



Published in final edited form as:

*Crit Care Med.* 2013 March ; 41(3): 886–896. doi:10.1097/CCM.0b013e31827bfc3c.

## Using existing data to address important clinical questions in critical care

Colin R. Cooke, MD, MSc, MS<sup>1,2,3</sup> and Theodore J. Iwashyna, MD, PhD<sup>1,2,4</sup>

<sup>1</sup>Department of Medicine, Division of Pulmonary & Critical Care Medicine, University of Michigan, Ann Arbor, MI

<sup>2</sup>Institute for Healthcare Policy and Innovation, University of Michigan, Ann Arbor, MI

<sup>3</sup>Center for Healthcare Outcomes & Policy, University of Michigan, Ann Arbor, MI

<sup>4</sup>VA Center for Clinical Management Research, Ann Arbor, MI

### Abstract

**Objective**—With important technological advances in healthcare delivery and the internet, clinicians and scientists now have access to overwhelming number of available databases capturing patients with critical illness. Yet investigators seeking to answer important clinical or research questions with existing data have few resources that adequately describe the available sources and the strengths and limitations of each. This article reviews an approach to selecting a database to address health services and outcomes research questions in critical care, examines several databases that are commonly used for this purpose, and briefly describes some strengths and limitations of each.

**Data Sources**—Narrative review of the medical literature.

**Summary**—The available databases that collect information on critically ill patients are numerous and vary in the types of questions they can optimally answer. Selection of a data source must not only consider accessibility, but also the quality of the data contained within the database, and the extent to which it captures the necessary variables for the research question. Questions seeking causal associations (e.g. effect of treatment on mortality) usually either require secondary data that contain detailed information about demographics, laboratories, and physiology to best address non-random selection or sophisticated study design. Purely descriptive questions (e.g. incidence of respiratory failure) can often be addressed using secondary data with less detail such as administrative claims. Though each database has its own inherent limitations, all secondary analyses will be subject to the same challenges of appropriate study design and good observational research.

**Conclusion**—The literature demonstrates that secondary analyses can have significant impact on critical care practice. While selection of the optimal database for a particular question is a necessary part of high-quality analyses, it is not sufficient to guarantee an unbiased study. Thoughtful and well-constructed study design and analysis approaches remain equally important pillars of robust science. Only through responsible use of existing data will investigators ensure that their study has the greatest impact on critical care practice and outcomes.

---

Corresponding author: Colin R. Cooke, MD, MSc, MS, 2800 Plymouth Rd, Building 16, Room 127W, Ann Arbor, MI, 48109, Ph: 734-615-9681, Fax: 734-998-7473, cookecr@umich.edu.

The authors have not disclosed any potential conflicts of interest

## Keywords

Epidemiology; intensive care; outcomes research; health services; health policy; statistical data analysis; factual databases; medical records systems

---

## Introduction

Clinicians and scientists have witnessed an unprecedented expansion in the publication of critical care studies employing observational designs. This expansion is perhaps most evident in studies using secondary data. *Secondary* data can be defined as data gathered for one reason (e.g. clinical trial) now being reemployed to answer a novel question (e.g. costs of care), whereas *primary* data are data collected specifically for the purposes of answering a novel question (1). Several factors are responsible for this expansion, including wide recognition of the importance of safety and quality improvement research (2-5), the ability to perform complex analytic tasks on personal computers, and growth in the cadre of investigators capable of performing secondary analyses through methods training(6, 7).

When done well, this work has contributed novel observations and changed practice in fundamental ways, improving outcomes for patients. At the bedside, the re-evaluation of pulmonary artery catheters—once a ubiquitous feature of nearly every medical ICU patient—began with a clever re-analysis of a clinical trial done to investigate other questions(8). Modern ICU organization—with its focus on high-volume centers of excellence—was shaped by scientific observations about the volume/outcome relationship(9, 10) and by the rigorous evaluation of the surgical and trauma center experience(11-13).

Secondary data analyses also provide essential raw material for key operations in healthcare. National priority setting about causes of death and clinical decision-making about prior probabilities of disease both depend on secondary data. For example, virtually every basic-science grant application on severe sepsis contextualizes the proposed work with national-scale epidemiology derived from administrative records(14, 15). Current policy concerns about healthcare overuse in the ICU such as excessive end-of-life spending and unexplained geographical variation in ICU use depend on secondary data analyses(16-20). Much of our understanding of racial/ethnic and insurance-based disparities(21-28), as well as the value of critical care(29-32), derive from secondary data analyses. This work has helped move the conversation about the causes for inequities in critical care away from personal opinion toward scientific evaluation and efforts to solve such problems.

This new scientific and clinical importance of secondary data analysis has regrettably been accompanied by numerous examples of poorly designed studies utilizing datasets ill equipped to answer the research questions posed of them(33). A major contributor to the evolution of secondary analyses is the dramatic growth in existing critical care databases and the ease with which one can access them. Despite the attractiveness of such data for many purposes, there have been few references to turn to that discuss available data sources relevant to critical care to facilitate informed choices by prospective investigators(1, 7, 34-36).

In this review, we examine several existing critical care data sources commonly used for secondary data analysis in critical care, and present a practical approach to the selection of a database based upon the strengths of the source. We limit our discussion to databases used for the conduct of clinical epidemiology, health service, policy and outcomes research, rather than discuss data derived from genetic analyses, “-omics”, or other bench science. Although there are important critical care data resources outside of the United States(37-39), we focus on resources within. Our target audience includes both investigators seeking to

answer research questions in critical care, but also readers of the medical literature interested in ways to better appraise the data sources selected in published studies. Finally, we focus on the secondary data available for answering well-formed questions, but do not seek to review fundamentals of good scientific study design.

## Why use secondary data?

Investigators who employ secondary data to answer clinical questions can capitalize on several of its advantages compared to primary data. First, secondary analysis of data promotes efficient use of research investments particularly when performed on biologic data or data otherwise overly expensive to collect. Second, there are some questions (e.g. where randomization is unethical or measuring “real-world” practice) that often can *only* or are *most efficiently* answered by secondary data analysis. Third, some secondary data, such as large registries and administrative data, may provide greater generalizability due to a much greater scope than studies collecting primary data—potentially generalizing to regions, states, or even the nation. Fourth, it may be feasible to carry out a secondary data analysis for questions where primary data collection is too onerous, such as those that consider 5- or 10-year follow-ups. Similarly, scientists with appropriate statistical training but with limited grant funding may find secondary data analysis more feasible as a first approach to a problem. Fifth, because some secondary analyses employ administrative data that are very large, they may provide a much more precise estimate of effect than smaller primary studies, particularly for rare diagnoses. Sixth, when secondary data cannot be used to answer the appropriate scientifically rigorous and clinically relevant question, they may play an important supplementary role to assess the plausibility and likelihood of success of a large-scale primary data collection effort; secondary data analysis may be a particularly cost-effective way to obtain such preliminary data. For example, the secondary analysis of the SUPPORT study, which suggested pulmonary artery catheters were harmful, formed the basis for several randomized trials. Finally, secondary analyses of administrative data may be more relevant to policymakers and therefore support the translation of scientific discoveries into improved care. For example, Medicare stakeholders base policy decision on research that is conducted using data from Medicare patients.

