Published in final edited form as:

West J Nurs Res. 2017 September ; 39(9): 1271–1288. doi:10.1177/0193945916665814.

Abstracting ICU Nursing Care Quality Data from the Electronic Health Record

Jennifer B. Seaman, PhD, RNh*, Anna C. Evans, MD†, Andrea M. Sciulli, MS, PA-S‡, Amber E. Barnato, MD, MPH, MS§, Susan M. Sereika, PhD||, and Mary Beth Happ, PhD, RN, FGSA, FAAN¶

*Department of Critical Care Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

†Division of Gastroenterology, Hepatology, and Nutrition, Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

‡Seton Hill University, Greensburg, PA, USA

§Section of Decision Sciences, Division of General Internal Medicine, Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

||Department of Health and Community Systems, University of Pittsburgh School of Nursing, Pittsburgh, PA, USA

¶Center of Excellence in Critical and Complex Care, The Ohio State University College of Nursing, Columbus, OH, USA

Abstract

The electronic health record is a potentially rich source of data for clinical research in the intensive care unit setting. We describe the iterative, multi-step process used to develop and test a data abstraction tool, used for collection of nursing care quality indicators from the electronic health record, for a pragmatic trial. We computed Cohen's kappa coefficient (κ) to assess inter-rater agreement or reliability of data abstracted using preliminary and finalized tools. In assessing the reliability of study data (n=1440 cases) using the finalized tool, 108 randomly selected cases (10% of first half sample; 5% of last half sample) were independently abstracted by a second rater. We demonstrated mean κ values ranging from 0.61 to 0.99 for all indicators. Nursing care quality data can be accurately and reliably abstracted from the electronic health records of intensive care unit patients using a well-developed data collection tool and detailed training.

Keywords

critical care; tool development; data collection tool; nursing practice; quality indicators; electronic health record; intervention testing

Corresponding author: Jennifer B. Seaman, PhD, RN, Critical Care Research Fellow, Department of Critical Care Medicine, University of Pittsburgh School of Medicine, 607 Scaife Hall, 3550 Terrace St., Pittsburgh, PA 15261; Ph. (412) 352-6545; Fax. (412) 383-7293; jbs31@pitt.edu.

Conflict of Interests

The authors declare that there is no conflict of interest.

There is tremendous interest in shortening the timeline from discovery to implementation in healthcare research; yet such progress requires the ability of investigators to rapidly and efficiently conduct translational work, determining the effectiveness of new protocols, as well as testing their implementation and dissemination (Woolf, 2008). The utility of electronic health record (EHR) data as a source for process and outcome data for research has been well demonstrated (Al-Rawajfah, Aloush, & Hewitt, 2015; Behier, Reynier, Bertoye, & Vray, 2010; Elkhenini et al., 2015; Flaatten, 2012; Glavan, Engelberg, Downey, & Curtis, 2008; Kahn, Gunn, Lorenz, Alvarez, & Angus, 2014; Liddy, Wiens, & Hogg, 2011), and this approach is particularly well suited for research in the intensive care unit (ICU) setting as the ICU medical record contains round-the-clock, structured documentation of patient assessments and care provided. Measurement of nurse-sensitive quality of care indicators from the EHR has the potential to provide dense, patient-level data over time, thereby offering powerful insight into care processes, patient safety, and patient outcomes and can serve as a valuable methodology in translational research.

The Use of EHR Data for Research Poses Multiple Challenges

The capacity for automated queries and reporting of EHR data for research purposes has long been promoted as a major advantage to health record automation and was an often cited rationale for conversion from analog paper-based records (Agency for Healthcare Research and Quality, 2014). In ideal circumstances, the use of automated queries and reporting of EHR quality data can greatly improve data accuracy; deliver information more efficiently; and incur lower costs, compared to manual abstraction. The use of automated data extraction approaches is increasingly an option as around 80% of hospitals report some level of EHR use within their ICUs (HIMSS Analytics, 2014). However, collection of EHR data for the purposes of translational research depends upon multiple factors including the structure of the EHR as well as the cost and availability of resources needed to extract data. Not all EHRs are structured using standardized definitions for common indicators, and some trials may utilize measures that are more complex or have not been standardized (Samuels, McGrath, Fetzer, Mittal, & Bourgoine, 2015). In addition, not all institutions have the biomedical informatics technology infrastructure to support easy data retrieval, and the costs of customized data extraction can be prohibitive. In our study, EHR structure, data retrieval costs, and time required to obtain the extracted data were impediments to an automated approach. And while manual collection of care quality data via the EHR was labor intensive, it was nonetheless a more feasible choice than direct observation.

Purpose

We describe the development and testing of the data collection tool and standard operating procedure (SOP) for manual data abstraction from the EHR of selected nursing care quality indicators used in a multi-site ICU practice change implementation study. We also describe the training process for abstractors and the process for assessing data quality. While we demonstrate the utility of this approach, we also discuss the drawbacks involved.

