Published in final edited form as:

Int J Med Inform. 2018 March ; 111: 45–50. doi:10.1016/j.ijmedinf.2017.12.014.

Electronic Medical Records as a Replacement for Prospective Research Data Collection in Postoperative Pain and Opioid Response Studies

Todd Lingren¹, Senthilkumar Sadhasivam², Xue Zhang³, and Keith Marsolo¹

¹Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, USA

²Department of Anesthesia, Cincinnati Children's Hospital Medical Center, Cincinnati, USA

³Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, USA

Abstract

BACKGROUND AND AIM—Many clinical research studies claim to collect data that are also captured in the electronic medical record (EMR). We evaluate the potential for EMR data to replace prospective research data collection.

METHODS—Using a dataset of 358 surgical patients enrolled in a prospective study, we examined the completeness and agreement of EMR and study entries for several variables, including the patient's stay in the post-operative care unit (PACU), surgical pain relief and pain medication side effects.

RESULTS—For all variables with a completeness percentage, values were greater than 96%. For the adverse event variables, we found slight to substantial agreement (Cohen's kappa), ranging from 0.19 (nausea) to 0.48 (respiratory depression) to 0.73 (emesis).

CONCLUSION—The potential to use EMR data as a replacement for prospective research data collection shows promise, but for now, should be evaluated on a variable-by-variable basis.

Keywords

Electronic Medical Records; Data Quality; Dimensional Measurement Accuracy; Opioid Analgesics; Postoperative Pain; Respiratory Depression

Corresponding author details: Dr. K. Marsolo, Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, 3333 Burnet Ave., Cincinnati, 45229, USA, keith.marsolo@cchmc.org.

AUTHOR CONTRIBUTIONS

KM and SS conceived of and designed the study. TL and XZ performed data analysis and interpretation. KM obtained the data from the electronic health record and drafted the study conclusions. All authors contributed to and approved of the final manuscript.

ETHICAL APPROVALS:

This study was approved by the Institutional Review Board of Cincinnati Children's Hospital Medical Center.

COMPETING INTERESTS

The authors report no competing interests.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. INTRODUCTION¹

One of the potential benefits of electronic medical records (EMRs) is that capturing data electronically allows the information to be more readily used for purposes other than patient care.[1–3] It remains an open question, however, whether information captured during the course of healthcare operations is truly fit for use in other contexts.[4] Variations in clinical workflows, the ability to document the same information in multiple places in the EMR, and minimal requirements in terms of required fields means that a critical evaluation is necessary before deciding whether to use data captured during the course of clinical care are suitable for research.[5–6]

Assessing the quality and completeness of EMR data is a challenge,[7–10] and the lack of assessment can cause bias in the results.[11] A recently published data quality framework establishes a number of data quality metrics (e.g., conformance, completeness, plausibility) that can be applied to a given dataset variable as well as the concept of internal and external validation for each (i.e., verification and validation, respectively).[12] It can be relatively straightforward to establish conformance and completeness measures for a given variable (i.e., is administrative sex always coded as “M” or “F” or “U”, and does every patient have a value), but plausibility measures are much more challenging, especially when looking for an external comparator (e.g., does my institution have similar rates of diabetic patients as the general population for given age/sex parameters). At an average academic medical center, hundreds of datasets of EMR data are generated for use in retrospective research every year, yet it is rare for that dataset to be accompanied with a quality scorecard,[13] despite such calls from within the informatics community.[14] While one might expect each study team to perform their own quality assessments before running any analyses, it is unlikely that they would always have a sufficient understanding of how variations in clinical workflows and documentation practices can impact the downstream quality and availability of a given piece of data.