## Selecting a data source

The issues that a researcher confronts when evaluating a database for secondary analysis are the same that readers of the literature must consider when assessing the quality of a data source used in a published study. However, because investigators ultimately need to obtain the data in addition to critiquing it, we approach the database evaluation and acquisition process from the perspective of a researcher.

## Data quality

When confronted with the overwhelming number of existing data sources, investigators must first consider the ability of a given data source to sufficiently address the research question of interest. This involves characterizing the quality and overall susceptibility to bias of the data source. While this process largely remains a subjective task(40, 41), in 2003 a group of investigators from the United Kingdom developed a framework to assess the quality of secondary databases(42). The framework included two aspects characterizing database quality: *coverage* and *accuracy*(Table 1). While useful in principle, this framework is not sufficient to identify an optimal database. By placing equal value on each aspect of database quality, it ignores how the potential for inadequacies in a single domain may be fatal to a study. Most obviously, a database may be perfect in all domains yet lack the outcome variable of interest.

## Taxonomy of the question

A more practical approach to the selection of a data source considers the needs imposed by the research question. Through articulating a well-defined research question, an investigator will know which variables are needed to define the population, the treatment or exposure, the outcomes, and those needed for adjustment (i.e. confounding variables), such as demographics and severity of illness(43). One can then consider whether clinically detailed versus clinically sparse data are needed to address the question. Many quantitative research questions can be lumped into five overlapping groups that describe the goals of the study (Table 2). The goals of the study often dictate the need for clinically rich versus sparse information. When investigators seek to determine the causal relationship between a risk factor and outcome or compare outcomes across specific treatments, greater clinical information is usually necessary to address confounding – that is, account for any variable that may distort the relationship between the observed exposure and outcome(44). Often, the most important confounding variable is severity of illness(7, 44).

In contrast, more descriptive questions that characterize the epidemiology of disease, clinical practice, health service use, and health care spending, are not usually limited by confounding because comparisons between different groups are less often performed. For example, Wiener and colleagues used national hospital discharge claims to examine temporal changes in use of pulmonary artery catheters(45). The absence of clinically detailed information in this study did not impact the importance of this result given its descriptive nature.

While the taxonomy of questions in Table 2 may be helpful to articulating the needs for a given question, we do not intend it to be prescriptive or without exception. For example, there are many excellent examples of investigators employing clinically sparse data to examine causal relationships(10, 46) and clinically rich information to examine disease incidence(47). Clinically rich information often has the benefit of being able to address many types of questions, but the same is not necessarily true for clinically sparse information. When sparse data are used to examine causal relationships more sophisticated methods to address patient selection and confounding are needed(44, 48-51).

## Mapping the question to the data

Only after an investigator has decided on the variables needed to address the question and has considered the importance of clinically rich versus sparse data, should he or she investigate the available databases in which to study the question. A prudent approach toward finding a source requires searching of the web, the literature, and discussions with investigators who have used secondary data. Once potential sources are identified, an important next step is to determine if access to the data is feasible. Some secondary data sources can be downloaded from the web free of charge, but others have fees that can range from \$20 to over \$100,000. Independent of access fees, some sources require navigating administrative hurdles, including vetting the research proposal by an oversight committee, or require collaboration with a scientist that has access to the data. While in no way comprehensive, Table 3 describes the several critical care databases organized by the degree of clinical detail available within each and qualitatively describes the accessibility of each data source.

## Available data sources

### Publicly available clinical trials and cohort studies

Nationally funded randomized controlled trials (RCT) and large-scale prospective cohort studies usually collect data with considerable clinical detail, including clinical physiology,

severity of illness, and patient outcomes. One of the most important existing repositories for critical care RCTs and cohort studies is the National Heart, Lung, and Blood Institute's Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC)(52). BioLINCC provides data from over 80 clinical and epidemiological studies. These include many prominent critical care studies from the last few decades, particularly the ARDSNet RCTs conducted from 1996 through 2006(53-58)—use of which has resulted in dozens of secondary analyses.

### **Electronic medical record**

The electronic medical record (EMR) has great promise to become the future source of many secondary data analyses(59, 60). Unfortunately, several important barriers hamper the current realization of the research potential of the EMR, including difficulty in extracting information from free text, and compatibility of systems across hospitals(60). Nevertheless, there have been successes.

The Department of Veterans Affairs created the Inpatient Evaluation Center (IPEC), an infrastructure for improving the quality of care in VA medical centers that includes data on all inpatients in over 100 hospitals extracted from the VA's EMR. This data source includes an excellent risk-adjustment measure and has been used to study the organization and quality of care within the VA(61, 62), develop risk-adjustment models(63, 64), and determine the impact of infection control measures on outcomes(65, 66). Kaiser Permanente of Northern California has similarly rich data on its large network of community hospitals(67).

An additional EMR-based resource is the Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II) database(68). This publicly available, deidentified repository includes minute-to-minute data for over 30,000 patients admitted to an ICU in Beth Israel Deaconess Medical Center. Published studies using the MIMIC-II database have examined several aspects of ICU care such as, developing and validating high-fidelity risk-adjustment models(69, 70), and characterizing providers' response to hypotension(71, 72). Users can gain access to MIMIC-II via the internet.

### **Quality Improvement and Benchmarking**

Several existing data sources that were created for benchmarking or quality improvement purposes provide clinically rich data on ICU patients. Perhaps the most famous of these is the APACHE database(73). By maintaining the gold standard for risk-adjustment, the APACHE database provides rich clinical information for patients in hospitals that voluntarily contribute data. Investigators have used this source to answer questions about the impact of organizational features on patient outcomes(74, 75), variation in ICU admission practice(76), volume-outcome relationships(9), among others. Cerner, the owner of APACHE, also previously maintained the now unavailable Project IMPACT(77, 78).

A relatively new data source of critically ill patients is the eICU Research Institute(79). Although designed to allow off-site intensivist involvement in remote ICUs, telemedicine systems also standardize disparate data from participating ICUs(80). Phillips eICU (formerly VISICU), currently the largest vendor of ICU telemedicine, created the eICU Research Institute in collaboration with health-care providers and academia(80, 81). The University of Maryland School of Pharmacy Pharmaceutical Research Computing Center (PRC) is the first academic partner with access to this database.

## Registries

Although the line distinguishing a registry from quality improvement or benchmarking database is somewhat arbitrary, registries usually focus on a single disease or syndrome and are often used by their participants to benchmark their own data to that of others. For example, the National Trauma Data Bank maintains the largest nationally representative sample of patients experiencing trauma. Data fields include demographics, vital signs from the ED and EMS, abbreviated injury scale, procedure and diagnosis codes, ICU and ventilator days, among other characteristics. Prominent past studies employing this data source have looked at the impact of helicopter transport(82), the pulmonary artery catheter(83), and prehospital fluids(84) on outcomes of trauma.

The American Heart Association Get with the Guidelines (GWTG) maintains several registries capturing patients that often require critical care services. The GWTG-Resuscitation collects information on consecutive patients with in-hospital cardiac arrests, defined by the absence of a central palpable pulse, apnea, and unresponsiveness(85). Extensive data surrounding the arrest and the post hospital course is collected including outcomes of return to spontaneous circulation, neurologic status, and survival to discharge. Recent studies employing this dataset include analyses examining cardiac arrest among patients with pneumonia(86), variation in hospital cardiac arrest rates(87, 88) and in the time to defibrillation(89), and racial differences in outcomes after arrest(90).

