

# Comparison of Manual versus Automated Data Collection Method for an Evidence-Based Nursing Practice Study

M.D. Byrne<sup>1</sup>; T.R. Jordan<sup>2</sup>; T. Welle<sup>3</sup>

<sup>1</sup>Saint Catherine University, Nursing, Saint Paul, Minnesota, United States; <sup>2</sup>OptumInsight, Provider Consulting, Eden Prairie, Minnesota, United States; <sup>3</sup>Saint Cloud Hospital, Clinical Utilization, Saint Cloud, Minnesota, United States

## Keywords

Automation, comparative study, data collection, evidence-based nursing, nursing methodology research

## Summary

**Objective:** The objective of this study was to investigate and improve the use of automated data collection procedures for nursing research and quality assurance.

**Methods:** A descriptive, correlational study analyzed 44 orthopedic surgical patients who were part of an evidence-based practice (EBP) project examining post-operative oxygen therapy at a Midwestern hospital. The automation work attempted to replicate a manually-collected data set from the EBP project.

**Results:** Automation was successful in replicating data collection for study data elements that were available in the clinical data repository. The automation procedures identified 32 “false negative” patients who met the inclusion criteria described in the EBP project but were not selected during the manual data collection. Automating data collection for certain data elements, such as oxygen saturation, proved challenging because of workflow and practice variations and the reliance on disparate sources for data abstraction. Automation also revealed instances of human error including computational and transcription errors as well as incomplete selection of eligible patients.

**Conclusion:** Automated data collection for analysis of nursing-specific phenomenon is potentially superior to manual data collection methods. Creation of automated reports and analysis may require initial up-front investment with collaboration between clinicians, researchers and information technology specialists who can manage the ambiguities and challenges of research and quality assurance work in healthcare.

## Correspondence to:

Matthew D. Byrne, PhD, RN, CPAN  
Department of Nursing  
Saint Catherine University  
Email: [mdbyrne@stkate.edu](mailto:mdbyrne@stkate.edu)  
Phone: 651-690-6761  
Fax: 651-690-6941

**Appl Clin Inf** 2013;4: 61–74

DOI: 10.4338/ACI-2012-09-RA-0037

received: September 27, 2012

accepted: January 9, 2013

published: February 13, 2013

**Citation:** Byrne MD, Jordan TR, Welle T. Comparison of Manual versus Automated Data Collection Method for an Evidence-Based Nursing Practice Study. *Appl Clin Inf* 2013; 4: 61–74

<http://dx.doi.org/10.4338/ACI-2012-09-RA-0037>

## 1. Background

### 1.1 Data Re-Use for Nursing Research, Quality Improvement Projects, and Practice Change

The expanded need for integrating evidence into clinical nursing practice, particularly through the use of the Electronic Health Records (EHR) and Clinical Data Repository (CDR), necessitates the exploration of new methodologies for quickly gathering, evaluating, and responding to clinical agency data. The traditional, solitary use of manual data collection may inhibit the ability of an organization to promptly inform practice changes and continuously measure the impact of these changes on patient outcomes. Recent research has suggested that automated data collection may be a synergistic approach that will help to reduce the time and effort required to collect and review data, while also increasing the quality of the data [1, 2].

### 1.2 Automation of Data Collection

Data collection for quality assurance and research can be an intensive use of resources. There are multiple potential sources of error for studies relying on manual data collection, particularly if data collection is done across multiple care units and with multiple data collectors, even if the data collection is done within one clinical agency. Automation of data collection has been shown to conserve limited temporal resources, improve quality and accuracy of data, and potentially allow for the on-going and rapid evaluation of policy or practice changes through report generation [2].

Automated data collection can be an excellent means of re-using clinical data that may have initially been captured only for care documentation, billing, or regulatory reporting. The re-use of this clinical data via queries and automated reports is not a new means of carrying out clinical research or answering quality assurance questions. However, broadening these practices to address nursing phenomena and nursing research questions has been identified as a part of the nursing informatics and biomedical informatics research agendas [3]. Most contemporary studies in which data re-use is being examined or used are primarily focused on serving physician information needs or for regulatory and reimbursement purposes; all of which are vitally important, but do not serve the knowledge work and needs of nurses [4-6].