Methods

Overview

Herein we report our experience with developing and implementing a tool for nursing care quality data abstraction from the EHR to assess outcomes of a pragmatic trial of the SPEACS-2 communication skills training intervention (Study of Patient-Nurse Effectiveness with Assisted Communication Strategies-2). The study employed a single-blind, randomized, crossover cluster (stepped-wedge) design and was conducted in 6 specialty ICUs across two hospitals within an academic health system. The sample was accrued by quarter (30 patients/ICU/quarter), and data were collected over an 8-quarter (2-year) period (Q1–Q8). This study was approved by the University of Pittsburgh Institutional Review Board; and a waiver of informed consent was obtained for abstraction of health record data (PRO09060348). Details of the trial are reported by Happ and colleagues (2015).

Given the sample size (N=1440), multi-site design, project budget, daily measures, and pragmatic intent of the study, the most feasible choice for care quality data collection for this large implementation was a manual retrospective abstraction of existing clinical data collected as part of the EHR, rather than prospective data collection. This approach was further supported by the presence of a shared EHR system among the study ICUs. The use of automated data retrieval was not only cost prohibitive, but the structure and patterns of nursing documentation in the EHR did not support an automated approach. For example, the nursing assessment section of the EHR contained multiple free-text fields, and nurses made frequent use of free-text documentation to clarify their assessment findings or better describe nursing care they delivered. Furthermore, the timeframe for receiving the extracted data was unclear. Therefore, we iteratively developed, refined, and evaluated an EHR data abstraction tool for retrospective assessment of the effectiveness of the SPEACS-2 program on patient-level care quality measures.

Measure selection-rationale and definitions

We based our choice of care quality measures on their hypothesized conceptual relationship(s) to successful and effective nurse-patient communication; endorsement by quality and safety standards bodies (American Nurses Association, 2014); and evidence they could be reliably assessed and abstracted from the EHR. We selected the following quality outcome measures: heavy sedation, physical restraint, pain presence, highest daily pain score, unplanned extubation and ICU-acquired pressure ulcers. For each we present the conceptual definition as well as the theoretical and methodological support for its inclusion:

Heavy sedation—Mechanically ventilated patients in the ICU receive sedation to decrease anxiety and restlessness and to facilitate ventilator synchrony and treatment. Because response to sedative medications varies widely among individual patients, our conceptual definition refers to the *state* of heavy sedation. A state of heavy sedation is a condition in which the patient has such depression the central nervous system (CNS) that attendant reduction in sensory and motor function results in minimal responsiveness to stimulation and inability to communicate with caregivers. Sedative medication can cause delirium, perpetuate the agitation-sedation cycle (Pandharipande et al., 2006; Pisani et al., 2009), and

prolong the duration of mechanical ventilation (MV)(Arroliga et al., 2005). Reductions in the use of heavy sedation, on the other hand, offer the potential for improved nurse-patient communication and are in alignment with current practice guidelines (Barr et al., 2013). Our team previously applied EHR data on neurological assessment to categorize levels of sedation-agitation maintaining IRR > .90 (Tate et al., 2005).

Physical restraint—Physical restraint refers to the application of unilateral or bilateral device(s) to restrict free movement of the upper/lower extremities. By restricting freedom of movement, physical restraints create barriers to a patient's ability to communicate via gestures, writing and communication boards, leading to feelings of frustration and anxiety. The use of physical restraints can cause both physiological and psychological harm; and physical restraints have been associated with increased stress response (cortisol, blood pressure, and heart rate), decreased immunological response, nerve impingement, and dehumanization (Strumpf & Evans, 1988; Sullivan-Marx, 1995; Wong & Chien, 2005). As a result, restraint use has been designated and tracked as a quality indicator by both the National Database of Nursing Quality Indicators (NDNQI) and the National Quality Forum (NQF) (American Nurses Association, 2014; Montalvo, 2007) Our previous research showed a negative relationship between physical restraint and nurse-patient communication (Happ, Tuite, Dobbin, DiVirgilio-Thomas, & Kitutu, 2004). Chart audits identify higher rates of restraint than observation since unlike direct observation, chart auditing provides restraint use data (incidence) over a 24 hour period. Edwards and colleagues (2006) found significant concordance ($p<.05$), but not complete agreement, between paired estimates of physical restraint prevalence using an observational tool and chart audits.

Pain presence and intensity—Pain has been conceptualized as a negative state, composed of both physical and affective components, in which there is fluctuation in intensity in response to both internal and environmental factors (Gélinas, Fortier, Viens, Fillion, & Puntillo, 2004). Effective pain management leads to improved outcomes among ICU patients (Chanques et al., 2010). Consistent self-report of pain intensity by patients led to a decrease in underestimation of the pain by staff; therefore, more effective communication of pain intensity *by patients* should result in more effective pain management(Happ et al., 2004). In prior research, the SPEACS intervention showed improvement in success of communication about pain and other symptoms (Happ et al., 2011). Clinical record review has been used in prior research to measure the presence of nursing pain assessment, compliance with pain rating protocols, and concordance between patient self-report and documentation of pain. (Gélinas et al., 2004) This is based on the work of Susan Beck, who with the support of the Robert Wood Johnson Foundation and NDNQI developed a quality indicator for pain management (American Nurses Association, 2014; Pett et al., 2013).