## Administrative data/utilization claims

Administrative data are data collected on patients during encounters with the healthcare system and are most often collected for billing insurers. For hospitalized patients, this usually includes data from the Uniform Billing 04 sheet (UB-04) which collects facility charges during an inpatient stay(91). Although elements vary by payer, this form typically collects demographics including payer, admission source, ICD-9-CM diagnosis and procedure codes, DRG codes, some CPT and/or HCPCS codes, length of stay, disposition, hospital identifier, and detailed charges for aspects of the hospital stay (e.g. ICU room and board, pharmacy). Although encounter-specific, claims can often be linked allowing one to trace an individual's course through inpatient, post-discharge, and outpatient facilities.

The two main sources of administrative data include insurers and government agencies interested in tracking healthcare use. For example, Medicare provides research claims for all aspects of care among its close to 50 million beneficiaries across the entire United States, a segment of the population that accounts for a majority of critical care use(92) and of intrinsic public policy interest. Long-term mortality and longitudinal utilization can be tracked in Medicare files. Access to Medicare data is relatively expensive if one's research question requires individual-level linkage across hospitals or outpatient claims; in contrast, one year's standard inpatient file, so-called "MedPAR" files, can cost less than \$1,000. MedPAR includes information about the inpatient stay typically present on the UB-04 form, including diagnosis, procedure, and DRG codes, ICU or CCU stay, hospital charges, and hospital discharge disposition. Data about skilled nursing stays are also included. However, information about outpatient visits, physician charges, durable medical equipment, and hospice care are in separate files. Investigators have used Medicare data to examine long term survival of respiratory failure(93), epidemiology of sepsis(94), cognitive outcomes among critically ill patients(95), and epidemiology of long-term acute care use(96).

In contrast to Medicare, some data sources include data from all payers. These include various state health departments or agencies such as the CDC that maintain national surveys of inpatient care(97). The Healthcare Cost and Utilization Project (HCUP), the largest collection of all-payer inpatient care data in the US maintains one of the most accessible

sources of administrative data(98). Investigators can access over 95% of ER visits and hospital discharges from individual states using HCUP's State Emergency Database or the State Inpatient Database, or use HCUP's Nationwide Inpatient Database to examine questions in a nationally representative sample of hospitals and patients. Data are also available for children. Readmissions are tracked in several states; however, follow-up to out-of-hospital patient-centered outcomes is often impossible. HCUP has been used to examine variation in ICU use(19, 99), stroke risk among patients with atrial fibrillation and sepsis(100), longitudinal trends in PA catheter use(45), and impact of marital status on sepsis outcomes(101). Finally, private groups or insurers also maintain research files that can be purchased at significant costs. These include MarketScan, a data source representing diverse claims from over 100 private payers(102), and Premier Perspective, the nations largest inpatient drug utilization database. Premier Perspective is unique in its collection of time stamped data about medications delivered during an inpatient stay. Investigators have capitalized on this unique attribute to examine the impact of activated protein C on mortality in sepsis(103), and the quality of care among patients admitted with COPD exacerbations(104, 105).

## Linking data sources

Often a single dataset may provide only part of the information that is necessary to conduct a successful analysis. In such situations investigators can either supplement the data source by collecting additional data or link two or more existing data sources. For example, Treggiari and colleagues successfully linked an existing ARDS database to a prospective survey of ICU directors to determine the relationship between physician staffing and outcomes of ARDS(106).

The often-easier option involves the linking of two independent but preexisting data sources that together have the necessary information for the question. Occasionally, this linkage has already been done prior to obtaining the data. For example, the Health and Retirement Study collects data on the sources beneficiaries use to pay for services, health status, and other economic and family variables from nationally representative samples of older Americans. An existing link to Medicare files allows one to identify survey respondents who were hospitalized with critical illness. Iwashyna et al.(95) used this data, and Barnato et al.(107) capitalized on the similar Medicare Current Beneficiary Study to examine disability of long-term survivors of critical illness. Such linkage offers an unusual opportunity to examine outcomes for rarer diseases with prospectively collected pre-morbid data(108). An additional pre-linked data source is the Surveillance Epidemiology and End Results linked to Medicare files (SEER-Medicare)(109). SEER collects information on cancer incidence, prevalence and survival from specific geographic areas containing 28 percent of the US population. Through an existing link to Medicare inpatient files, one can examine the intersection between cancer and critical care, such as the relationship between critical illness and long-term survival among lung cancer patients(110). When links are not already in place, investigators can often establish them provided identifying information is present within the data. For example, Seymour and colleagues linked paramedic run sheets with WA state hospital discharge claims to study prehospital risk factors for ICU admission and hospital mortality(111).

## Limitations of available databases

While many databases described above are limited by their lack of clinical detail, all have additional unique limitations. Most RCTs in critical care enroll only a small fraction of eligible patients, which may threaten generalizability of secondary analyses using RCTs as a data source. Many registries or databases collected for quality improvement and

benchmarking efforts include only volunteer hospitals. These non-random hospitals may be highly motivated to improve care for their patients, are often geographically clustered, and are more likely to be teaching hospitals, factors that threaten generalizability of studies employing these data sources(7). Administrative data are limited by the often-unknown validity of ICD-9-CM or other billing codes for identifying critical illness, variable number of ICD-9-CM codes collected across hospitals(112), temporal instability in coding practice(113), biases due to provider efforts to maximize payment(114), among others. Investigators using these databases should consider how these limitations might bias their results and include strategies in their analyses that address these weaknesses. These limitations suggest that the optimal approach to an avenue of research uses secondary data for the questions that they are uniquely suited to address, but turns to primary data for other aspects of the key clinical questions.

Finally, and perhaps most crucially, databases only provide the raw materials to address a research question but do nothing to ensure a study is appropriately designed and conducted. Observational research—indeed, all research—regardless of the study design or data source, is subject to a variety of biases in addition to the issues of confounding.

## Speculations about the future

A major barrier to optimal care for all critically ill patients is absence of a centralized repository of data on critically ill patients in the US—despite the fact that such a barrier is surmountable. Policymakers and scientists have used available registries of patients with cardiac disease, including those described above, not only to increase guideline concordant care, but also to gain important insights about the care for patients with congestive heart failure, myocardial infarction, and cardiac arrest(87, 115-117). Leaders within the American College of Surgeons have driven continuing improvements in trauma and surgical outcomes through trauma registries and the National Surgical Quality Improvement Program registry(118, 119). Registries have even been successfully implemented in combat zones to improve care of wounded soldiers(120). Armed with high-quality clinical practice guidelines, policymakers in cardiology and surgery have developed clinical registries by capitalizing on strong leadership and financing from professional societies. Despite the existence of guidelines for management of some critically ill populations, such as sepsis, leaders in critical care have been less successful in their efforts to create comparably accessible and comprehensive registries(121). As leaders within professional critical care societies strive to guide practice through publication and implementation of guidelines, they should continue to pursue parallel efforts to track the populations targeted by their guidelines to ensure that optimal care is being delivered in the real world. As we have witnessed in cardiology and surgery, secondary analyses of such critical care registries could realize further gains in the care for our patients.