A small number of studies have demonstrated the value of automation in transforming raw data from the EHR and CDR into actionable knowledge that may be used to advance the work of the registered nurse. Keenan et al., described the HANDS Project, which was an effort to improve the quality of patient care by providing nurses with a core set of clinical data. Their work identified both the value of using nursing data and the challenges of creating data models that could inform practice [7]. Cho et al. described a study in which the CDR was utilized to analyze nursing practice variations, staffing, and interventions related to pressure ulcers [8]. Their work demonstrated the unique nature of nursing terminology and practice and the impact of a nursing discipline-specific focus for the creation of algorithms and examination of the CDR. Purvis and Brenny-Fitzpatrick shared their success in pulling clinical data from the EHR into electronic reports that were utilized by clinical nurse specialists to identify high-risk geriatric patients [9]. Nursing clinical practice should be improved and informed through the automation of data collection and the re-use of clinical data. The creation of automation algorithms to do this work, though, requires researchers and programmers to consider nursing-specific practices, interventions and terminology.

### 1.3 Initial Evidence-Based Practice Project

The modeling of data collection for this study was patterned after a nursing evidence-based practice (EBP) project carried out at St. Cloud Hospital; a 489-bed regional medical center in the CentraCare Health System and perennial recipient of the American Nurses Credentialing Center's Magnet designation for excellence in nursing [10]. The EBP project challenged existing culture by examining the habitual use of post-operative oxygen in an effort to better apply finite healthcare resources and limit the potential for post-operative complications. The methodology of the initial EBP study consisted

of manual collection of clinical data by the primary researcher and nursing staff in the orthopedic and post-anesthesia care units. Approximately 2-4 nurses were involved in this process per patient studied, depending on factors such as shift change.

## 2. Objectives

The objective of this study was to investigate the potential for automating the collection of data for a nursing EBP project that was previously carried out at St. Cloud Hospital using a manual data collection method. The study examined the pool of actual and potential data elements, as well as the methodological processes of the initial EBP project, in order to build the query algorithm and to determine if replication by query of the clinical data repository was feasible and accurate. The data collection from the manual method and the automated method were compared to identify the etiology of any discrepancies and determine feasibility of using an automated method for future study of nursing practice issues.

## Methods

### 3.1 Methodology for Manual Data Collection

The initial EBP project utilized a convenience sample of 44 total knee or hip replacement surgery patients (► Appendix-A) who were admitted to the peri-operative setting at St. Cloud Hospital from November 15, 2011 to December 12, 2011. The sample excluded minors (<18 years old) and the elderly (>75 years old). Patients with an American Society of Anesthesiologists (ASA) Physical Status Classification score of one or two were included while those patients with scores of four or greater were excluded. Patients with an ASA Physical Status Classification of three were included as long as they had no prior medical history of respiratory disease (► Appendix-B). The use of ASA Physical Status Classification satisfied internal stakeholder requests to exclude organizational patients with elevated risk for adverse outcomes following anesthesia.

The initial data were collected on paper forms by a number of different nurses on the post-anesthesia and orthopedic care units, as well as by the primary nurse conducting the EBP project. Data were abstracted from a variety of sources (► Figure 1) and entered into a spreadsheet which contained the following data elements [11]:

- Age
- Gender
- Weight
- Length of stay
- Surgical procedure
- ASA Physical Status Classification score
- Anesthesia type
- Anesthesia duration
- Oxygen saturation (SpO<sub>2</sub>): initial pre-operative, initial post-operative, 30 minute post-operative, and four hour post-operative
- Oxygen therapy device
- Oxygen flow rate

### 3.2 Methodology for Automated Data Collection

The main goal of the methodology in the current study was to attempt to replicate the data in the manual database using automated procedures (i.e. report creation) as a means of comparison. The first step in replicating the manual sample was to ascertain which data elements of interest were being captured within the hospital's electronic health record [12]. A member of the agency's clinical documentation team helped complete the mapping of manual data elements to their automated counterparts. The manual sample included several data elements which were abstracted from paper

sources or electronic free text. In the absence of robust natural language processing tools, the automated query instead substituted equivalent codified claims data from the hospital's billing system (► Figure 2).

Per agency protocols, data were extracted nightly from the electronic health record, transformed, and then loaded into a clinical data repository used for analytical reporting. The second step in replicating the manual sample was to create an algorithm by using a structured query language (SQL) stored procedure to manipulate the data in the CDR and achieve the researchers' design goal of creating one tuple (datamart record/row) of output per relevant clinical encounter (► Appendix-C). This manipulation included the pivoting and partitioning of recurring flow sheet values and inferential associations between data elements [13, 14]. The results were stored in a data mart to facilitate further online analytical processing by the researchers.

### 3.3 Methodology for Statistical Analysis

The researchers evaluated the resulting data from two perspectives. The first perspective was an analysis of the data element values, namely comparing the results from the manual and automated samples. The second perspective was an analysis of patient sets. Measures of sensitivity and specificity for the manual sample were calculated by comparing the patient set from the original EBP project versus the patient set that the automated query would have identified via application of the study's inclusionary and exclusionary criteria against the entire patient population in the clinical data repository. Statistical techniques for these analyses included frequency distributions, correlations, independent sample t-tests, difference scores, and the raw number of matches between data elements [15].

