

The Reliability of Electronic Health Record Data Used for Obstetrical Research

Molly R. Altman¹ Karen Colorafi² Kenn B. Daratha²

¹Department of Family Health Care Nursing, University of California, School of Nursing, San Francisco, California, United States

²Department of Family Health Care Nursing, Washington State University, College of Nursing, Spokane, Washington, United States

Address for correspondence Molly R. Altman, PhD, CNM, MPH, 128 Banff Way, Petaluma, CA 94954, United States (e-mail: molly.altman@ucsf.edu).

Appl Clin Inform 2018;9:156–162.

Abstract

Background Hospital electronic health record (EHR) data are increasingly being called upon for research purposes, yet only recently has it been tested to examine its reliability. Studies that have examined reliability of EHR data for research purposes have varied widely in methods used and field of inquiry, with little reporting of the reliability of perinatal and obstetric variables in the current literature.

Objective To assess the reliability of data extracted from a commercially available inpatient EHR as compared with manually abstracted data for common attributes used in obstetrical research.

Methods Data extracted through automated EHR reports for 3,250 women who delivered a live infant at a large hospital in the Pacific Northwest were compared with manual chart abstraction for the following perinatal measures: delivery method, labor induction, labor augmentation, cervical ripening, vertex presentation, and postpartum hemorrhage.

Results Almost perfect agreement was observed for all four modes of delivery (vacuum assisted: kappa = 0.92; 95% confidence interval [CI] = 0.88–0.95, forceps assisted: kappa = 0.90; 95%CI = 0.76–1.00, cesarean delivery: kappa = 0.91; 95% CI = 0.90–0.93, and spontaneous vaginal delivery: kappa = 0.91; 95%CI = 0.90–0.93). Cervical ripening demonstrated substantial agreement (kappa = 0.77; 95% CI = 0.73–0.80); labor induction (kappa = 0.65; 95%CI = 0.62–0.68) and augmentation (kappa = 0.54; 95%CI = 0.49–0.58) demonstrated moderate agreement between the two data sources. Vertex presentation (kappa = 0.35; 95%CI = 0.31–0.40) and post-partum hemorrhage (kappa = 0.21; 95%CI = 0.13–0.28) demonstrated fair agreement.

Conclusion Our study demonstrates variability in the reliability of obstetrical data collected and reported through the EHR. While delivery method was satisfactorily reliable in our sample, other examined perinatal measures were less so when compared with manual chart abstraction. The use of multiple modalities for assessing reliability presents a more consistent and rigorous approach for assessing reliability of data from EHR systems and underscores the importance of requiring validation of automated EHR data for research purposes.

Keywords

- ▶ reliability
- ▶ electronic health records
- ▶ Cohen's kappa

received
August 2, 2017
accepted after revision
January 1, 2018

Copyright © 2018 Schattauer

DOI <https://doi.org/10.1055/s-0038-1627475>.
ISSN 1869-0327.

Background and Significance

Electronic health records (EHRs) have been widely adopted in the United States healthcare system for management of patient clinical data. President George W. Bush's 2004 Executive Order called for the development and implementation of a nationwide, interoperable health information technology infrastructure that could be used to improve the quality and efficiency of healthcare, calling for the creation of an EHR for all Americans by the year 2014.¹ After years of sluggish adoption rates, that mandate was eventually funded in 2009 by the Obama administration's stimulus legislation, The American Recovery and Reinvestment Act (ARRA). The HITECH Act was embedded in ARRA and authorized the Centers for Medicare and Medicaid Services along with the Office of the National Coordinator for Health Information Technology to establish the EHR Incentive Program. Meaningful Use, as it is commonly known, has been wildly successful in spurring the adoption of EHR in hospital settings. The last decade has seen a rapid increase in the use of EHRs in clinical practice from a low of 9% in 2008 to a high of 84% in 2015.² More than 96% of non-federal acute care hospitals have adopted a basic EHR, a nine-fold increase since 2008.²

Hospital EHR data are disseminated in report format for a variety of purposes, including clinical decision making, administrative evaluation, and research purposes; however, only recently have we begun to question its reliability. Reliability, the degree to which the result of a measurement or calculation is considered accurate, reflects the qualities of trustworthiness and consistency in data performance. A well-known feature of EHR systems is the ability to document the same thing in multiple places. For end users, flexibility and customization are highly desirable, but these can be cumbersome and inaccurate for non-clinical applications that rely on data being captured in one discrete field on a structured form. More often than not, a great deal of clinical information in the EHR is recorded in free-text fields or dictated narrative notes and therefore not captured using automated processes. Compounding the problem, individual EHR installations can alter structured templates and fail to alter "out of the box" vendor-generated reports that make it possible to extract data automatically by pulling documented discrete data elements. In this way, we run the risk that auto-generated EHR reports under report cases of interest.