## Conclusions

Through secondary data analyses, investigators have provided a large contribution to the understanding of disease and health care delivery in critical care. This past work is an important reminder that rigorous observational science is not only possible it is essential to further improve the care delivered to critically ill patients. Scientists using existing data for research also promote a more efficient research agenda because they maximize the knowledge that can be gained from the past, often expensive efforts to gather data(122). Investigators using secondary data must carefully consider the advantages and disadvantages of each potential data source prior to selecting one or more for their analyses. Through applying a rigorous approach to database selection and data quality assessment, investigators will be well on their way ensure that their study will have the greatest impact.



## Acknowledgments

**Funding:** Support for this work was provided in part by a grant from the Agency For Healthcare Research and Quality (K08 HS020672, Dr. Cooke), the National Heart, Lung, and Blood Institute (K08 HL091249, Dr. Iwashyna), and U.S. Department of Veterans Affairs Health Services Research & Development Services (IIR 11-109, Dr. Iwashyna). The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the US government.

## References

1. Wunsch H, Harrison DA, Rowan K. Health services research in critical care using administrative data. *J Crit Care.* 2005; 20(3):264–269. [PubMed: 16253796]
2. Kohn, KT.; Corrigan, JM.; Donaldson, MS., editors. *To Err Is Human: Building a Safer Health System.* Washington, DC: Committee on Quality of Health Care in America, Institute of Medicine, National Academy Press; 1999.
3. Chassin MR, Galvin RW. The urgent need to improve health care quality. Institute of Medicine National Roundtable on Health Care Quality. *Jama.* 1998; 280(11):1000–1005. [PubMed: 9749483]
4. Pronovost P, Wu AW, Dorman T, et al. Building safety into ICU care. *J Crit Care.* 2002; 17(2):78–85. [PubMed: 12096370]
5. Leape LL, Berwick DM. Five years after To Err Is Human: what have we learned? *Jama.* 2005; 293(19):2384–2390. [PubMed: 15900009]
6. Curtis JR, Rubenfeld GD, Hudson LD. Training pulmonary and critical care physicians in outcomes research: should we take the challenge? *Am J Respir Crit Care Med.* 1998; 157(4 Pt 1):1012–1015. [PubMed: 9563712]
7. Rubenfeld GD, Angus DC, Pinsky MR, et al. Outcomes research in critical care: results of the American Thoracic Society Critical Care Assembly Workshop on Outcomes Research. The Members of the Outcomes Research Workshop. *Am J Respir Crit Care Med.* 1999; 160(1):358–367. [PubMed: 10390426]
8. Connors AF Jr, Speroff T, Dawson NV, et al. The effectiveness of right heart catheterization in the initial care of critically ill patients. SUPPORT Investigators. *Jama.* 1996; 276(11):889–897. [PubMed: 8782638]
9. Kahn JM, Goss CH, Heagerty PJ, et al. Hospital volume and the outcomes of mechanical ventilation. *N Engl J Med.* 2006; 355(1):41–50. [PubMed: 16822995]
10. Kahn JM, Ten Have TR, Iwashyna TJ. The relationship between hospital volume and mortality in mechanical ventilation: an instrumental variable analysis. *Health Serv Res.* 2009; 44(3):862–879. [PubMed: 19674428]
11. Nathens AB, Jurkovich GJ, Maier RV, et al. Relationship between trauma center volume and outcomes. *Jama.* 2001; 285(9):1164–1171. [PubMed: 11231745]
12. Birkmeyer JD, Siewers AE, Finlayson EV, et al. Hospital volume and surgical mortality in the United States. *N Engl J Med.* 2002; 346(15):1128–1137. [PubMed: 11948273]
13. MacKenzie EJ, Rivara FP, Jurkovich GJ, et al. A national evaluation of the effect of trauma-center care on mortality. *N Engl J Med.* 2006; 354(4):366–378. [PubMed: 16436768]
14. Angus DC, Linde-Zwirble WT, Lidicker J, et al. Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care. *Crit Care Med.* 2001; 29(7): 1303–1310. [PubMed: 11445675]
15. Martin GS, Mannino DM, Eaton S, et al. The epidemiology of sepsis in the United States from 1979 through 2000. *N Engl J Med.* 2003; 348(16):1546–1554. [PubMed: 12700374]
16. Wennberg J, Gittelsohn. Small area variations in health care delivery. *Science.* 1973; 182(117): 1102–1108. [PubMed: 4750608]
17. Wennberg JE, Fisher ES, Stukel TA, et al. Use of hospitals, physician visits, and hospice care during last six months of life among cohorts loyal to highly respected hospitals in the United States. *BMJ.* 2004; 328(7440):607. [PubMed: 15016692]
18. Barnato AE, Herndon MB, Anthony DL, et al. Are regional variations in end-of-life care intensity explained by patient preferences?: A Study of the US Medicare Population. *Med Care.* 2007; 45(5):386–393. [PubMed: 17446824]

19. Seymour CW, Iwashyna TJ, Ehlenbach WJ, et al. Hospital-level variation in the use of intensive care. *Health Serv Res.* 2012; 47(5):2060–2080. [PubMed: 22985033]
20. Seymour CW, Kahn JM. Addressing the Growth in Intensive Care : Comment on “Intensive Care Unit Admitting Patterns in the Veterans Affairs Health Care System”. *Arch Intern Med.* 2012:7–9.
21. Rapoport J, Teres D, Steingrub J, et al. Patient characteristics and ICU organizational factors that influence frequency of pulmonary artery catheterization. *Jama.* 2000; 283(19):2559–2567. [PubMed: 10815120]
22. Smedley, BD.; Stith, AY.; Nelson, AR., et al. *Unequal treatment: confronting racial and ethnic disparities in health care.* Washington, D.C.: National Academy Press; 2003.
23. Mayr FB, Yende S, D'Angelo G, et al. Do hospitals provide lower quality of care to black patients for pneumonia? *Crit Care Med.* 2010; 38(3):759–765. [PubMed: 20009756]
24. Cooke CR, Nallamotheu B, Kahn JM, et al. Race and timeliness of transfer for revascularization in patients with acute myocardial infarction. *Med Care.* 2011; 49(7):662–667. [PubMed: 21677592]
25. Erickson SE, Vasilevskis EE, Kuzniewicz MW, et al. The effect of race and ethnicity on outcomes among patients in the intensive care unit: a comprehensive study involving socioeconomic status and resuscitation preferences. *Crit Care Med.* 2011; 39(3):429–435. [PubMed: 21187746]
26. Lyon SM, Benson NM, Cooke CR, et al. The Effect of Insurance Status on Mortality and Procedural Utilization in Critically Ill Patients. *Am J Respir Crit Care Med.* 2011
27. Cooke CR, Erickson SE, Eisner MD, et al. Trends in the incidence of noncardiogenic acute respiratory failure: the role of race. *Crit Care Med.* 2012; 40(5):1532–1538. [PubMed: 22511134]
28. Lane-Fall MB, Iwashyna TJ, Cooke CR, et al. Insurance and racial differences in long-term acute care utilization after critical illness. *Crit Care Med.* 2012; 40(4):1143–1149. [PubMed: 22020247]
29. Kahn JM, Angus DC. Reducing the cost of critical care: new challenges, new solutions. *Am J Respir Crit Care Med.* 2006; 174(11):1167–1168. [PubMed: 17110650]
30. Kahn JM, Rubenfeld GD, Rohrbach J, et al. Cost savings attributable to reductions in intensive care unit length of stay for mechanically ventilated patients. *Med Care.* 2008; 46(12):1226–1233. [PubMed: 19300312]
31. Cooke CR, Kahn JM, Watkins TR, et al. Cost-effectiveness of implementing low-tidal volume ventilation in patients with acute lung injury. *Chest.* 2009; 136(1):79–88. [PubMed: 19318673]
32. Fisher ES, Bynum JP, Skinner JS. Slowing the growth of health care costs--lessons from regional variation. *N Engl J Med.* 2009; 360(9):849–852. [PubMed: 19246356]
33. Terris DD, Litaker DG, Koroukian SM. Health state information derived from secondary databases is affected by multiple sources of bias. *J Clin Epidemiol.* 2007; 60(7):734–741. [PubMed: 17573990]
34. Black N. Developing high quality clinical databases. *Bmj.* 1997; 315(7105):381–382. [PubMed: 9277596]
35. Best AE. Secondary data bases and their use in outcomes research: a review of the area resource file and the Healthcare Cost and Utilization Project. *J Med Syst.* 1999; 23(3):175–181. [PubMed: 10554733]
36. Oinonen MJ, Sansguiri V, Smith M. National data sources on the care and outcomes of patients with community acquired pneumonia. *J Med Syst.* 2000; 24(5):267–277. [PubMed: 11103357]
37. Harrison DA, Brady AR, Rowan K. Case mix, outcome and length of stay for admissions to adult, general critical care units in England, Wales and Northern Ireland: the Intensive Care National Audit & Research Centre Case Mix Programme Database. *Crit Care.* 2004; 8(2):R99–111. [PubMed: 15025784]
38. Scales DC, Guan J, Martin CM, et al. Administrative data accurately identified intensive care unit admissions in Ontario. *J Clin Epidemiol.* 2006; 59(8):802–807. [PubMed: 16828673]
39. Garland A, Yogendran M, Olafson K, et al. The accuracy of administrative data for identifying the presence and timing of admission to intensive care units in a Canadian province. *Med Care.* 2012; 50(3):e1–6. [PubMed: 22270100]
40. Sorensen HT, Sabroe S, Olsen J. A framework for evaluation of secondary data sources for epidemiological research. *Int J Epidemiol.* 1996; 25(2):435–442. [PubMed: 9119571]

41. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol.* 2005; 58(4):323–337. [PubMed: 15862718]
42. Black N, Payne M. Directory of clinical databases: improving and promoting their use. *Qual Saf Health Care.* 2003; 12(5):348–352. [PubMed: 14532366]
43. Koepsell, TD.; Weiss, NS. *Epidemiologic methods : studying the occurrence of illness.* Oxford; New York: Oxford University Press; 2003.
44. Wunsch H, Linde-Zwirble WT, Angus DC. Methods to adjust for bias and confounding in critical care health services research involving observational data. *J Crit Care.* 2006; 21(1):1–7. [PubMed: 16616616]
45. Wiener RS, Welch HG. Trends in the use of the pulmonary artery catheter in the United States, 1993-2004. *Jama.* 2007; 298(4):423–429. [PubMed: 17652296]
46. Scales DC, Thiruchelvam D, Kiss A, et al. The effect of tracheostomy timing during critical illness on long-term survival. *Crit Care Med.* 2008; 36(9):2547–2557. [PubMed: 18679113]
47. Li G, Malinchoc M, Cartin-Ceba R, et al. Eight-year trend of acute respiratory distress syndrome: a population-based study in Olmsted County, Minnesota. *Am J Respir Crit Care Med.* 2011; 183(1):59–66. [PubMed: 20693377]
48. Shadish, WR.; Cook, TD.; Campbell, DT. *Experimental and quasi-experimental designs for generalized causal inference.* Boston, MA: Houghton Mifflin Company; 2002.
49. Hernan MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology.* 2006; 17(4):360–372. [PubMed: 16755261]
50. Austin PC. The relative ability of different propensity score methods to balance measured covariates between treated and untreated subjects in observational studies. *Med Decis Making.* 2009; 29(6):661–677. [PubMed: 19684288]
51. Austin PC, Lee DS. The concept of the marginally matched subject in propensity-score matched analyses. *Pharmacoepidemiol Drug Saf.* 2009; 18(6):469–482. [PubMed: 19319923]
52. National Heart, Lung, and Blood Institute's Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC). [cited Sept 9, 2012] Available from: <http://biolincc.nhlbi.nih.gov/home>
53. Ketoconazole for early treatment of acute lung injury and acute respiratory distress syndrome: a randomized controlled trial. The ARDS Network. *Jama.* 2000; 283(15):1995–2002. [PubMed: 10789668]
54. Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. The Acute Respiratory Distress Syndrome Network. *N Engl J Med.* 2000; 342(18):1301–1308. [PubMed: 10793162]
55. Randomized, placebo-controlled trial of lisofylline for early treatment of acute lung injury and acute respiratory distress syndrome. *Crit Care Med.* 2002; 30(1):1–6. [PubMed: 11902249]
56. Brower RG, Lanken PN, MacIntyre N, et al. Higher versus lower positive end-expiratory pressures in patients with the acute respiratory distress syndrome. *N Engl J Med.* 2004; 351(4):327–336. [PubMed: 15269312]
57. Wheeler AP, Bernard GR, Thompson BT, et al. Pulmonary-artery versus central venous catheter to guide treatment of acute lung injury. *N Engl J Med.* 2006; 354(21):2213–2224. [PubMed: 16714768]
58. Wiedemann HP, Wheeler AP, Bernard GR, et al. Comparison of two fluid-management strategies in acute lung injury. *N Engl J Med.* 2006; 354(24):2564–2575. [PubMed: 16714767]
59. Varon J, Marik PE. Clinical information systems and the electronic medical record in the intensive care unit. *Curr Opin Crit Care.* 2002; 8(6):616–624. [PubMed: 12454551]
60. Rubenfeld GD. Using computerized medical databases to measure and to improve the quality of intensive care. *J Crit Care.* 2004; 19(4):248–256. [PubMed: 15648042]
61. Render ML, Freyberg RW, Hasselbeck R, et al. Infrastructure for quality transformation: measurement and reporting in veterans administration intensive care units. *BMJ Qual Saf.* 2011; 20(6):498–507.
62. Cooke CR, Kennedy EH, Wiitala WL, et al. Despite variation in volume, Veterans Affairs hospitals show consistent outcomes among patients with non-postoperative mechanical ventilation\*. *Crit Care Med.* 2012; 40(9):2569–2575. [PubMed: 22732289]