In this work, we investigate the quality of the documentation relating to post-operative analgesic response in the EMRs of children who underwent tonsillectomies at Cincinnati Children’s Hospital Medical Center (CCHMC), between the years of 2009 and 2016. We had a unique opportunity to cross-check the accuracy of the records because these patients were also enrolled in a clinical research study, whereby a member of the research staff was present throughout the child’s stay in a post-operative care unit (PACU), collecting clinical outcome data in real-time, independent of the health care providers documenting the patient’s vitals, outcomes and clinical status. This study database thus provides us with an external gold standard for comparison.

Previous work in this area have reported mixed results. Studies have found anesthesia documentation practices to be substandard when compared to established guidelines,[15] and when comparing the rates of adverse events, some studies have found manually-reported

¹Abbreviations: EMR (Electronic Medical Record), PACU (Post Anesthesia Care Unit), CCHMC (Cincinnati Children’s Hospital Medical Center), MAR (Medication Administration Record), RD (Respiratory Depression), PONV (Post-Operative Nausea or Vomiting), NRS (Numeric Rating Scale), FLACC (Face Legs Activity Cry Consolability)

events to be underreported to those obtained from electronic sources,[16] which others have found the opposite.[17] Some of this variation may be explained by the nature of the adverse event and whether it relies on a data stream that is automated (i.e., a feed from a device monitor) versus one that requires manual entry (e.g., vital signs), which speaks to the importance of understanding clinical documentation practices and how they affect whether the underlying dataset actually supports the research question being proposed. Given the paucity of studies that utilize an external comparator, we believe that our ability to compare data extracted from a mature, comprehensive commercial EMR with a prospectively-collected research study database adds to the body of knowledge in this area.

2. METHODS

2.1 Dataset description

Five hundred and four children between the ages of 6 and 15 undergoing outpatient adenotonsillectomy with standard perioperative anesthetic, surgical, and nursing care, were enrolled in a clinical study (approved by the CCHMC Institutional Review Board) in which clinical research staff, whose training included daily observations of clinical processes over a one month period, carefully documented the children's post-operative response to analgesics,[18, 19] including any side effects.[19–23] The PACU nursing staff entered similar data into the EMR (Epic) as part of their normal patient care workflow (we did not consider anesthesia or other inpatient documentation). This involves recording vital signs and other data in EMR “flowsheets” every 15 minutes, as well as any administered medications, which are documented in the Medication Administration Record (MAR). The information documented in the flowsheets includes the patient's pulse, blood pressure, temperature, respiratory status (including respiratory rate), oxygen saturation level, need for supplemental oxygen and airway support, pain assessments, sedation level, nausea and vomiting, and optionally, any additional notes. We also included the nursing notes associated with the PACU stay as part of our analysis. CCHMC uses a single EMR that is live across all parts of the clinical enterprise. The EMR implementation occurred through a phased roll-out that began in 2007. Each clinic/specialty had the opportunity to customize their EMR configurations by creating fields (typically flowsheets) to allow them to discretely capture the data elements that are needed to support their work in measuring and improving process measures and patient outcomes. The EMR module used by the nursing staff to document information on the patient's stay in the PACU was not live at CCHMC until mid-January 2010, so patients with surgeries prior to that go-live date were excluded, reducing our dataset to 358 patients. The specific variables that we chose to compare are defined and summarized in Table 1.

2.2 Variable description and analysis

For the variables in Table 1, we computed the completeness of each in both the study database and the EMR, defining completeness as whether a value exists or, in the case of a composite variable, whether it can be derived from its relevant subcomponents, for each patient in our dataset. Multiple values for a variable were only counted once. For some variables, particularly post-operative emesis (vomiting), only positive entries are routinely captured. Negative entries may occur, but are less frequent. We see similar patterns with

post-operative nausea, though negative entries can be found in the clinical narrative. As a result, we do not report a completeness value for these variables, as it is difficult to determine a true denominator.