## 4. Results

### 4.1 Replicating the Patient Sample

The manual sample from the initial EBP project consisted of 44 patients. The automated query and subsequent validation revealed that 40 of these patients were true positives and the other four were false positives. Of the four false positives, one of the patients exceeded the age limit while the other three had an ASA Physical Classification Score of three along with prior medical history of respiratory disease (an exclusionary criterion). The automated query also identified 32 false negative patients who met the EBP project's inclusion criteria, yet were not chosen for the study. These patients were reviewed by the primary researcher to verify that they would have indeed met the inclusion criteria.

The patients who were included during the manual data collection were compared with those who were identified as accidentally excluded by using independent sample t-tests and analyzing the variables of age, weight, anesthesia duration, and oxygen saturation. All variable distributions for each group were examined to assure normality before further analyses were conducted. The only variable that demonstrated a statistically significant difference between the two samples was anesthesia duration ( $t = 2.201$ ,  $df = 71$ ,  $p = 0.031$ ), though the 13.4 minute time difference in the sample means (excluded = 142.9 min, included = 129.5 min) may not have been clinically significant in the study context.

### 4.2 Additional Data Element Findings

All patient level demographic variables (age, gender, weight) and the vast majority of encounter level surgical variables (procedure, ASA physical classification score, anesthesia type) were identical between the manual and automated samples. Length of stay contained one discrepancy, which was due to a math error during manual data collection. The descriptions for the surgical procedure and anesthesia type were more specific in the manual collection data set than the automated sample due to the prior discussed decision to query codified billing data rather than free text surgical case log notes.

Five data elements (anesthesia duration and SpO<sub>2</sub> measurements at four intervals) were analyzed in greater detail using univariate and bivariate analysis. Anesthesia duration contained six discrepancies, which were due to math and transcription errors during manual data collection. Despite these errors, the anesthesia duration data sets were strongly correlated ( $r = 0.98$ ,  $p < 0.001$ ), with descriptive statistics echoing this finding (manual data collection mean = 133.14, automated query mean = 132.52).

Descriptive statistics for the SpO<sub>2</sub> data indicated that the results for the manual versus automated methods were also very close across time (► Table 1). However, Pearson correlations (► Table 2) and the raw number of matched/unmatched values (► Table 3) between the two methods revealed a pattern of discrepancies that emerged and grew with the passage of time. SpO<sub>2</sub> scores were clustered closer together early in the pre-operative readings but became increasingly disparate in the post-operative readings on the care unit.

## 5. Discussion

### 5.1 Theories Behind Methodological Differences

The statistical and theoretical challenge in analyzing the results was that there was no true “gold standard” data set that would allow for the researchers to judge one method as superior to the other. However, several themes emerged during the analysis which could help explain the difference in results between the manual and automated data sets.

#### 5.1.1 Data Collection Methods

The attempted automation of SpO<sub>2</sub> data collection and subsequent comparison of results between the manual and automated data sets proved to be the most challenging aspect of the study. It is possible that both the manual data collection and the automated query produced accurate, yet disparate, information (for the SpO<sub>2</sub> data in particular) by relying on different sources for the data. While the automated query could only abstract secondary data from the EHR, the initial EBP study also gathered primary data from bedside devices (pulse oximeters and vital sign monitors) and care interventions (oxygen therapy) and referenced the paper-based intra-operative anesthesia record. As a result, some of the data from the initial EBP study may have never been entered into the EHR. Validating these discrepancies is a time consuming endeavor and often produces fruitless results due to the difficulty in distinguishing between inaccurate data and data that is accurate but missing from the EHR.

#### 5.1.2 Timing Issues

Data collectors may have differed in their knowledge of patient event times. In the manual data collection, SpO<sub>2</sub> values may have been entered by patient care assistants during their routine workflow (patient rounding). These recordings may not have coincided with the exact system time stamps utilized by the automated query or the time intervals prescribed by the study protocol.

Furthermore, psychological studies have suggested that the general populace prefers the use of round numbers as cognitive reference points [16, 17]. Manual data collectors may have approximated the timing of patient events, ostensibly to simplify the calculation of the 30 minute and four hour post-operative time intervals. Automated queries offer much simpler and more precise alternatives through the use of built-in SQL functions like *DATEADD* and *DATEDIFF*.