63. Render ML, Kim HM, Welsh DE, et al. Automated intensive care unit risk adjustment: results from a National Veterans Affairs study. *Crit Care Med.* 2003; 31(6):1638–1646. [PubMed: 12794398]
64. Render ML, Deddens J, Freyberg R, et al. Veterans Affairs intensive care unit risk adjustment model: validation, updating, recalibration. *Crit Care Med.* 2008; 36(4):1031–1042. [PubMed: 18379226]
65. Jain R, Kralovic SM, Evans ME, et al. Veterans Affairs initiative to prevent methicillin-resistant *Staphylococcus aureus* infections. *N Engl J Med.* 2011; 364(15):1419–1430. [PubMed: 21488764]
66. Render ML, Hasselbeck R, Freyberg RW, et al. Reduction of central line infections in Veterans Administration intensive care units: an observational cohort using a central infrastructure to support learning and improvement. *BMJ Qual Saf.* 2011; 20(8):725–732.
67. Liu V, Turk BJ, Ragins AI, et al. An electronic SAPS3-based risk adjustment score for critical illness in an integrated healthcare system. *Crit Care Med.* 2012 in press.
68. Saeed M, Villarreal M, Reisner AT, et al. Multiparameter Intelligent Monitoring in Intensive Care II: a public-access intensive care unit database. *Crit Care Med.* 2011; 39(5):952–960. [PubMed: 21283005]
69. Celi LA, Tang RJ, Villarreal MC, et al. A Clinical Database-Driven Approach to Decision Support: Predicting Mortality Among Patients with Acute Kidney Injury. *J Healthc Eng.* 2011; 2(1):97–110. [PubMed: 22844575]
70. Hunziker S, Celi LA, Lee J, et al. Red cell distribution width improves the simplified acute physiology score for risk prediction in unselected critically ill patients. *Crit Care.* 2012; 16(3):R89. [PubMed: 22607685]
71. Hug CW, Clifford GD, Reisner AT. Clinician blood pressure documentation of stable intensive care patients: an intelligent archiving agent has a higher association with future hypotension. *Crit Care Med.* 2011; 39(5):1006–1014. [PubMed: 21336136]
72. Lee J, Kothari R, Ladapo JA, et al. Interrogating a clinical database to study treatment of hypotension in the critically ill. *BMJ Open.* 2012; 2(3)
73. Zimmerman JE, Kramer AA, McNair DS, et al. Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. *Crit Care Med.* 2006; 34(5):1297–1310. [PubMed: 16540951]
74. Lott JP, Iwashyna TJ, Christie JD, et al. Critical illness outcomes in specialty versus general intensive care units. *Am J Respir Crit Care Med.* 2009; 179(8):676–683. [PubMed: 19201923]
75. Wallace DJ, Angus DC, Barnato AE, et al. Nighttime intensivist staffing and mortality among critically ill patients. *N Engl J Med.* 2012; 366(22):2093–2101. [PubMed: 22612639]
76. Iwashyna TJ, Kramer AA, Kahn JM. Intensive care unit occupancy and patient outcomes. *Crit Care Med.* 2009; 37(5):1545–1557. [PubMed: 19325466]
77. Kilgannon JH, Jones AE, Shapiro NI, et al. Association between arterial hyperoxia following resuscitation from cardiac arrest and in-hospital mortality. *Jama.* 2010; 303(21):2165–2171. [PubMed: 20516417]
78. Gershengorn HB, Li G, Kramer A, et al. Survival and functional outcomes after cardiopulmonary resuscitation in the intensive care unit. *J Crit Care.* 2012; 27(2):421 e429–421 e417. [PubMed: 22227081]
79. McShea M, Holl R, Badawi O, et al. The eICU research institute - a collaboration between industry, health-care providers, and academia. *IEEE Eng Med Biol Mag.* 2010; 29(2):18–25. [PubMed: 20659837]
80. Halpern NA. From telemedicine to a critical care database: a new resource for national benchmarking. *Chest.* 2011; 140(5):1111–1113. [PubMed: 22045874]
81. Lilly CM, Zuckerman IH, Badawi O, et al. Benchmark data from more than 240,000 adults that reflect the current practice of critical care in the United States. *Chest.* 2011; 140(5):1232–1242. [PubMed: 21868469]
82. Galvagno SM Jr, Haut ER, Zafar SN, et al. Association between helicopter vs ground emergency medical services and survival for adults with major trauma. *Jama.* 2012; 307(15):1602–1610. [PubMed: 22511688]

83. Friese RS, Shafi S, Gentilello LM. Pulmonary artery catheter use is associated with reduced mortality in severely injured patients: a National Trauma Data Bank analysis of 53,312 patients. *Crit Care Med.* 2006; 34(6):1597–1601. [PubMed: 16607232]
84. Haut ER, Kalish BT, Cotton BA, et al. Prehospital intravenous fluid administration is associated with higher mortality in trauma patients: a National Trauma Data Bank analysis. *Ann Surg.* 2011; 253(2):371–377. [PubMed: 21178760]
85. Peberdy MA, Kaye W, Ornato JP, et al. Cardiopulmonary resuscitation of adults in the hospital: a report of 14720 cardiac arrests from the National Registry of Cardiopulmonary Resuscitation. *Resuscitation.* 2003; 58(3):297–308. [PubMed: 12969608]
86. Carr GE, Yuen TC, McConville JF, et al. Early cardiac arrest in patients hospitalized with pneumonia: a report from the American Heart Association's get with the Guidelines-Resuscitation Program. *Chest.* 2012; 141(6):1528–1536. [PubMed: 22194592]
87. Goldberger ZD, Chan PS, Berg RA, et al. Duration of resuscitation efforts and survival after in-hospital cardiac arrest: an observational study. *Lancet.* 2012 epub 5 Sept 2012.
88. Merchant RM, Yang L, Becker LB, et al. Variability in case-mix adjusted in-hospital cardiac arrest rates. *Med Care.* 2012; 50(2):124–130. [PubMed: 22249921]
89. Chan PS, Nichol G, Krumholz HM, et al. Hospital variation in time to defibrillation after in-hospital cardiac arrest. *Arch Intern Med.* 2009; 169(14):1265–1273. [PubMed: 19636027]
90. Chan PS, Nichol G, Krumholz HM, et al. Racial differences in survival after in-hospital cardiac arrest. *Jama.* 2009; 302(11):1195–1201. [PubMed: 19755698]
91. Medicare Claims Processing Manual, Chapter 25 - Completing and Processing the Form CMS-1450 Data Set. cited Available from: <http://www.cms.gov/Regulations-and-Guidance/Guidance/Manuals/downloads/clm104c25.pdf>
92. Angus DC, Kelley MA, Schmitz RJ, et al. Caring for the critically ill patient. Current and projected workforce requirements for care of the critically ill and patients with pulmonary disease: can we meet the requirements of an aging population? *Jama.* 2000; 284(21):2762–2770. [PubMed: 11105183]
93. Wunsch H, Guerra C, Barnato AE, et al. Three-year outcomes for Medicare beneficiaries who survive intensive care. *JAMA.* 2010; 303(9):849–856. [PubMed: 20197531]
94. Iwashyna TJ, Cooke CR, Wunsch H, et al. Population burden of long-term survivorship after severe sepsis in older Americans. *J Am Geriatr Soc.* 2012; 60(6):1070–1077. [PubMed: 22642542]
95. Iwashyna TJ, Ely EW, Smith DM, et al. Long-term cognitive impairment and functional disability among survivors of severe sepsis. *Jama.* 2010; 304(16):1787–1794. [PubMed: 20978258]
96. Kahn JM, Benson NM, Appleby D, et al. Long-term acute care hospital utilization after critical illness. *Jama.* 2010; 303(22):2253–2259. [PubMed: 20530778]
97. US Department of Health and Human Services, Public Health Service, National Center for Health Statistics National Hospital Discharge Survey 1979-2005. [Last accessed Dec 20, 2010] Multi-year Public-Use Data File Documentation. Available at: <http://www.cdc.gov/nchs/about/major/hdasd/nhdsdes.htm>
98. Clancy CM. Let the data be our guide: trends and tools for research on health care utilization. *Health Econ.* 2012; 21(1):19–23. [PubMed: 22147624]
99. Gershengorn HB, Iwashyna TJ, Cooke CR, et al. Variation in use of intensive care for adults with diabetic ketoacidosis\*. *Crit Care Med.* 2012; 40(7):2009–2015. [PubMed: 22564962]
100. Walkey AJ, Wiener RS, Ghobrial JM, et al. Incident stroke and mortality associated with new-onset atrial fibrillation in patients hospitalized with severe sepsis. *Jama.* 2011; 306(20):2248–2254. [PubMed: 22081378]
101. Seymour CW, Iwashyna TJ, Cooke CR, et al. Marital status and the epidemiology and outcomes of sepsis. *Chest.* 2010; 137(6):1289–1296. [PubMed: 20173054]
102. Bonafede MM, Suaya JA, Wilson KL, et al. Incidence and cost of CAP in a large working-age population. *Am J Manag Care.* 2012; 18(7):380–387. [PubMed: 22823532]
103. Ernst FR, Johnston JA, Pulgar S, et al. Timing of drotrecogin alfa (activated) initiation in treatment of severe sepsis: a database cohort study of hospital mortality, length of stay, and costs. *Curr Med Res Opin.* 2007; 23(1):235–244. [PubMed: 17207305]