To determine a mention of nausea or vomiting in the nursing notes associated with the PACU stay encounter, we searched using regular expressions matching full or partial keywords including (nausea, nauseous, vomiting, vomit, emesis, n/v). Negation was also considered, using a modified version of the NegEx algorithm to exclude false positive mentions.[24] If a patient had positive mentions and negative mentions, the patient was considered to have a net positive mention score. If the patient only had negative mentions, the patient was considered to have a negative mention score. In the emesis-related flowsheets (e.g., Emesis Occurrence, Emesis Description), any entry that did not indicate a lack of emesis (i.e., a value of '0' in the Emesis Occurrence field) was considered to be a positive mention.

For total opioid use, we considered the lack of a medication administration record in the EMR to be equivalent to a 0 in the study database (i.e., no opioids given). Since all inpatient medication administrations are captured in the EMR, we are confident that the lack of a record indicates that the event did not occur.

In evaluating a patient's pain, a variety of rating scales were used in the PACU. In our analysis, we focused on Numeric Rating Scale (NRS) scores, as the NRS is more subjective than the others that were used (e.g., FLACC, OUCHER, FACES). There were instances when a patient had an NRS score recorded in the study, but not in the EMR. Because every patient had some form of pain measurement recorded in the EMR, we do not believe that this reflects an incomplete data field. As a result, when computing completeness, we consider the denominator for NRS pain scores to be the number of patients with any pain score recorded in the EMR.

To assess the agreement between the entries in the study database and the EMR for a patient when a value existed in both, we generated Bland-Altman plots for the numeric variables and computed the Cohen kappa coefficient for positive entries of the binary variables (e.g., adverse events).[25]

3. RESULTS

3.1 Completeness

As shown in Table 2, for almost all measurements under consideration, there were comparable degrees of completeness between the EMR and the study database. For every variable where we computed completeness, we found the rate to be >97%.

3.2 Agreement

Figure 1 contains the Bland-Altman plots for the numeric variables, illustrating the agreement between the variables as recorded in the EMR and the study database. Figure 1a shows the time until the post-anesthesia criteria for patient discharge were met,[13] and Figure 1b, the duration of PACU stay. On average, the time until the criteria for discharge

were met is 7.75 minutes more in the EMR than in the study, with 96.1% of the data within the 95% CL (95% CL: -57.17, 41.70), and a total of 14 outliers. There was one discrepant instance where the EMR record far exceeded the study record that is out of range on the figure (295 min). For the duration of PACU stay, the mean difference was smaller, at 1.71 minutes, with 97.2% of the data within the 95% CL (95% CL: -62.12, 58.69). There were two discrepant instances with a difference greater than 200 minutes. One where the study recorded duration of PACU stay far exceeded that available from the EMR (324 min), and another where the EMR record far exceeded the study PACU stay duration (466 min).

Figure 1c shows the Bland-Altman plot for the total equivalent dosage of the three analgesics (morphine, fentanyl, and hydromorphone) and Figure 1d, the Bland-Altman plot for the numerical pain scores. The mean difference between the EMR- and study-recorded total dose is 0.012 mg, with 96.3% of the data within the 95% CL (95% CL: -1.12, 1.14). For the numerical pain scores, the mean difference is 1.89 with 98.8% of the data within the 95% CL (95% CL: -5.00, 8.79). None of the four Bland-Altman plots shows any discernable patterns. Figure 2 shows the scatter plots of actual values for EMR and study data.

To assess the agreement between EMR and study data for the adverse events (emesis, nausea and RD), we use the Cohen's kappa test. The results of these comparisons are shown in Table 3.

Looking more in-depth at the adverse event variables, we found there were 26 patients recorded as having had at least one episode of emesis in the EMR, which is about as many emesis events recorded in the study database (25). There are more than 5 times as many positive mentions of nausea in the study database than in the EMR (51 vs. 10), where nausea is infrequently documented. Of these 10 nausea EMR mentions, only seven are true positives relative to the study data. The kappa results show substantial agreement for emesis (0.73, 95% CL: 0.58, 0.87) and slight agreement for nausea (0.19, 95% CL: 0.06, 0.32) [21]. If we look at both post-operative nausea and emesis (PONV) as a single event, we find 56 instances in the study database and 31 in the EMR. The overall agreement for the single event is lower than for emesis alone, due to the performance of the nausea result, with a kappa of 0.009 (95% CL: -0.02, 0.19). The simple approach to NLP did not recapture sensitivity sufficiently, and added more false positives than expected. Without NLP, the kappa for PONV results is 0.49 (95% CL: 0.36, 0.63), showing moderate agreement.