Unplanned extubation—Unplanned extubation refers to the accidental dislodgement or patient self-removal of the endotracheal or tracheostomy tube. Unplanned extubation is associated with restlessness, agitation and frustration (Atkins et al., 1997; da Silva & Fonseca, 2012; Grap, Glass, & Lindamood, 1995). Unplanned extubation is a costly, traumatic and potentially harmful event. Our research model purports that by facilitating

communication with nonvocal MV patients, their restlessness, agitation, and frustration will be relieved and unplanned extubation will be reduced.

ICU – acquired pressure ulcers—Hospital acquired pressure ulcers are a quality of care indicator targeted by the Center for Medicare and Medicaid Services (CMS) as a non-reimbursable adverse event (i.e., “never event”). Hospital acquired pressure ulcers lead to increased suffering and cost. Pressure ulcer (PU) prevalence in ICU settings varies from 4% to 49%, whereas the incidence of pressure ulcers ranges from 3.8% to 12.4% in the ICU (Shahin, Dassen, & Halfens, 2008). An ICU stay is associated with a doubling of risk for developing a PU among hospitalized older adults.(Baumgarten et al., 2008) In reviewing 356 video recordings of ICU patient-nurse interaction for an observational study to measure "Symptom Management, Patient-Caregiver Communication, and Outcomes", we noticed that nurses routinely offered turning and repositioning in response to patient complaints of generalized discomfort or pain (Happ & Tate, unpublished). This observed communication-care pattern suggests that when patients are able to communicate effectively about pain and/or discomfort, position changes are initiated. Thus, improved patient-nurse communication may result in greater attention to repositioning that prevents decubitus ulcer formation. A study by Gunningberg, Dahm, and Ehrenberg showed that the use of the EHR for documenting pressure ulcers, as opposed to paper forms, led to an increase in the their documentation as well as improved documentation accuracy by the nurses (2008).

We describe the operational definition and measurement for each measure in Table 1.

Preliminary tool development: Mapping indicators onto data fields in the EHR

At the outset of the project, the research team drafted a pilot data collection tool which contained the target data elements related to the variables selected. The pilot instrument consisted of sections for both single time point data (e.g., demographics) and repeated daily measurement of quality indicators, see Tables 1 and 2. We first located data elements within the EHR, and revised the preliminary tool so that nomenclature was consistent with the EHR. In situations where data could be stored in multiple locations within the EHR, we evaluated all locations for data accuracy and ease of abstraction and accuracy (Utter et al., 2011). For example, to determine if any ICU acquired pressure ulcer had occurred, one could review the bedside nursing documentation over the course of the ICU stay. Another option, however, was to review the documentation by the enterostomal therapy (ET) nurses, who are automatically (electronically) consulted when there is any skin breakdown that is Stage II or greater (National Pressure Ulcer Advisory Council, 2016). The process in such cases was as follows: we (tool developers J.S. and A.E.) evaluated the options for accuracy of the data and efficiency of abstraction, with input from expert clinicians (most often clinical nurse specialists) if needed; we presented the options to the study team for evaluation; and the study team made a consensus decision. In the case of pressure ulcer data, the ET nurse documentation was more accurate in terms of differentiation of wound type (pressure ulcer vs. other forms of skin breakdown) and wound staging. In addition, ET nurse documentation was easily isolated in a single view, allowing for more efficient and reliable abstraction. Once the optimal location of each data element was determined, we developed a

corresponding standardized operating procedure (SOP) containing detailed instructions and EHR screenshots to guide the data collection process.

Initial Testing and Refinement

To pilot test the data collection tool and SOP, two patient cases were selected and we (J.S. and A.E.) worked separately to complete data collection for the first case, noting any questions or ambiguities. After independently abstracting the chart, we compared results. We brought discrepancies that we were unable to resolve to the study team to adjudicate. We made necessary clarifications and corrections to the SOP and repeated the process with the second record.

Examples of refinements made during this iterative testing phase included: adding categories to the list of admission diagnoses; expanding the process for determining presence of heavy sedation; and specifying situations where one type of documentation should be weighted more heavily/supersede another. For example, we determined that free-text neurological descriptions should supersede vague “forced choice” assessment descriptors from the neurological assessment drop-down menu in the EHR. This was based on our observation that the descriptor “sedated” was being used to refer to the use of sedating medication as well as the patient being in a sedated state. Since our objective was to identify when patients were in a heavily sedated state, we needed to examine a wider array of criteria to accurately determine the actual level of sedation/ arousal. For example, the nurse may have chosen the descriptor “sedated” but in the free text section included “nodding head yes/no to questions”. If there was no additional documentation about level of sedation, and no conflicting documentation (such as Glasgow Coma Scale motor score <6) (Teasdale & Jennett, 1974) we would conclude that the patient was awake and heavy sedation was not present.