104. Lindenauer PK, Pekow PS, Lahti MC, et al. Association of corticosteroid dose and route of administration with risk of treatment failure in acute exacerbation of chronic obstructive pulmonary disease. *Jama*. 2010; 303(23):2359–2367. [PubMed: 20551406]
105. Rothberg MB, Pekow PS, Lahti M, et al. Antibiotic therapy and treatment failure in patients hospitalized for acute exacerbations of chronic obstructive pulmonary disease. *Jama*. 2010; 303(20):2035–2042. [PubMed: 20501925]
106. Treggiari MM, Martin DP, Yanez ND, et al. Effect of intensive care unit organizational model and structure on outcomes in patients with acute lung injury. *Am J Respir Crit Care Med*. 2007; 176(7):685–690. [PubMed: 17556721]
107. Barnato AE, Albert SM, Angus DC, et al. Disability among elderly survivors of mechanical ventilation. *Am J Respir Crit Care Med*. 2011; 183(8):1037–1042. [PubMed: 21057004]
108. Iwashyna TJ, Netzer G, Langa KM, et al. Spurious inferences about long-term outcomes: the case of severe sepsis and geriatric conditions. *Am J Respir Crit Care Med*. 2012; 185(8):835–841. [PubMed: 22323301]
109. SEER-Medicare Linked Database. [cited 2012 Sept 5] Available from: <http://healthservices.cancer.gov/seermedicare/>
110. Slatore CG, Cecere LM, Letourneau JL, et al. Intensive care unit outcomes among patients with lung cancer in the surveillance, epidemiology, and end results-medicare registry. *J Clin Oncol*. 2012; 30(14):1686–1691. [PubMed: 22473159]
111. Seymour CW, Kahn JM, Cooke CR, et al. Prediction of critical illness during out-of-hospital emergency care. *Jama*. 2010; 304(7):747–754. [PubMed: 20716737]
112. Iezzoni LI. Using administrative diagnostic data to assess the quality of hospital care. Pitfalls and potential of ICD-9-CM. *Int J Technol Assess Health Care*. 1990; 6(2):272–281. [PubMed: 2203703]
113. Lindenauer PK, Lagu T, Shieh MS, et al. Association of diagnostic coding with trends in hospitalizations and mortality of patients with pneumonia, 2003-2009. *Jama*. 2012; 307(13):1405–1413. [PubMed: 22474204]
114. Riley GF. Administrative and claims records as sources of health care cost data. *Med Care*. 2009; 47(7 Suppl 1):S51–55. [PubMed: 19536019]
115. Brindis RG, Fitzgerald S, Anderson HV, et al. The American College of Cardiology-National Cardiovascular Data Registry (ACC-NCDR): building a national clinical data repository. *J Am Coll Cardiol*. 2001; 37(8):2240–2245. [PubMed: 11419906]
116. Fonarow GC, Abraham WT, Albert NM, et al. Influence of a performance-improvement initiative on quality of care for patients hospitalized with heart failure: results of the Organized Program to Initiate Lifesaving Treatment in Hospitalized Patients With Heart Failure (OPTIMIZE-HF). *Arch Intern Med*. 2007; 167(14):1493–1502. [PubMed: 17646603]
117. Hernandez AF, Greiner MA, Fonarow GC, et al. Relationship between early physician follow-up and 30-day readmission among Medicare beneficiaries hospitalized for heart failure. *Jama*. 2010; 303(17):1716–1722. [PubMed: 20442387]
118. Ghaferi AA, Birkmeyer JD, Dimick JB. Variation in hospital mortality associated with inpatient surgery. *N Engl J Med*. 2009; 361(14):1368–1375. [PubMed: 19797283]
119. Nathens AB, Cryer HG, Fildes J. The American College of Surgeons Trauma Quality Improvement Program. *Surg Clin North Am*. 2012; 92(2):441–454. x–xi. [PubMed: 22414421]
120. Gawande, A. *Better : a surgeon's notes on performance*. 1st. New York: Metropolitan; 2007.
121. Cook SF, Visscher WA, Hobbs CL, et al. Project IMPACT: results from a pilot validity study of a new observational database. *Crit Care Med*. 2002; 30(12):2765–2770. [PubMed: 12483071]
122. Ioannidis JP. The importance of potential studies that have not existed and registration of observational data sets. *Jama*. 2012; 308(6):575–576. [PubMed: 22871867]
123. Fan E, Checkley W, Stewart TE, et al. Complications From Recruitment Maneuvers in Patients With Acute Lung Injury: Secondary Analysis from the Lung Open Ventilation Study. *Respir Care*. 2012
124. Gershengorn HB, Wunsch H, Wahab R, et al. Impact of nonphysician staffing on outcomes in a medical ICU. *Chest*. 2011; 139(6):1347–1353. [PubMed: 21393397]

125. Watkins TR, Nathens AB, Cooke CR, et al. Acute respiratory distress syndrome after trauma: Development and validation of a predictive model\*. *Crit Care Med.* 2012; 40(8):2295–2303. [PubMed: 22809905]
126. Merchant RM, Yang L, Becker LB, et al. Incidence of treated cardiac arrest in hospitalized patients in the United States. *Crit Care Med.* 2011; 39(11):2401–2406. [PubMed: 21705896]
127. Checkley W, Brower R, Korpak A, et al. Effects of a clinical trial on mechanical ventilation practices in patients with acute lung injury. *Am J Respir Crit Care Med.* 2008; 177(11):1215–1222. [PubMed: 18356562]
128. Prasad M, Iwashyna TJ, Christie JD, et al. Effect of work-hours regulations on intensive care unit mortality in United States teaching hospitals. *Crit Care Med.* 2009; 37(9):2564–2569. [PubMed: 19623042]
129. Lyon SM, Kahn JM, Wunsch H, et al. The Impact of Massachusetts Health Insurance Reform on ICU Utilization. *Am J Respir Crit Care Med.* 2012; 185:A1494. Meeting Abstracts.
130. Pronovost PJ, Goeschel CA, Colantuoni E, et al. Sustaining reductions in catheter related bloodstream infections in Michigan intensive care units: observational study. *Bmj.* 2010; 340:c309. [PubMed: 20133365]
131. Hall WB, Willis LE, Medvedev S, et al. The implications of long-term acute care hospital transfer practices for measures of in-hospital mortality and length of stay. *Am J Respir Crit Care Med.* 2012; 185(1):53–57. [PubMed: 21940788]