For overall RD (mild and severe), the kappa is -0.02 (95% CL: -0.037, 0.003), or less than chance. For severe RD, the kappa is better, 0.22 (95% CL: 0.097, 0.35), which indicates fair agreement. RD is a composite measure derived from two different flowsheet variables: respiratory rate and oxygen saturation. Using the definition listed in Table 1, 83 patients in the study database were flagged as having some form of RD (either mild or severe). When applying the same criteria to the EMR, however, only 14 were found to have a RD. One possible explanation why so few patients are inferred as having had a RD using their EMR data is the fact that the expected frequency of documentation for respiratory vital signs during the PACU stay is rather sparse, at 15 minute intervals (these measures are continuously monitored, however). Therefore, it is conceivable that if a patient's oxygen

saturation or respiratory rate dips, supplementary oxygen would be administered quickly enough for both of those measures to register as normal at the next reading.

In order to determine how much of an impact the infrequent documentation of vital signs might have had on the results, we evaluated the time difference between the documentation of subsequent respiratory vital signs in the EMR. The median value is 900 seconds, or 15 minutes, as expected. However, there are a significant number of records where the sampling is sparser than that. This presents an actionable opportunity for a quality improvement intervention to ensure that the nursing staff adhere to established care guidelines and document vitals no fewer than 15 minutes apart.

To remove the sparse recording of vital signs as a factor and to increase sensitivity of the RD algorithm in the EMR, we modified the definition of RD to include those patients who received supplemental oxygen, in any form, more than 20 minutes after entering the PACU. This allows us to better detect the intervention to the RD. Using this revised criteria, we flag an additional 45 patients, bringing the total to 59. The kappa results comparing severe and mild RD in the study database to any RD in the EMR (severe + mild + supplemental oxygen) improve to 0.38 (95% CL: 0.26, 0.5). Looking at subsets of RD provides even better results. Comparing patients with severe RD in the study database with those that were flagged as severe RD or receiving supplemental oxygen in the EMR (59 patients vs. 59), we see a kappa of 0.48 (95% CL: 0.35, 0.6), indicating moderate agreement.

4. DISCUSSION

Our results illustrate the issues one must consider in using data from the EMR to replace prospective research data collection. When a variable is typically populated only when an event occurred, such as with emesis, the lack of documentation of the negative case may cause analysis issues, depending on the study. Aside from elements like nausea and emesis, we found that overall completeness was quite good, though there was a fair amount of variability in terms of agreement. This makes it difficult to know a priori whether a given research study variable can be replaced with data from the EMR. For example, looking at the time until the PACU discharge criteria were met, we found the completeness percentage to be 100% in both the study and the EMR. There was a mean difference of almost 7 minutes in terms of agreement, however. This discrepancy is likely due in part to the fact that it is clinically acceptable to record the discharge time in the EMR anywhere within 15 minutes of the actual time. Such imprecision may be acceptable for some studies, but it would be almost impossible to know that it exists without an external validation like the one conducted in this study.