#### 5.1.3 Exception Handling

Human logic and the computer algorithm may have differed in their handling of exceptions such as null values and workflow variations. Human users understand clinical context and have the freedom to reference data that falls outside of the direct scope of the study in order to gain additional insight. This allows them to exercise discretion in the face of ambiguity. This flexibility is a substantial benefit in the complex and often unpredictable healthcare environment as it is challenging to code for every possible contingency.

However, humans frequently fail to apply this judgment in a consistent manner. The rigidity of the computer algorithm ensures that the study results will not contain two different interpretations of the exact same event. Furthermore, errors in query logic are often easily fixed, after which the results may be quickly re-generated. The process of “fixing” manually gathered data is much more tedious and could undermine study results altogether. Manual data collection is also subject to the realities of staffing levels, workflow routines, and the ebb and flow of patient care needs which might necessitate prioritization of time and energy to different patients at different times.

## 5.2 Significance and Implications of Research

The results of this study have clinical, scientific, technical, and financial implications that are vitally important to consider in ensuring that future efforts using EHRs can inform and enhance the knowledge work of nurses.

### 5.2.1 Clinical Provider Time

The actual clinical provider time was not examined in this study. However, the clinical time of the registered nurse is often an expensive and limited resource in the hospital setting. Any time taken to do data collection, even if it is just a few minutes, is time that could be applied to patient care or other professional practice activities. Although research is vitally important and definitively falls under the professional practice umbrella of nursing, the ability of modern EHRs and software tools should be leveraged to replace manual data collection which takes a small but nonetheless meaningful toll on nurses' time. This is particularly relevant in an era where EHRs are a source of frustration for nurses who often express that clinical documentation and time spent at the computer is taking time away from direct patient care [18, 19]. Future research could build upon the results of this study by quantifying and comparing the actual costs of developing an automated algorithm against the benefits in terms of actual cost savings in clinical provider time.

### 5.2.2 Scientific

The study contributes to the body of research that has identified some of the challenges in building algorithms that allow for description, inference, and understanding of nursing practice phenomena; in this case, oxygen therapy in orthopedic surgical patients. The ability of automated data collection methods to gather data pertinent to the work of nurses in a more accurate and consistent manner opens new possibilities for the scientific advancement of nursing practice, quality assurance/improvement activities, and regulatory reporting. This proof of concept work could be applied to Surgical Care Improvement Projects, patient outcome evaluation, and process improvement work in the future. The study validates what is already known about some of the pitfalls of human error that arise with manual data collection and secondary use of data from documentation artifacts but with nursing practice as the context for this study [20-23].

There were valuable insights gained regarding how to improve both manual and automated data collection. Manual data collection must continue as some variables in this study were still on paper (i.e. all of the intra-operative anesthesia care record) and manual validation may be needed to ensure that the automated algorithms are appropriately sensitive and specific. Automated collection, when carefully created and modeled from manual collection, can save time, improve accuracy and speed, and allow for ongoing and immediate response to surveillance as evidenced in medical and public health focused research [1, 2, 24]. Automation and re-use for physician and medical practice must continue to be examined to ensure that the findings carry over to nursing practice context.

### 5.2.3 Technical

Health information technology possesses the elusive potential to help bridge the “inferential gap” inherent to clinical decision making. Retrospective, population-focused analytical processing via the clinical data repository will accelerate the creation of evidence. This evidence, in turn, can and should be made available at the point of care through real-time, patient-focused transactional processing via clinical decision support systems (CDSS) [25, 26].

However, as this study illustrates, the quality of evidence produced is highly dependent upon the underlying data and medical data is renowned for its unique set of constraints and challenges [27].



Thus far, the industry's capacity to accumulate data has outpaced its ability to transform the raw data into actionable knowledge [28]. Furthermore, the relationship between the quality of data and the cost of obtaining it (i.e. administrative burden) is often conceptualized as a zero-sum game, with gains in one area coming at the expense of the other. A limitation of this study is that it represents one clinical agency's experience with a particular vended system. There is wide variability in the EHR systems' ability to automate data collection, generate reports, and handle database preparation. Future research efforts should examine opportunities that hold the promise of concurrently advancing both the primary and secondary uses of data with broader implications for a variety of vended systems.

