbert A, Wijlaars L, Zylbersztejn A, Cromwell D, Hardelid P. Data Resource Profile: Hospital Episode Statistics Admitted Patient Care (HES APC). *Int J Epidemiol* 2017; **46**: 1093-i.
40. Kovacs MJ, Hawel JD, Rekman JF, Lazo-Langner A. Ambulatory management of pulmonary embolism: a pragmatic evaluation. *J Thromb Haemost* 2010; **8**: 2406-11.
41. Tang H, Dong A, Yan L. Day surgery versus overnight stay laparoscopic cholecystectomy: A systematic review and meta-analysis. *Dig Liver Dis* 2015; **47**: 556-61.
42. Houchens R. *Missing data methods for the NIS and the SID*: US Agency for Healthcare Research and Quality (AHRQ); 2015.
43. United States Census Bureau. *Annual Estimates of the Resident Population for the United States, Regions, States, and Puerto Rico: April 1, 2010 to July 1, 2017 (NST-EST2017-01)*. State population totals and components of change: 2010-2017 2018 [cited 2018 9th October]; Available from: <https://www.census.gov/data/tables/2017/demo/popest/state-total.html>

Figure 1

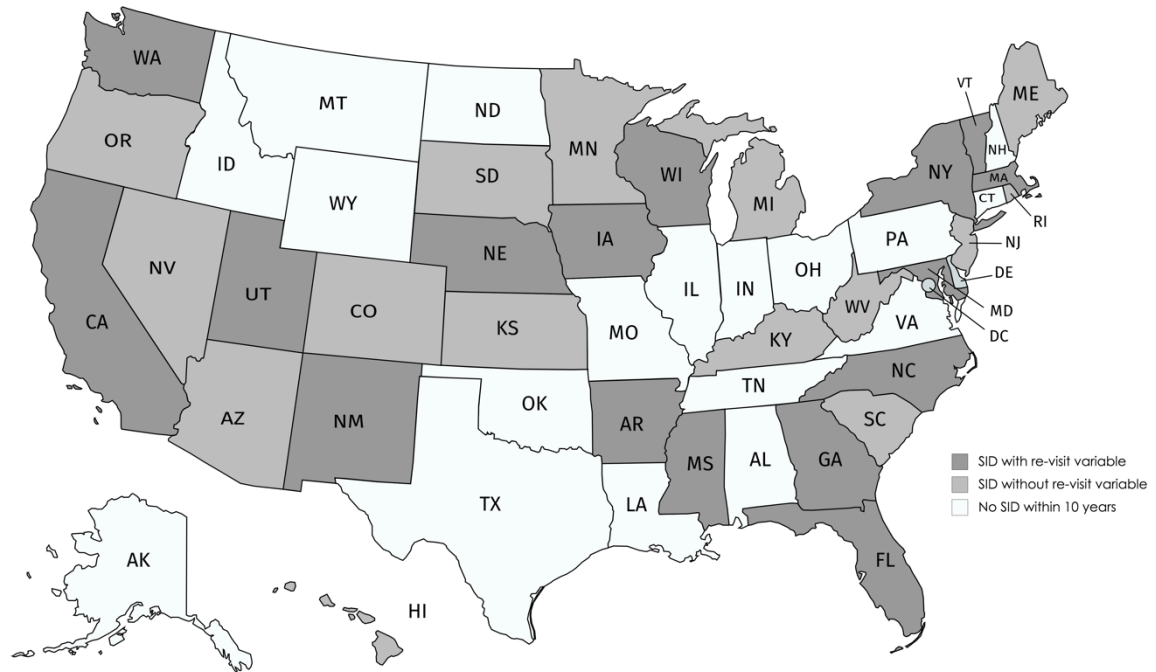


Figure 2