In the next step, we (J.S. and A.E.) completed four more cases independently and then followed the same process of comparison of abstracted data, discussion and adjudication of discrepancies, and revision of the SOP and/or data collection tool accordingly. Upon achieving consensus on all data elements for these four cases, we used these four abstracted cases as the “gold standard” against which to compare future abstractors during training. The multi-step iterative process we used to develop and test the tool is depicted in Figure 1.

Analyses

To evaluate inter-rater agreement or reliability (IRR) of patient-level nursing care quality data collected in the study data, we computed simple or weighted Cohen’s coefficient kappa (κ) for: heavy sedation, restraint use, pain presence, highest daily pain score, and unplanned extubation on a daily basis for the period of ICU observation, and ICU-acquired pressure ulcer for the entire period of ICU observation. The data were analyzed using SAS (version 9.3, SAS Institute, Inc., Cary NC).

Inter-rater Reliability Assessment

Phase I—development and testing—For IRR assessment of the preliminary tool, we trained four graduate students without prior experience with the EHR to use the abstraction tool. After initial training, these four abstractors independently abstracted data from the four

“gold standard” charts. We calculated Cohen’s κ for each abstractor against the “gold standard” and summarized the mean κ for each quality indicator across the four abstractors. For items with mean $\kappa < 0.6$, the team discussed discrepancies and modified the tool and SOP to address sources of ambiguity or inaccuracy. With the data abstraction tool finalized, we trained additional abstractors and began the abstraction of patient-level care quality data for the SPEACS-2 study. Ultimately a total of 12 abstractors were trained across the duration of the data collection phase (8 Quarters). Of the two study team members involved in the tool’s development, one (J.S.) consistently served as the second rater for IRR assessment and was responsible for tool refinement, SOP development, and training. The other (A.E.) abstracted study data.

Phase II—quality assurance—In the second phase of IRR assessment we randomly selected 10% (n=18) of Quarter 1 cases (n=180) for independent abstraction by a second data collector. We used a stratified sampling scheme for case selection to ensure equal sampling of ICU units and abstractors. We computed Cohen’s κ statistics for the co-abstracted cases, and the team discussed the results. Feedback was provided to the abstractors to remediate any deficiencies in reliability of collected data. Cohen’s kappa coefficient was calculated for each day and averaged over the days for each quality indicator assessed on a daily basis.

We continued to conduct IRR assessment as described above on 10% of randomly selected cases through Quarter 4, evaluating the results after each quarter’s reliability statistics had been calculated. To maintain reliability of abstracted data, we employed consistent data collectors for the last 5 quarters of the study, and the team reviewed and discussed significant discrepancies. The majority of the EHR data collection (1126/1440) was conducted by a single abstractor (A.S.). For Quarter 5 through Quarter 8, reliability testing was conducted for 5% of randomly selected cases, as stability in IRR had been achieved. This yielded a total sample of 108 cases for IRR assessment.

Results

Phase I—Development and Testing

The results of the IRR assessment for the preliminary tool by four raters across four cases are displayed in Table 3. Results showed excellent IRR for restraint use, awake-AM (12:00am-11:59am), and unplanned extubation ($\kappa = 1.00, 0.84$, and 1.00 , respectively); substantial IRR for pain, pain score, and awake-PM (12:00pm-11:59pm) ($\kappa = 0.63, 0.73$, and 0.70 , respectively) and moderate reliability for heavy sedation (any in 24 hour period) ($\kappa = 0.56$). Reliability for ICU-acquired pressure ulcer was also excellent ($\kappa = 1.00$); however, there were no positive cases (those with ICU-acquired pressure ulcer) in the “gold standard” set to robustly assess abstraction of this measure. An additional finding of early IRR testing was the selection, by nurses, of the response option, “Unable to Communicate” when documenting the presence of pain and pain rating score for mechanically ventilated patients. Since the purpose of the study was to improve communication with mechanically ventilated patents, we added this descriptor as a variable indicating an incomplete pain assessment and collected these data to determine if the use of “Unable to Communicate” decreased after the

intervention. Reliability testing for the “unable to communicate” pain variable showed excellent agreement across raters ($\kappa = .88$).

Phase II—Quality Assurance

As reported in Table 4, analysis of IRR for study data from Quarters 1 through 8 showed substantial to excellent IRR over multiple quarters of data collection for all indicators except ICU-acquired pressure ulcer. Early results for ICU-acquired pressure ulcer, were problematic, with poor agreement ($\kappa = .05$). We provided additional training to the data abstractors and the data collection SOP was updated to provide more detailed direction. Detailed IRR assessment indicated some continued variability between individual raters in comparison with the standard for the heavy sedation item, especially early in the ICU stay. However, in general, IRR was maintained or improved for all outcome variables except ICU acquired pressure ulcer. Ultimately we re-abstracted pressure ulcer data for Quarters 1 through 4, which resulted in an improved agreement ($\kappa = .79$) for data from Quarters 1 through 4 and acceptable cumulative agreement for Quarters 1 through 8 ($\kappa = .61$).