More problematic are the discrepancies we found related to RD, which is a rare, but potentially life-threatening complication that can result from the use of opioids. After performing a review of the charts of the patients who were flagged as having a RD in the clinical study database but not in the EMR, we found that the discrepancies were not due to a single factor, but result from a variety of issues. These include the fact that RD is a composite measure of oxygen saturation and respiratory rate that was developed for clinical research that does not exist as a single variable within the EMR, and more importantly, the

relatively infrequent rate of documentation of respiratory vitals in the EMR can miss events that are captured during real-time study collection. In addition, we would expect that research study staff would collect more information about RD, since they would record even those occurrences of mild RD that are deemed as negligible by clinical providers and not documented in the EMR. Clinical demands might have also limited the time available to document cases of mild and resolved RD. This illustrates why it may not be possible to use EMR data to directly replace clinical research variables even if they are purported to measure the same thing. In the case of RD, we actually had better results supplementing vitals with proxy measures, like the use of supplemental oxygen, to identify cases. Thus, it is crucial to understand the clinical workflows of a given unit and how those workflows impact a given variable's fitness for use before using any EMR data for research.

5. CONCLUSIONS

While the secondary use of EMR data is promising, one must be aware of the limitation of the data. In this work, we were able to compare EMR data with manually collected records from a clinical study of 358 pediatric surgical patients. We found that certain well-defined measures, such as the duration of a patient's PACU stay, agreed very well. We saw less agreement with adverse events like nausea and RD. This was due in part to the differences in recording practices between the study and the EMR, as well as the subjectivity on the part of clinical staff on what is recorded in the EMR. As a result, proxy measures may be necessary in order to detect certain events. While our work represents the results from a single institution's EMR, our findings are similar to studies that looked at using anesthesia information management systems for detection of adverse events like emesis.[17] We believe that more study in this area is needed, in particular more multi-center studies that involve external benchmarks. Ultimately, we believe that it will be necessary to develop development of a more comprehensive framework that can be used to help determine for what types of research a given EMR variable might be suited, if any. This would go beyond the simple metrics of conformance, completeness and plausibility, and take into account documentation practices and variation in workflow so that investigators would be able to assess whether a variable or set of variables would support a certain study design for a given patient cohort and time period.

Acknowledgments

FUNDING

This work was supported in part by the following grants: UL1TR000077 from the National Center for Advancing Translational Sciences (NCATS) and U01HG006828 from the National Human Genome Research Institute (NHGRI). It was also supported with funding from Cincinnati Children's Hospital Medical Center (CCHMC) from the following sources, the Center for Pediatric Genomics, the Place Outcomes Research Award, the Translational Research Award and with funding from the Department of Anesthesia and the Division of Biomedical Informatics. The views expressed in this work are those of the authors.

Marietta Labbato was involved in the collection of the research study data.