#### 5.2.4 Financial

The creation of these report algorithms required an initial upfront investment in time from researchers, clinicians, and programmers. This time would generally not be ongoing as would be the case with manual data collection over comparable time frames. Manual methods require the work of multiple nurses for collection and entry of data, including analysis with every step of practice and policy change or shift in outcomes of interest. The automated method, once developed, needs only to be updated with software, documentation, or practice changes as a part of routine maintenance of any decision support or report creation programs. In many cases, the programming that underpins such reports can be modified and/or expanded to addend existing reports or to build new reports, essentially paying the investment forward. Further, because automated reports may reduce or eliminate many of the previously identified human errors, the financial and practice pitfalls well known in the CDSS field may be avoided [19, 29, 30]. Rapid turnaround and analysis of data allows health care units and organizations to quickly respond to disease, infection, or risk trends rather than waiting longer and dealing with potential manual errors which bear financial and quality costs.

## 6. Conclusions

Pressures to improve quality and safety with shrinking budgets and reimbursements require clinicians and information technology (IT) specialists to rapidly cycle clinical data to inform practice. Nursing practice has lagged behind in these efforts and automation, as opposed to traditional manual data collection, was shown here to be a potentially sustainable, accurate, and expandable means of studying clinical phenomena relevant to nursing practice. The upfront investment of time and energy to create reports that can be easily maintained and upgraded will minimize interruptions to nursing care while capitalizing on the expensive investment in IT infrastructure and the EHR. In the long and short term, automation efforts can bear the fruits of improved patient outcomes through rapid data re-use and analysis to identify deficiencies and errors while evaluating changes in nursing practice.

#### Clinical Relevance Statement

Automation of data gathering for research and quality improvement efforts examining nursing practice and phenomenon is an accurate and precise method that can replace manual data collection methods. Clinicians, particularly nurses, must collaborate with IT specialists to ensure that nursing practice issues are addressed by leveraging the expensive and sophisticated modern IT infrastructures and the EHR. Automation has the potential to rapidly inform nursing practice and to save valuable time and money by up-front investment that can be sustainably maintained and upgraded.

#### Conflicts of Interest

The authors declare that they have no conflicts of interest in the research.

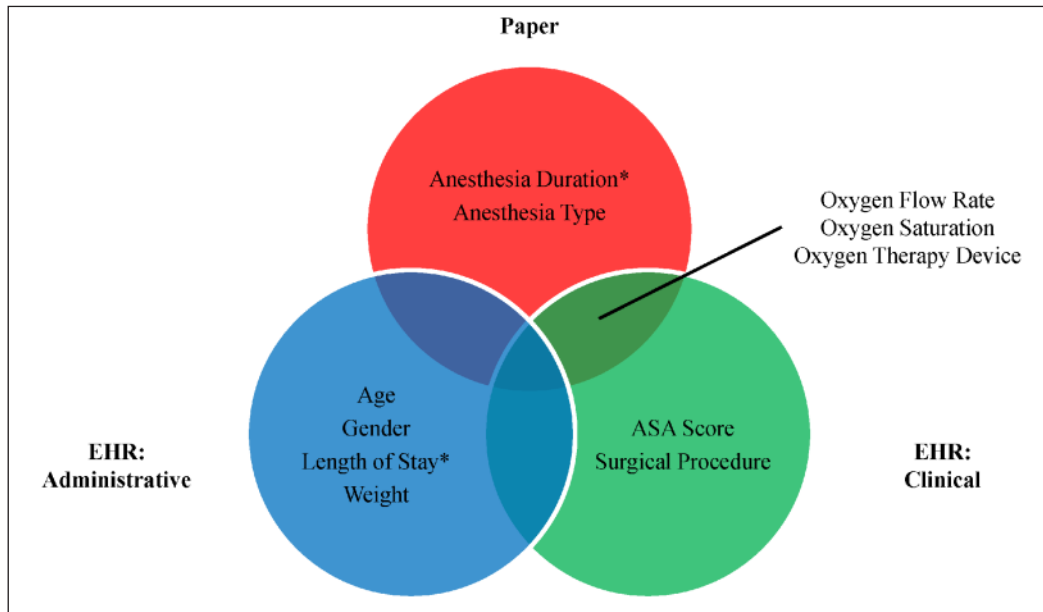
**Protection of Human Subjects**

The study was performed in compliance with the World Medical Association Declaration of Helsinki on Ethical Principles for Medical Research Involving Human Subjects, and was reviewed and approved by the Nursing Research Review Board and the Institutional Review Board of CentraCare Health System. All patient data would meet federal qualifications for de-identification and was securely stored, transmitted, and accessed by no parties other than the Principal Investigator, Co-Principal Investigator, and statistician.