Discussion

Given the current funding climate, nurse scientists are increasingly challenged to find novel means to conduct their research (Conn et al., 2015). This work on tool development and reliability testing contributes to nursing science in the critical care setting by demonstrating the feasibility of a manual approach to abstraction of care quality indicators from the EHR. This methodology facilitated the collection of data necessary to answer the research questions posed in the SPEACS-2 study (Happ et al., 2015). This tool also shows how EHR abstraction can reliably provide rich and detailed longitudinal data for the systematic study of quality issues and practice change implementation. With such detailed data it is possible to explore patterns and trends in care quality and patient outcomes and determine correlations with different patient-level or unit-level conditions such as diagnosis, time of day, provider mix, etc. In effect, EHR data represents a source of big data for the purposes of quality monitoring and improvement as well as practice-level implementation science research.

However, the use of a manual approach to abstraction was very labor intensive. And while this option was still less expensive than the available automated data retrieval services, it was not optimal. The ability to rapidly assess the effectiveness of interventions; improve quality processes and outcomes; and measure the sustainability of effects only becomes possible with automated approaches. The expanded use of EHRs as a means to monitor and improve quality is a provision in the Affordable Care Act, and increasingly EHRs are being used in health care quality research. As EHRs become the dominant repository for healthcare information, including records of hospital-based care, clinicians can potentially leverage this resource to develop and test novel models for continuous quality monitoring and improvement. Yet, the promise of the EHR as a powerful tool for quality improvement and translational research requires more than simply an automated patient record; the EHR must have the necessary structure and be embedded as part of larger health information technology infrastructure in order to realize that promise (Samuels, 2012). We noted that

frequently the nuances of each patient's situation were described in free text by nursing staff, making this information impossible to extract using automated methods. The process described in this paper was time-consuming and produced a tool with limited generalizability, demonstrating the extent to which translational research is impeded by manual approaches. Implementation science research has demonstrated the need for interventions tailored to meet the needs of a variety of care contexts and the ability to measure effectiveness and sustainability over time in dynamic environments (Chambers, Glasgow, & Stange, 2013). This level of flexibility and responsiveness cannot realistically be achieved using manual approaches. Next steps specific to this work are to test these definitions and indicators using automated data extraction.

Despite the potential of the EHR as a data source, it is important to recognize the limitations of using abstracted EHR data for quality improvement and especially for patient outcomes research (Terry et al., 2010). Validity of the data obtained from the EHR depends heavily on the accuracy of the clinician's assessment and documentation. Data selected and available from the patient record may be inadequate or represent invalid proxy measures for the phenomenon of interest (e.g., awake, alert as indicators of "no heavy sedation").

We discovered that nurses commonly used "Unable to Communicate" - an option from the EHR drop down menu for pain assessment. Although sedation and waning consciousness are clearly factors impeding pain communication and accurate assessment, for some, "Unable to Communicate" may be a habitual default used instead of arousing the patient and applying assistive communication techniques to ascertain pain presence, location and intensity. Similarly, Swan identified that ICU clinicians routinely recorded "Unable to Assess" neurological status in lieu of a thorough examination of arousability (Swan, 2014). A program to improve screening of neurological status, which was administered as part of training in the use of the Confusion Assessment Method for the ICU (CAM-ICU) resulted in improvements in nurse attentiveness to patient arousal before administering the CAM-ICU. The circumstances (heavy sedation, intubation type, etc.) and use of "Unable to Communicate" deserve further study. We collected data on the use of the term, "Unable to Communicate," as an additional measure of communication improvement during mechanical ventilation in ICU.

Thorough knowledge of the EHR system and confirmation of the validity of selected indicators/variables is required. Because most health record documentation is not entered into required fields, missing data becomes a significant problem (Al-Rawajfah et al., 2015; Terry et al., 2010). In addition, data that are manually abstracted require a robust process for tool validation and assessment of data reliability. To advance the use of EHR data for research, further testing and refinement of quality indicators is needed. Finally, broad organizational support, including the input of clinical practice experts from multiple disciplines and adequate IT resources, including data analytics, will be needed to fully integrate quality improvement into the ICU EHR and support ongoing translational research.

We have demonstrated that data on patient-level care quality indicators can be accurately and reliably abstracted from the EHR of ICU patients using a manual data collection approach. An iterative development process ensured a robust instrument, adequate to collect the

desired data across a variety of practice settings and types of illness while maintaining the targeted levels of IRR: a Cohen's κ at or above 0.6 – 0.7, a level generally accepted as “at least substantial agreement” (Landis & Koch, 1977). Thorough training on the use of the abstraction tool and SOP using a “training to competency” approach resulted in a high level of reliability across 12 individual abstractors. This method represents an accurate, but not necessarily efficient or cost effective means to collect detailed quality of care data. Further testing of these indicators, using automated data extraction would facilitate future practice-based, patient-focused research in the ICU setting.