References

1. Geissbuhler A, Safran C, Buchan I, Bellazzi R, Labkoff S, Eilenberg K, Leese A, Richardson C, Mantas J, Murray P, De Moor G. Trustworthy reuse of health data: a transnational perspective. *International journal of medical informatics*. 2013; 82:1–9. [PubMed: 23182430]
2. Coorevits P, Sundgren M, Klein GO, Bahr A, Claerhout B, Daniel C, Dugas M, Dupont D, Schmidt A, Singleton P, De Moor G, Kalra D. Electronic health records: new opportunities for clinical research. *Journal of internal medicine*. 2013; 274:547–560. [PubMed: 23952476]
3. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nat Rev Genet*. 2012; 13:395–405. [PubMed: 22549152]
4. van der Lei J. Use and abuse of computer-stored medical records. *Methods of information in medicine*. 1991; 30:79–80. [PubMed: 1857252]
5. Ancker, JS., Shih, S., Singh, MP., Snyder, A., Edwards, A., Kaushal, R. H. investigators. Root causes underlying challenges to secondary use of data. *AMIA Annual Symposium proceedings AMIA Symposium*; 2011; 2011. p. 57-62.
6. Botsis T, Hartvigsen G, Chen F, Weng C. Secondary Use of EHR: Data Quality Issues and Informatics Opportunities. *AMIA Joint Summits on Translational Science proceedings AMIA Joint Summits on Translational Science*. 2010; 2010:1–5.
7. Kahn, MG., Eliason, BB., Bathurst, J. Quantifying clinical data quality using relative gold standards. *AMIA ... Annual Symposium proceedings/AMIA Symposium*. AMIA Symposium; 2010. p. 356-360.
8. Kahn MG, Raebel MA, Glanz JM, Riedlinger K, Steiner JF. A pragmatic framework for single-site and multisite data quality assessment in electronic health record-based clinical research. *Med Care*. 2012; 50(Suppl):S21–29. [PubMed: 22692254]
9. Weiskopf NG, Hripcsak G, Swaminathan S, Weng C. Defining and measuring completeness of electronic health records for secondary use. *Journal of biomedical informatics*. 2013; 46:830–836. [PubMed: 23820016]
10. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *J Am Med Inform Assoc*. 2013; 20:144–151. [PubMed: 22733976]
11. Hripcsak G, Knirsch C, Zhou L, Wilcox A, Melton G. Bias associated with mining electronic health records. *J Biomed Discov Collab*. 2011; 6:48–52. [PubMed: 21647858]
12. Kahn MG, Callahan TJ, Barnard J, Bauck AE, Brown J, Davidson BN, Estiri H, Goerg C, Holve E, Johnson SG, Liaw ST, Hamilton-Lopez M, Meeker D, Ong TC, Ryan P, Shang N, Weiskopf NG, Weng C, Zozus MN, Schilling L. A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. *Egms*. 2016; 4:1244. [PubMed: 27713905]
13. Mackenzie SL, Wyatt MC, Schuff R, Tenenbaum JD, Anderson N. Practices and perspectives on building integrated data repositories: results from a 2010 CTSA survey. *J Am Med Inform Assoc*. 2012
14. Hersh WR, Cimino J, Payne PRO, Embi P, Logan J, Weiner M, Bernstam EV, Lehmann H, Hripcsak G, Hartzog T, Saltz J. Recommendations for the use of operational electronic health record data in comparative effectiveness research. *EGEMS (Washington, DC)*. 2013; 1:1018.
15. Elhalawani I, Jenkins S, Newman N. Perioperative anesthetic documentation: Adherence to current Australian guidelines. *J Anaesthesiol Clin Pharmacol*. 2013; 29:211–215. [PubMed: 23878444]
16. Benson M, Junger A, Michel A, Sciuk G, Quinzio L, Marquardt K, Hempelmann G. Comparison of manual and automated documentation of adverse events with an Anesthesia Information Management System (AIMS). *Studies in health technology and informatics*. 2000; 77:925–929. [PubMed: 11187690]
17. Simpao AF, Pruitt EY, Cook-Sather SD, Gurnaney HG, Rehman MA. The reliability of manual reporting of clinical events in an anesthesia information management system (AIMS). *J Clin Monit Comput*. 2012; 26:437–439. [PubMed: 22618299]