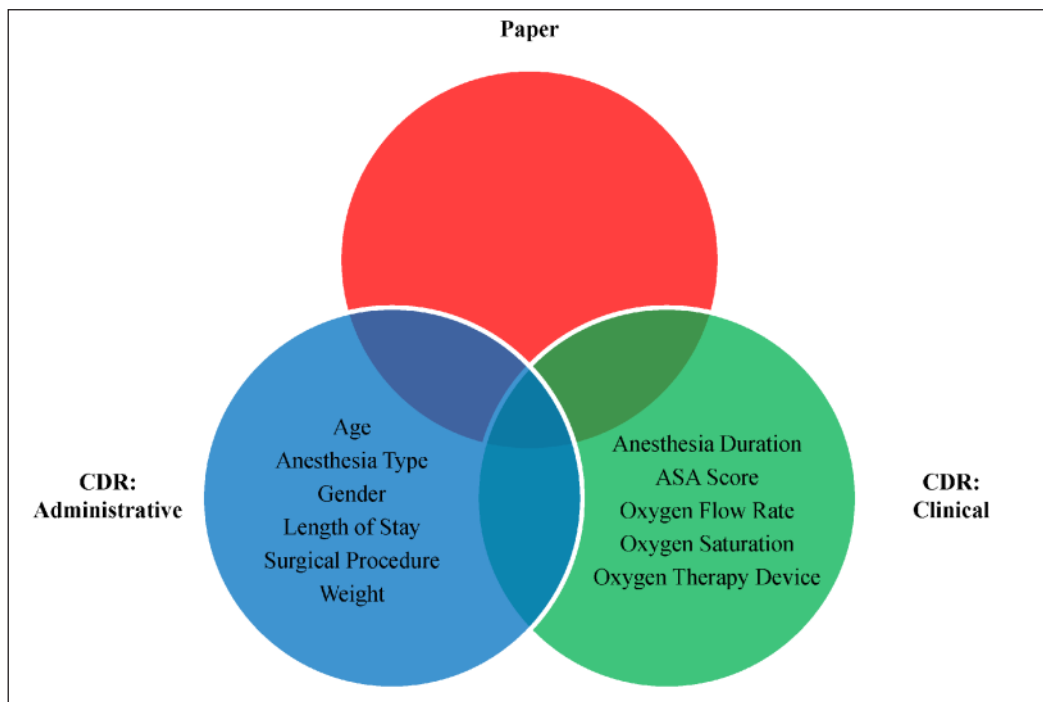
**Acknowledgements**

The authors want to acknowledge the nursing and information technology administrators of CentraCare Health System for their support of this research; especially, Amy Porwoll, Linda Chmielewski, Roberta Basol, and Naomi Schneider. The authors would also like to thank Jeffrey Ferranti for advisement, Sandra Pettingell for statistical analysis, Brenda Hansen for identification of electronic data elements, and all of the staff who participated in the original evidence-based nursing practice project upon which this research study was based.





**Fig. 1 Manual Data Collecting Methodology.** "Paper" source include anesthesia record, bedside devices, and EBP data collection tool. "EHR: Administrative" sources include registration, admission/discharge/transfer, and billing information. "EHR: Clinical" sources includes surgical case log and flowsheet documentation. Overlap regions indicate that data was collected from a combination of multiple sources; \*indicates that manual calculations were required.



**Fig. 2 Automated Data Collection Methodology.** Sources classified in same manner as Figure 2, with the exception that CDR replaces EHR as the electronic source.

Variable	Statistic	Manual (EBP)	Automated (Query)
PreOp Initial SpO <sub>2</sub>	Mean	96.40	96.48
	Median	96.00	96.00
	StDev	1.892	1.898
	Range	92–100	92–100
PostOp Arrival SpO <sub>2</sub>	Mean	95.72	96.14
	Median	96.00	96.50
	StDev	3.119	2.799
	Range	88–100	91–100
PostOp 30 Min SpO <sub>2</sub>	Mean	95.63	95.80
	Median	95.00	96.00
	StDev	2.928	2.777
	Range	89–100	90–100
PostOp 4 Hr SpO <sub>2</sub>	Mean	95.62	96.50
	Median	96.00	96.00
	StDev	2.145	2.124
	Range	91–100	92–100

**Table 1**  
SpO<sub>2</sub> Descriptive Statistics (Univariate Analysis)

Variable	Pearson R Value	p-Value
PreOp Initial SpO <sub>2</sub>	0.997	0.000
PostOp Arrival SpO <sub>2</sub>	0.884	0.000
PostOp 30 Min SpO <sub>2</sub>	0.868	0.000
PostOp 4 Hr SpO <sub>2</sub>	0.469	0.010

**Table 2**  
SpO<sub>2</sub> Pearson Correlations (Bivariate Analysis)

**Table 3** SpO<sub>2</sub> Raw Number Matched Values

	PreOp Initial	PostOp Initial	PostOp 30 Min	PostOp 4 Hr
Match	42	30	28	18
No Match	2	14	16	26
% Match	95.5	68.2	63.6	40.9