Acknowledgments

The SPEACS-2 study was supported by a Robert Wood Johnson Foundation Interdisciplinary Nursing Quality Research Initiative (INQRI) Grant (#66633). Author was supported by the NIH (F31-NR014078); The Mayday Fund; and a John A. Hartford Foundation Archbold Pre-doctoral Fellowship (2011–13).

References

- Agency for Healthcare Research and Quality. A robust health data infrastructure. 2014. (AHRQ Publication No. 14-0041-EF). Retrieved from https://www.healthit.gov/sites/default/files/ptp13-700hhs_white.pdf
- Al-Rawajfah OM, Aloush S, Hewitt JB. Use of electronic health-related datasets in nursing and health-related research. *Western Journal of Nursing Research*. 2015; 37:952–983. [PubMed: 25421544]
- Arroliga A, Frutos-Vivar F, Hall J, Esteban A, Apezteguía C, Soto L, Anzueto A. Use of sedatives and neuromuscular blockers in a cohort of patients receiving mechanical ventilation. *Chest*. 2005; 128(2):496–506. [PubMed: 16100131]
- Atkins PM, Mion LC, Mendelson W, Palmer RM, Slomka J, Franko T. Characteristics and outcomes of patients who self-extubate from ventilatory support: A case-control study. *Chest*. 1997; 112(5): 1317–1323. [PubMed: 9367475]
- Barr J, Fraser GL, Puntillo K, Ely EW, Gelinas C, Dasta JF, Jaeschke R. Clinical practice guidelines for the management of pain, agitation, and delirium in adult patients in the intensive care unit. *Critical Care Medicine*. 2013; 41(1):263–306. [PubMed: 23269131]
- Baumgarten M, Margolis DJ, Localio AR, Kagan SH, Lowe RA, Kinosian B, Mehari T. Extrinsic risk factors for pressure ulcers early in the hospital stay: A nested case-control study. *Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*. 2008; 63(4):408–413.
- Behier JM, Reynier JC, Bertoye PH, Vray M. Clinical trials and E-health: Impact of new information technology applied to clinical trials (including source data-medical records) and to human and drug research. *Thérapie*. 2010; 65(4):317–322. 309–315. [PubMed: 20854753]
- Chambers DA, Glasgow RE, Stange KC. The dynamic sustainability framework: Addressing the paradox of sustainment amid ongoing change. *Implementation Science*. 2013; 8:117. [PubMed: 24088228]
- Chanques G, Viel E, Constantin JM, Jung B, de Lattre S, Carr J, Jaber S. The measurement of pain in intensive care unit: Comparison of 5 self-report intensity scales. *Pain*. 2010; 151(3):711–721. [PubMed: 20843604]
- Conn VS, Topp R, Dunn SL, Hopp L, Jadack R, Jansen DA, Moch SD. Science on a shoestring: Building nursing knowledge with limited funding. *Western Journal of Nursing Research*. 2015; 37:1256–1268. [PubMed: 25601935]
- da Silva PS, Fonseca MC. Unplanned endotracheal extubations in the intensive care unit: Systematic review, critical appraisal, and evidence-based recommendations. *Anesthesia and Analgesia*. 2012; 114(5):1003–1014. [PubMed: 22366845]
- Edwards N, Danseco E, Heslin K, Ploeg J, Santos J, Stansfield M, Davies B. Development and testing of tools to assess physical restraint use. *Worldviews on Evidence-Based Nursing*. 2006; 3(2):73–85. [PubMed: 17040512]