18. Sadhasivam S, Chidambaran V, Ngamprasertwong P, Esslinger HR, Prows C, Zhang X, Martin LJ, McAuliffe J. Race and unequal burden of perioperative pain and opioid related adverse effects in children. *Pediatrics*. 2012; 129:832–838. [PubMed: 22529273]
19. Sadhasivam S, Chidambaran V, Olbrecht VA, Esslinger HR, Zhang K, Zhang X, Martin LJ. Genetics of pain perception, COMT and postoperative pain management in children. *Pharmacogenomics*. 2014; 15:277–284. [PubMed: 24533707]
20. Sadhasivam S, Zhang X, Chidambaran V, Mavi J, Pilipenko V, Mersha TB, Meller J, Kaufman KM, Martin LJ, McAuliffe J. Novel associations between FAAH genetic variants and postoperative central opioid-related adverse effects. *Pharmacogenomics J*. 2015; 15:436–442. [PubMed: 25558980]
21. Sadhasivam S, Chidambaran V, Olbrecht VA, Costandi A, Clay S, Prows CA, Zhang X, Martin LJ. Opioid-related adverse effects in children undergoing surgery: unequal burden on younger girls with higher doses of opioids. *Pain Med*. 2015; 16:985–997. [PubMed: 25521773]
22. Sadhasivam S, Chidambaran V, Zhang X, Meller J, Esslinger H, Zhang K, Martin LJ, McAuliffe J. Opioid-induced respiratory depression: ABCB1 transporter pharmacogenetics. *Pharmacogenomics J*. 2015; 15:119–126. [PubMed: 25311385]
23. Biesiada J, Chidambaran V, Wagner M, Zhang X, Martin LJ, Meller J, Sadhasivam S. Genetic risk signatures of opioid-induced respiratory depression following pediatric tonsillectomy. *Pharmacogenomics*. 2014; 15:1749–1762. [PubMed: 25493568]
24. Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. A simple algorithm for identifying negated findings and diseases in discharge summaries. *Journal of biomedical informatics*. 2001; 34:301–310. [PubMed: 12123149]
25. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med*. 2005; 37:360–363. [PubMed: 15883903]

SUMMARY TABLE

Research report – addition to literature:

A. What is already known

- Electronic medical records are used to collect many of the same variables used in prospective postoperative pain and opioid response research studies
- The use of EMR data as a replacement for prospective research data collection could drastically reduce the resources required for a given study

B. What this article adds

- We found high levels of completeness and agreement for a number of variables
- Composite variables (e.g., combining several variables) or those that signify an adverse event, showed far lower levels of agreement
- For certain studies, such as genomic and outcomes research, EMR data may be of sufficient quality to obviate the need for prospective research data collection

Highlights

- We evaluate the potential for EMR data to replace prospective research data collection.
- We compared 358 surgical patients enrolled in a prospective study with the completeness and agreement of EMR and study entries for several variables, including the patient's stay in the post-operative care unit (PACU), surgical pain relief and pain medication side effects.
- Completeness values were greater than 96%.
- For adverse event variables, we found slight to substantial agreement
- Using EMR data as a replacement for prospective research data collection shows promise for some variables, but more study is needed.

Bland-Altman Plots

Figure 1.

Bland-Altman plots comparing EMR and study database values for the numeric clinical variables when a value was present in both sources: (a) the time until Anesthesia Department's criteria for patient discharge from PACU were met, (b) the total duration of the PACU stay, (c) the total dose of post-operative analgesics received, and (d) numerical (NRS) pain scores. The solid line indicates the mean difference; the dashed lines correspond to the 95% confidence limits.

Figure 2.

Scatter plots comparing agreement between EMR and study databases for (a) the time until Anesthesia Department's criteria for patient discharge from PACU were met, (b) the total duration of the PACU stay, (c) the total dose of post-operative analgesics received, and (d) numerical (NRS) pain scores. The solid line indicates perfect agreement.

Table 1

Description of the clinical variables compared in this study.