## References

1. Ferranti J, Horvath MM, Cozart H, Whitehurst J, Eckstrand J, Pietrobon R, et al. A multifaceted approach to safety: The synergistic detection of adverse drug events in adult inpatients. *Journal of Patient Safety* 2008; 4(3): 184-190.
2. Inacio MC, Paxton EW, Chen Y, Harris J, Eck E, Barnes S, et al. Leveraging electronic medical records for surveillance of surgical site infection in a total joint replacement population. *Infect Control Hosp Epidemiol* 2011; 32(4): 351-359.
3. Bakken S, Stone PW, Larson EL. A nursing informatics research agenda for 2008-18: Contextual influences and key components. *Nurs Outlook* 2008; 56(5): 206, 214.e3.
4. McBride AB. Nursing and the informatics revolution. *Nurs Outlook* 2005; 53(4): 183, 191; discussion 192.
5. Lang NM. The promise of simultaneous transformation of practice and research with the use of clinical information systems. *Nurs Outlook* 2008; 56(5): 232-236.
6. Kaplan B. Evaluating informatics applications some alternative approaches: Theory, social interactionism, and call for methodological pluralism. *Int J Med Inform* 2001; 64(1): 39-56.
7. Keenan GM, Stocker JR, Geo-Thomas AT, Soparkar NR, Barkauskas VH, Lee JL. The HANDS project: Studying and refining the automated collection of a cross-setting clinical data set. *Comput Inform Nurs* 2002; 20(3): 89-100.
8. Cho I, Park HA, Chung E. Exploring practice variation in preventive pressure-ulcer care using data from a clinical data repository. *Int J Med Inform* 2011; 80(1): 47-55.
9. Purvis S, Brenny-Fitzpatrick M. Innovative use of electronic health record reports by clinical nurse specialists. *Clin Nurse Spec*. 2010; 24(6): 289-294.
10. St. Cloud Hospital [Internet]. St. Cloud (MN): CentraCare Health System. [cited 2012 Nov 7]. Available from: <http://www.centracare.com/hospitals/sch/>.
11. Office Excel [CD-ROM]. Version 12. Redmond (WA): Microsoft Corporation; 2007.
12. Hyperspace [CD-ROM]. Version 2010 IU2. Verona (WI): Epic Systems Corporation; 2010.
13. Toad for Data Analysts [CD-ROM]. Version 3.1. Round Rock (TX): Dell Incorporated, Quest Software; 2012.
14. Crystal Reports [CD-ROM]. Version 12. Walldorf (Germany): SAP AG, Business Objects; 2008.
15. SPSS Base [CD-ROM]. Version 14.0. Armonk (NY): IBM Corporation, SPSS; 2005.
16. Pope D, Simonsohn U. Round numbers as goals: Evidence from baseball, SAT takers, and the lab. *Psychol Sci* 2011; 22(1): 71-79.
17. Bhattacharya U, Holden CW, Jacobsen S. Penny wise, dollar foolish: Buy-sell imbalances on and around round numbers. *Management Science* 2012; 58(2): 413-431.
18. Petrovskaya O, McIntyre M, McDonald C. Dilemmas, tetralemmas, reimagining the electronic health record. *ANS Adv Nurs Sci* 2009; 32(3): 241-251.
19. Ash JS, Sittig DF, Campbell EM, Guappone KP, Dykstra RH. Some unintended consequences of clinical decision support systems. *AMIA Annu Symp Proc* 2007: 26-30.
20. Bonney W. Is it appropriate, or ethical, to use health data collected for the purpose of direct patient care to develop computerized predictive decision support tools? *Stud Health Technol Inform* 2009; 143: 115-121.
21. Braaf S, Manias E, Riley R. The role of documents and documentation in communication failure across the perioperative pathway. A literature review. *Int J Nurs Stud* 2011; 48(8): 1024-1038.
22. Elkin PL, Trusko BE, Koppel R, Speroff T, Mohrer D, Sakji S, et al. Secondary use of clinical data. *Stud Health Technol Inform* 2010; 155: 14-29.
23. Fadlalla AM, Golob J, J., Claridge JA. The surgical intensive care-infection registry: A research registry with daily clinical support capabilities. *Am J Med Qual* 2009; 24(1): 29-34.
24. Salem H, Roth E, Fornango J. A comparison of automated data collection and manual data collection for toxicology studies. *Arch Toxicol Suppl* 1983; 6: 361-364.
25. Stewart WF, Shah NR, Selna MJ, Paulus RA, Walker JM. Bridging the inferential gap: The electronic health record and clinical evidence. *Health Aff (Millwood)*. 2007; 26(2): w181-w191.
26. Ledbetter CS, Morgan MW. Toward best practice: Leveraging the electronic patient record as a clinical data warehouse. *J Healthc Inf Manag* 2001; 15(2): 119-131.
27. Cios KJ, Moore GW. Uniqueness of medical data mining. *Artif Intell Med* 2002; 26(1-2): 1-24.
28. Ferranti JM, Langman MK, Tanaka D, McCall J, Ahmad A. Bridging the gap: Leveraging business intelligence tools in support of patient safety and financial effectiveness. *J Am Med Inform Assoc* 2010; 17(2): 136-143.