- Elkhenini HF, Davis KJ, Stein ND, New JP, Delderfield MR, Gibson M, Bakerly ND. Using an electronic medical record (EMR) to conduct clinical trials: Salford Lung Study feasibility. *BMC Medical Informatics and Decision Making*. 2015; 15:8. [PubMed: 25880660]
- Flaatten H. The present use of quality indicators in the intensive care unit. *Acta Anaesthesiologica Scandinavica*. 2012; 56(9):1078–1083. [PubMed: 22339772]
- Gélinas C, Fortier M, Viens C, Fillion L, Puntillo K. Pain assessment and management in critically ill intubated patients: A retrospective study. *American Journal of Critical Care*. 2004; 13(2):126–136. [PubMed: 15043240]
- Glavan BJ, Engelberg RA, Downey L, Curtis JR. Using the medical record to evaluate the quality of end-of-life care in the intensive care unit. *Critical Care Medicine*. 2008; 36(4):1138–1146. [PubMed: 18379239]
- Grap MJ, Glass C, Lindamood MO. Factors related to unplanned extubation of endotracheal tubes. *Critical Care Nurse*. 1995; 15(2):57–65.
- Gunningberg L, Dahm MF, Ehrenberg A. Accuracy in the recording of pressure ulcers and prevention after implementing an electronic health record in hospital care. *Quality and Safety in Health Care*. 2008; 17(4):281–285. [PubMed: 18678726]
- Happ MB, Garrett K, Thomas DD, Tate J, George E, Houze M, Sereika S. Nurse-patient communication interactions in the intensive care unit. *American Journal of Critical Care*. 2011; 20(2):e28–e40. [PubMed: 21362711]
- Happ MB, Sereika SM, Houze MP, Seaman JB, Tate JA, Nilsen ML, Barnato AE. Quality of care and resource use among mechanically ventilated patients before and after an intervention to assist nurse-nonvocal patient communication. *Heart and Lung*. 2015; 44(5):408.e402–415.e402. [PubMed: 26354859]
- Happ MB, Tate JA. Symptom Management, Patient-Caregiver Communication, and Outcomes. 2008 Unpublished raw data.
- Happ MB, Tuite P, Dobbin K, DiVirgilio-Thomas D, Kitutu J. Communication ability, method, and content among nonspeaking nonsurviving patients treated with mechanical ventilation in the intensive care unit. *American Journal of Critical Care*. 2004; 13(3):210–218. [PubMed: 15149055]
- HIMSS Analytics. EMR/SEHR adoption model scores by hospital type, 2nd quarter, 2014. Current EMRAM Scores. 2014 Retrieved from <http://www.himssanalytics.org/emram/scoreTrends.aspx>.
- Kahn J, Gunn SR, Lorenz HL, Alvarez J, Angus DC. Impact of nurse-led remote screening and prompting for evidence-based practices in the ICU. *Critical Care Medicine*. 2014; 42(4):896–904. [PubMed: 24201176]
- Landis J, Koch G. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*. 1977; 33(2):363–374. [PubMed: 884196]
- Liddy C, Wiens M, Hogg W. Methods to achieve high interrater reliability in data collection from primary care medical records. *The Annals of Family Medicine*. 2011; 9(1):57–62. [PubMed: 21242562]
- Montalvo I. The National Database of Nursing Quality Indicators™ (NDNQI®). OJIN: The Online Journal of Issues in Nursing. 2007 Sep 30. 12(3) Manuscript 2.
- National Pressure Ulcer Advisory Council. NPUAP pressure injury stages/categories. Educational and CLINICAL Resources. 2016 Retrieved from <http://www.npuap.org/resources/educational-and-clinical-resources/npuap-pressure-injury-stages/>.
- Pandharipande P, Shintani A, Peterson J, Pun BT, Wilkinson GR, Dittus RS, Ely EW. Lorazepam is an independent risk factor for transitioning to delirium in intensive care unit patients. *Anesthesiology*. 2006; 104(1):21–26. [PubMed: 16394685]
- Pett MA, Beck SL, Guo J-W, Towsley GL, Brant JM, Lavoie Smith EM, Donaldson GW. Confirmatory factor analysis of the Pain Care Quality Surveys (PainCQ®). *Health Services Research*. 2013; 48(3):1018–1038. [PubMed: 23205503]
- Pisani M, Murphy T, Araujo K, Slattum P, Van Ness P, Inouye S. Benzodiazepine and opioid use and the duration of intensive care unit delirium in an older population. *Critical Care Medicine*. 2009; 37(1):177–183. [PubMed: 19050611]

- Samuels JG. Abstracting pain management documentation from the electronic medical record: comparison of three hospitals. *Applied Nursing Research*. 2012; 25(2):89–94. [PubMed: 20974098]
- Samuels JG, McGrath RJ, Fetzner SJ, Mittal P, Bourgoine D. Using the electronic health record in nursing research: Challenges and opportunities. *Western Journal of Nursing Research*. 2015; 37:1284–1294. [PubMed: 25819698]
- Shahin ES, Dassen T, Halfens RJ. Pressure ulcer prevalence and incidence in intensive care patients: A literature review. *Nursing in Critical Care*. 2008; 13(2):71–79. [PubMed: 18289185]
- Strumpf NE, Evans LK. Physical restraint of the hospitalized elderly: Perceptions of patients and nurses. *Nursing Research*. 1988; 37(3):132–137. [PubMed: 3368353]
- Sullivan-Marx E. Psychological responses to physical restraint use in older adults. *Journal of Psychosocial Nursing and Mental Health Services*. 1995; 33(6):20–25.
- Swan JT. Decreasing inappropriate unable-to-assess ratings for the confusion assessment method for the intensive care unit. *American Journal of Critical Care*. 2014; 23(1):60–69. [PubMed: 24382618]
- Tate, J., Sereika, S., Swigart, V., Arnold, R., Hoffman, L., Roesch, T., Happ, MB. Exploring the relationship between anxiety and weaning from long-term mechanical ventilation (LTMV) treatment. Paper presented at the American Thoracic Society International Conference; San Diego, CA. 2005.
- Teasdale G, Jennett B. Assessment of coma and impaired consciousness: A practical scale. *The Lancet*. 1974; 304(7872):81–84.
- Terry AL, Chevendra V, Thind A, Stewart M, Marshall JN, Cejic S. Using your electronic medical record for research: A primer for avoiding pitfalls. *Family Practice*. 2010; 27(1):121–126. [PubMed: 19828572]
- Utter GH, Borzecki AM, Rosen AK, Zrelak PA, Sadeghi B, Baron R, Romano PS. Designing an abstraction instrument: Lessons from efforts to validate the AHRQ patient safety indicators. *Joint Commission Journal on Quality and Patient Safety*. 2011; 37(1):20–28. [PubMed: 21306062]
- Wong IS, Chien WT. Young medical patients' experience of physical restraint: An exploratory study. *Journal of Clinical Nursing*. 2005; 14(1):120–121. [PubMed: 15656859]
- Woolf SH. The meaning of translational research and why it matters. *JAMA: Journal of the American Medical Association*. 2008; 299(2):211–213. [PubMed: 18182604]