Variable Name	Variable Definition
Time to Meet Post-Anesthesia Care Unit (PACU) Discharge Criteria	Duration of a patient's stay in the PACU until Anesthesia Department's criteria for discharge were met, in minutes: <ol style="list-style-type: none"> 1 level of consciousness – easily arousable or awake 2 airway – patent with adequate air exchange 3 core body temperature 36-3 degrees Celsius 4 acceptable pain level (pain score <4) 5 hemodynamically stable 6 no significant opioid related adverse effects 7 surgical site without any bleeding or complications.
Total PACU Stay	Total duration of a patient's stay in PACU, in minutes.
Respiratory Depression (RD)	RD is a composite variable defined as follows: <i>Severe RD:</i> <ol style="list-style-type: none"> 1 Peripheral oxygen saturation (SpO₂) 90%, OR 2. 90% < SpO₂ 2 lt; 94% for at least 10 min., OR 3 SpO₂ 94%, AND respiratory rate 8 breaths/min <i>Mild RD:</i> <ol style="list-style-type: none"> 4 90% < SpO₂ < 94% for less than 10 min., OR 5 SpO₂ 94%, AND 8 < respiratory rate 10 breaths/min. <i>Revised RD:</i> <ul style="list-style-type: none"> • In addition to the definitions above, If patients receive supplemental oxygen (simple mask, blowby) > 20 minutes after arrival in the PACU.
Max Pain – Numerical Rating Scale (NRS)	Maximum pain score using the numerical rating scale, which ranges from 0 (no pain) to 10 (maximal pain). In both EMR and the PATA study, pain was assessed at different time points. The maximum pain score is the maximum over all pain scores recorded.
Total Opioid Analgesic Use	Combined dose of 3 analgesics: morphine, fentanyl, and hydromorphone, as documented in the Medication Administration Record (MAR), computed as follows: $\text{total dose} = \text{morphine dose} + (\text{fentanyl dose})/10 + (\text{hydromorphone dose})/0.2$
Emesis (post-operative vomiting)	The presence of any number of post-operative vomiting (emesis) episodes. Inferred from EMR if any of the following variables are filled: "Emesis Occurrence" (numeric), or "Emesis Amount", "Emesis Description", or "Emesis Frequency" (text). Can also be inferred from "Additional Detail" and nursing notes.

Table 2

Count of patients with data recorded in the EMR and study database for each clinical variable, as well as the corresponding percentage for completeness.

Variable Name (<i>n</i> = denominator)	Patients with data in EMR (Count)	Patients with data in study DB (Count)	Complete in EMR (%)	Complete in study (%)
Time to meet PACU Discharge Criteria (<i>n</i> = 358)	358	355	100	99.2
Total PACU Stay (<i>n</i> = 358)	358	355	100	99.2
Total Opioid Analgesic Use (<i>n</i> =358; includes patients who received no opioids)	358 (221 received opioids)	358 (221 received opioids)	100	100
RD (<i>n</i> = 358)	358 had values recorded for oxygen saturation & respiratory rate	350 had values recorded for oxygen saturation & respiratory rate	100	97.8
Max Pain NRS (<i>n</i> =329)	329 (296 had a non-zero pain score)	329 193 had a non-zero pain score	100	100

Table 3

Adverse Event Kappa Results comparing data recorded in the EMR and study database

Adverse event comparison	Source description	Positive mentions (EMR)	Positive mentions (study DB)	Cohen's Kappa coefficient (95% CL)
Emesis	EMR flowsheets + MAR + NLP vs. Study DB	26	25	0.73 (0.58, 0.87)
Nausea	EMR flowsheets + MAR + NLP vs. Study DB	10	51	0.19 (0.06, 0.32)
PONV (emesis OR nausea)	EMR flowsheets + MAR + NLP vs. Study DB	31	56	0.009 (−0.02, 0.19)
Emesis	EMR flowsheets + MAR, no NLP vs. Study DB	22	25	0.79 (0.67, 0.92)
Nausea	EMR flowsheets + MAR, no NLP vs. Study DB	7	51	0.18 (0.05, 0.31)
PONV (emesis OR nausea)	EMR flowsheets + MAR, no NLP vs. Study DB	25	56	0.49 (0.36, 0.63)
RD (mild)	EMR flowsheets vs. Study DB	4	24	−0.02 (−0.0037, −0.003)
RD (severe)	EMR flowsheet vs. Study DB	10	59	0.22 (0.097, 0.35)
RD (mild OR severe)	EMR flowsheets vs. Study DB	14	83	−0.02 (−0.37, 0.003)
RD (mild OR severe OR supplemental oxygen)	EMR flowsheets vs. Study DB	59	83	0.38 (0.26, 0.5)
RD (severe OR supplemental oxygen)	EMR flowsheets vs. Study DB	59	59	0.48 (0.35, 0.6)