**Table 1**  
**Aspects of databases that reflect the quality of data contained within**

<b>Quality Domain</b>	<b>Explanation</b>
<b>Coverage</b>	
Representativeness	How well does the data source represent the population that it intends to?
Completeness of recruitment	This feature measures the extent to which all eligible individuals have been included in the data collection scheme
Variables included	What is the extent of the data collected on each individual? Are demographic, exposure, outcome, and confounding variables present?
Completeness of variables	What is the extent of the missing data?
<b>Accuracy</b>	
Collection of raw data	Is the raw data collected or are aggregate averages collected?
Explicit definitions	Are the variables explicitly defined?
Explicit rules	Are there explicit rules for deciding how variables are recorded? For example, the timing of physiologic variables.
Reliability of coding	Was the reliability of coded conditions and interventions tested?
Independence of observations	Was the data recorder blinded to patient outcome at the time the data were collected?
Data validation	Were data validated using outside sources? Were there consistency checks?

Based upon Black, et al.(42)



**Table 2**  
**Assessing the research goal and types of questions answered using existing data**

Assessing the Research Goal	Types of Questions Answered	Need for Clinical Detail	Notes
Determine the causal relationship between risk factor and outcome	<ul style="list-style-type: none"> <li>What is the relationship between to recruitment maneuvers and complications in patients with ARDS?(123)</li> <li>What is the association between refractory hypotension and long-term cognitive function among survivors of sepsis?</li> </ul>	High	In select situations, one can use advanced methods for causal inference to attempt to overcome confounding by severity of illness such as propensity scores, instrumental variables, or quasi-experimental design.
Compare outcomes between two treatments, tests, or care delivery models (comparative effectiveness)	<ul style="list-style-type: none"> <li>Does the routine use of pulmonary-artery catheters improve outcomes for critically ill patients?(8)</li> <li>Do non-physician providers compared to physicians achieve better or worse mortality rates for their ICU patients?(124)</li> </ul>	High	Many therapies in critical care are confounded by indication - that is, the choice of treatment depends upon how sick patients are. Because most data sources do not collect the rationale behind medical decision making, overcoming this bias is difficult even with detailed data.
Determine the association between a risk factor or group of risk factors and outcome	<ul style="list-style-type: none"> <li>Are septic patients with a trial fibrillation at greater risk of stroke?(100)</li> <li>Can we predict ARDS after trauma?(125)</li> </ul>	Moderate	Predictive models are usually not evaluated on variables in model, but rather on their effectiveness at determining outcome
Describe epidemiology of disease, clinical practice, health service use, or health care spending	<ul style="list-style-type: none"> <li>What are the longitudinal trends in hospitalizations for sepsis?(94)</li> <li>What proportion of patients who receive CPR in the ICU are alive at six months?(126)</li> <li>To what extent does ICU admission for patients with DKA vary across hospitals?(99)</li> </ul>	Low	Descriptive epidemiology questions rarely require risk adjustment; comparisons between groups are not made. Such questions will also require richer data to approach well.
Evaluate policy or quality improvement intervention	<ul style="list-style-type: none"> <li>Did participation in a clinical trial of low tidal volume ventilation in ARDS change clinical practice?(127)</li> <li>Did resident work hour reform impact outcomes of ICU mortality?(128)</li> <li>Did health care reform in Massachusetts result in changes in ICU utilization rates?(129)</li> </ul>	Low	Although extensive clinical information is not often necessary for such analyses, when present it can greatly enhance the analysis.

**Table 3**  
**Description of features for example data sources containing critically ill patients**

Type of Data	Clinical Detail	Example Data Sources	Represented population	Accessibility/cost	Notes
Publicly-available studies	High	Acute respiratory Distress Syndrome Network(52)	Specific to individual studies	High/free	Application for data available online
Electronic medical record	High	Multiparameter Intelligent Monitoring in Intensive Care II(68)	>30,000 ICU patients at Beth Israel Deaconess since 2002	High/free	Minute-to-minute data on physiologic variables; time-stamped treatments
	High	Veterans Affairs, Inpatient Evaluation Center(62)	>100 Veterans hospitals	Low/variable	Allows for rich risk adjustment on par with APACHE. Use requires partnering with VA investigator.
	High	Epic, Cerner, local systems	Usually limited to local health system unless EMR system widely disseminated	Low/variable	
Quality improvement, benchmarking	High	Michigan Hospital Association Keystone initiative(130)	10 participating hospitals in Michigan	Low/-	Use usually requires partnering with Keystone investigator
	High	APACHE(75)	45 voluntarily hospitals	Low/-	Use requires partnering with Cerner
	High	eICU Research Institute (formerly VISICU)(79)	>1million patients in >180 non-federal voluntary hospitals	Low/-	Use requires partnering with University of Maryland.
Registries	Moderate to low	University Health System Consortium(131)	90% academic and affiliated hospitals in US	Moderate/-	
	Moderate	National Trauma Data Bank(82)	Nationally representative sample of trauma	Moderate/\$200/yr	Relatively inexpensive; obtained from the American College of Surgeons.
	Moderate	American Heart Association Get With the Guidelines Resuscitation(87)	>400 voluntarily participating hospitals	Moderate/free	
Administrative/utilization claims	Moderate	Premier Perspective(105)	Discharges of all payers from over 600 hospitals	Moderate/>\$50,000 - \$100,000	Nation's largest inpatient drug utilization database and uniquely includes timestamps for labs, tests, procedures
	Moderate to low	MarketScan(102)	>36 million discharges; >100 insurers	Low/thousands	
	Moderate to low	Medicare(93)	Currently enrolled beneficiaries	Moderate/\$900-\$20,000/file/year	Inpatient + skilled nursing combined file (MedPAR) ~ \$900-\$3500

Type of Data	Clinical Detail	Example Data Sources	Represented population	Accessibility/cost	Notes
	Low	National Hospital Discharge Database(27)	Nationally representative sample of more than 300 hospitals	High/free	Maintained by the CDC and downloadable from internet
	Low	Healthcare Cost and Utilization Project(98) Kids' Inpatient Database	Hospital discharges from 44 participating states	High/\$200-\$350/year	Price discounted for students
	Low	National Inpatient Sample	Nationally representative sample of hospital discharges	High/\$160-\$322/year	Price discounted for students
	Low	State Emergency Department Databases	ED discharges without admission from 37 participating states	High/\$35-\$1600/state/year	Price discounted for students
	Low	State Inpatient Databases	Hospital discharges from 44 participating states	High/\$35-\$3500/state/year	Price varies for profit/non-profit requestors