Figure 1.
Process for Tool Development and Testing

Table 1

Quality Indicators*

Quality Indicator	Definition
Heavy Sedation (any in 24 hrs.)	Evidence of heavy sedation <i>at any point</i> during the 24 hour interval as measured by: <ul style="list-style-type: none"> Modified Ramsay score 4–6 <i>or</i> Riker score of 1–2 <i>or</i> Nursing note description of unresponsiveness to verbal or tactile stimulation, or being comatose or anesthetized
Awake (AM)/(PM)	Awake for 8 out of 12 hours for 12:00am–11:59am or 12:00pm–11:59pm as defined by: <ul style="list-style-type: none"> Modified Ramsay score 1–3 <i>or</i> GCS motor score of 6 <i>or</i> Nursing note documentation of being alert, awake, arousable, responsive, or communicative
Restraint Use	All restraint devices used within the 24 hour interval being evaluated including: soft extremity restraints (specify number of limbs restrained), vests, waist belts, full side-rails, mitts, and enclosure beds.
Pain Presence	Presence of any pain during the 24 hour period being evaluated (Y/N)
Pain Score	Highest pain score on a scale of 1–10 (including half scores) for the 24 hour interval
Pain-Unable to Communicate	Any use of the descriptor “unable to communicate” in the pain assessment documentation during the 24 hour interval (Y/N)
Unplanned Extubation	Documentation of self-extubation or any dislodging of the endotracheal or tracheostomy tube that is not part of a routine, intentional extubation by the clinical staff.
ICU-Acquired Pressure Ulcer	Any pressure ulcer, Stage II or greater, occurring during the index ICU stay that was not documented on admission [cumulative for ICU stay]

* All items are daily observations and calculated as a proportion of days observed, except ICU-acquired Pressure Ulcer which is calculated a cumulative total for the ICU stay.

Table 2

Demographic and Clinical Characteristics

Demographic Characteristics	Definition
Age	Age of the patient on the day of hospital admission
Gender	Male/Female
Admission Date	Date of admission to the study ICU during which the subject was mechanically ventilated for 2 days and awake for one nursing shift. (Referred to as the Index ICU stay)
ICU Location	Name of the ICU to which the patient was admitted on the admission date
Admitting Diagnosis	Indication for admission to the ICU
Pre-hospital Functional Status	Functional ability related to ADLs & IADLs collected in the nursing admission assessment upon hospital admission
Admission Braden Score	Braden score assigned within 72 hours of admission to the hospital
Community-Acquired Pressure Ulcers	Documentation of any pressure ulcers (Stage II or greater) present on admission to the hospital

Table 3

IRR Results on Training Cases (n=16)

Indicator	mean Cohen's Kappa *
Restraint Use	1.00
Heavy Sedation (any in 24 hours)	.555
Awake (AM)	.843
Awake (PM)	.698
Pain Presence	.630
Pain Score	.729
Pain-Unable to Communicate	.883
Unplanned Extubation	1.00
ICU-acquired pressure ulcer	1.00

* Kappa Agreement: < 0 Less than chance agreement; 0.01–0.20 Slight agreement; 0.21–0.40 Fair agreement; 0.41–0.60 Moderate agreement; 0.61–0.80 Substantial agreement; 0.81–0.99 Almost perfect agreement (Landis & Koch, 1977).

Table 4

IRR Results on Quality Assurance Cases, Q1 (n=18) and Q1– Q8 (n=108)

Indicator	Mean Cohen's Kappa	
	<i>Q1</i>	<i>Q1–8</i>
Restraint Use	.930	.981
Heavy Sedation (any in 24 hours)	.885	.803
Awake (AM)	.931	.863
Awake (PM)	.793	.825
Pain Presence	.905	.861
Pain Score	.992	.991
Pain-Unable to Communicate	.884	.887
Unplanned Extubation	1.00	.989
ICU-acquired pressure ulcer	1.00 [*]	.610

^{*} Presence of ICU-acquired pressure ulcer after re-abstraction.