29. Ash JS, Sittig DF, Dykstra RH, Guappone K, Carpenter JD, Seshadri V. Categorizing the unintended socio-technical consequences of computerized provider order entry. *Int J Med Inf* 2007; 76: 21-27.
30. Ash JS, Berg M, Coiera E. Some unintended consequences of information technology in health care: The nature of patient care information system-related errors. *J Am Med Inform Assoc* 2004; 11(2): 104-112.

## Appendix-A

The International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) is a system used to categorize morbidity and mortality information by assigning standardized diagnosis and procedure codes (► Table A).

Diagnosis Related Groups (DRGs) are a patient classification scheme that is used to relate a hospital's case mix to the costs it is expected to incur. All Patient Refined DRGs (APR-DRG) are an expansion of the basic code set to better represent non-Medicare populations and incorporate severity of illness subclasses.

Together, these code sets are utilized for a variety of purposes including reimbursement and epidemiologic monitoring.

## Appendix-B

Elevated pre-operative anesthetic risk level was defined in this study as an ASA Physical Status Classification score that exceeds three (► Table B.1). This system is used to assess the fitness of patients before surgery.

Patients with an ASA score of three were only included if the patient did not have a prior medical history of respiratory disease as defined in ► Table B.2.

## Appendix-C

### Automated Query Algorithm

1. Create initial patient set using the following inclusionary criteria:
  - a. Admission to St. Cloud Hospital between 11/15/2011 – 12/12/2011.
  - b. ICD-9-CM procedure code (81.51 or 81.54) or APR-DRG (301 or 302) indicates total hip or knee replacement surgery.
2. Remove from patient set anyone who meets the following exclusionary criteria:
  - a. Age <18.
  - b. Age >75.
  - c. ASA Physical Status Classification score >3.
  - d. ASA Physical Status Classification score = 3 and ICD-9-CM diagnosis code in problem list indicates prior medical history of respiratory disease (490-496).
3. Retrieve administrative data elements including age, gender, weight, admission date, discharge date, anesthesia charge, and surgical procedure. Calculate length of stay as difference between admission and discharge dates.
4. Retrieve single-response clinical data elements including ASA score, anesthesia start time, and anesthesia end time. Calculate anesthesia duration as the difference between anesthesia start and end times.
5. Retrieve multiple-response clinical data elements (henceforth “flow sheets”) including oxygen saturation, device (if applicable), and flow rate (if applicable). Pivot and associate flow sheet entries with hospital unit in which they were recorded by comparing documentation time stamp against ADT time stamps.
6. Partition flow sheet entries by patient encounter and hospital unit. Sort and row number by documentation time stamp. Select the patient's first entry in the pre-operative unit and the first entry in the post-operative unit.
7. Add 30 minutes to documentation time stamp from initial post-operative reading to use as an estimate. Compare estimate to actual documentation time stamps and select the closest entry.
8. Repeat prior step for the 4 hour post-operative reading.

Description	ICD-9-CM Procedure	APR-DRG
Hip Replacement	81.51	301
Knee Replacement	81.54	302

**Table A.1**

Codes Identifying Total Joint Replacement Procedures

**Table B.1** ASA Physical Status Classification System

Score	Description
1	A normal healthy patient
2	A patient with mild systemic disease
3	A patient with severe systemic disease
4	A patient with severe systemic disease that is a constant threat to life.
5	A moribund patient who is not expected to survive without the operation.
6	A declared brain-dead patient whose organs are being removed for donor purposes.

Description	ICD-9-CM Diagnosis
Bronchitis, not specified as acute or chronic	490
Chronic bronchitis	491
Emphysema	492
Asthma	493
Bronchiectasis	494
Extrinsic allergic alveolitis	495
Chronic airway obstruction, not elsewhere classified	496

**Table B.2**

Codes Identifying Respiratory Disease

Note: All associated sub-codes were also included