To help assimilate data that are entered in different ways and retrieved by a variety of methods, innovative techniques have been tested to assess the reliability of extracted EHR data. These include (1) automated extractive text summarization methods of free text in assisting clinicians with clinical care,³ (2) reconciliation of registries to administrative data in hospital discharge databases,⁴ (3) the reporting of reliability of EHR data used in providing financial incentives for performance,⁵ (4) predictive models using EHR data for hospital readmission,⁶ and (5) clinical applications that test the reliability of EHR data in electronic surveillance systems to detect urinary tract infections⁷ and for improving diagnosis of gastroesophageal reflux in infants.⁸

Few studies have examined reliability of perinatal measures, and those reported have done so in large administrative databases using predominantly birth certificate and hospital discharge data⁹ and have not used a comparative reference group.¹⁰ Studies that did compare EHR data to manually extracted data chose to descriptively report findings rather than quantitatively assess reliability.¹¹ A 2012 systematic review examined the quality of perinatal measures using administrative and population-based datasets (not EHR however) and reported those perinatal measures that are reliably captured within these sources of data.¹² As EHR data become more readily available for research purposes within perinatal research, there is a need to assess the reliability of these data obtained from automated EHR reports for variables of interest, as well as to determine the best methodology for these reliability assessments.

Objective

The purpose of our study is to assess the reliability of data extracted from an established EHR as compared with manually abstracted data for common variables used in perinatal research including mode of delivery, labor induction and augmentation, fetal presentation, and postpartum hemorrhage, using measurement of sensitivity, specificity, and Cohen's kappa. We aim to provide insight into the reliability of EHR data in perinatal research and make recommendations for the adoption of techniques that can be used to report reliability of EHR data used in clinical and practice-based decision making.

Methods

Our current study is a secondary analysis from a larger retrospective study examining women who gave birth at a large, multi-payer institution in the Pacific Northwest, United States.¹³ Women were included if they gave birth between January and September 2013 at the study institution. By design, only women who had their delivery data within the EHR available for data extraction and whose charts were considered to be relatively complete with less than 20% missing data were included. Women who had a 'break-glass' privacy feature enabled on their EHR were excluded (→ Fig. 1).

The study took place at a large, multi-payer institution that has been using the Epic EHR system since 2011. All providers who attend deliveries within the hospital system are trained to document within the EHR system. In the original study, variables were collected using standard, automated reporting tools available within the Epic EHR. Variables not able to be captured using these reporting tools were collected by manual chart abstraction, and several variables of interest were captured by both chart abstraction and by the automated extraction process. The capture of variables by two different data collection modalities, with manual chart abstraction as the "gold standard," allowed for reliability testing of the available EHR extraction tools. The principal investigator, a certified nurse midwife and content

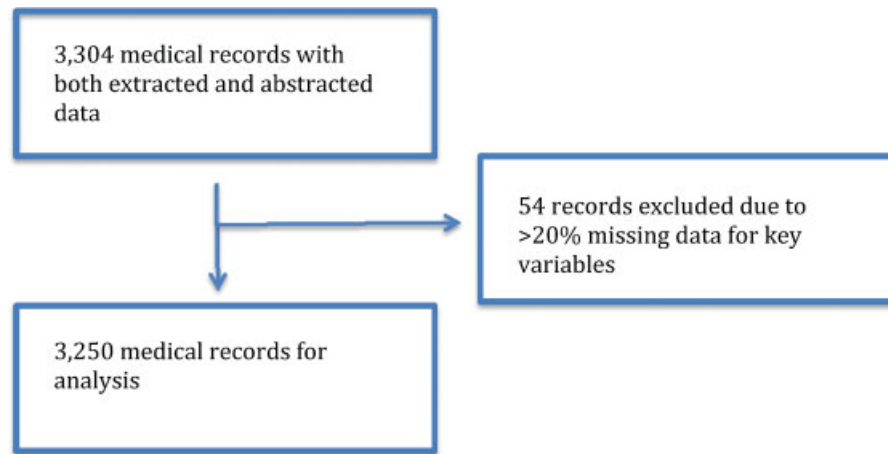


Fig. 1 Study chart depicting inclusion/exclusion criteria.

expert, performed the majority of the manual chart abstraction, with a subset of 10% of the medical records assessed by a separate content expert (an obstetrician) from the study institution for internal validity. A standardized data collection form was used and discrepancies were addressed. Concordance between data collectors was satisfactory for the study (Cohen’s kappa = 0.76), with discordance almost entirely due to variables incorrectly entered in multiple locations in the EHR.

Variables were chosen based on the parent study, which was a comparative analysis between different obstetrical providers for labor and delivery care in the hospital setting. Those variables collected by both modalities were chosen due to importance in the parent study (for inclusion criteria and outcomes of interest) and suspected inaccuracies in data capture using the standard reports. Manually abstracted data were obtained through free-text provider summaries, med-

ication administration records, and nursing documentation of intervention use, with secondary sources utilized in the case of missing data or for verification purposes (→Table 1).

Mode of delivery variables included spontaneous vaginal birth, cesarean delivery, vacuum-assisted delivery, and forceps-assisted delivery and were all defined as the method by which the woman delivered her infant. Labor induction was defined as the use of synthetic oxytocin (Pitocin) to initiate labor, labor augmentation was defined as the use of Pitocin to improve a labor that has already started spontaneously, and cervical ripening was defined as the use of any modality to prepare a woman’s cervix for induction prior to the initiation of Pitocin. Lastly, vertex presentation of the fetus was defined as the fetus presenting head down at the time of admission, and postpartum hemorrhage was defined as greater than 500cc estimated blood loss for a vaginal birth and greater than 1,000cc estimated blood loss for a cesarean delivery.

Table 1 Data abstraction processes

Variable of interest	Primary chart location	Secondary chart location
Cervical ripening (yes/no)	History and physical on admission (free text), medication administration record	Labor progress notes describing methods used (free text)
Labor induction (yes/no)	History and physical on admission (free text), medication administration record	Labor progress notes describing induction (free text)
Labor augmentation (yes/no)	Labor progress notes, medication administration record	N/A
Vertex presentation (yes/no)	History and physical on admission (form entry)	Delivery summary (free text), labor progress notes (free text)
Postpartum hemorrhage (yes/no)	Delivery summary (yes/no)	Delivery note (quantitative, transformed to yes/no)
Mode of delivery		
Vacuum assisted (yes/no)	Delivery summary	Delivery note (free text)
Forceps assisted (yes/no)	Delivery summary	Delivery note (free text)
Cesarean (yes/no)	Delivery summary	Delivery note (free text)
Spontaneous vaginal (yes/no)	Delivery summary, lack of other mode of delivery noted	Delivery note (free text)

Due to lack of a consistent methodology for assessment of reliability and quality of EHR data in the extant literature, we chose to use a combination of techniques: sensitivity, specificity, agreement, and Cohen's kappa. Using the manually abstracted data by content experts as the gold standard, sensitivities and specificities were calculated for the systematically extracted EHR data. However, given that there was a level of discordance between data abstractors due to the inherent structure of the EHR, we have also included agreement statistics to more fully and completely assess reliability. Cohen's kappa scores were calculated to assess agreement between the two data collection modalities. Kappa statistics approaching 1.0 indicate perfect agreement between data sources. A commonly cited scale was used to interpret kappa scores in this study, with almost perfect agreement as 0.81 to 0.99, substantial agreement as 0.61 to 0.80, moderate agreement as 0.41 to 0.60, and fair agreement as 0.21 to 0.40.¹⁴ All calculations were performed and confirmed using several modalities, including Excel, SPSS, and by hand calculations.

Results

There were a total of 3,304 medical records for which we had both extracted and abstracted data from the labor and delivery hospital stay during the study period, 3,250 of which included the variables of interest for this study (→ Fig. 1). Demographic and clinical characteristics of study participants are typical for a large hospital in the Pacific Northwest (→ Table 2). Average gestational age was over 39 weeks, with 97.6% of women with a singleton fetus and 95.0% with the fetus in vertex presentation. Nearly half of women were aged 20 to 29 with ~45% of women aged 30 to 9. While 58.0% of women were covered by commercial insurance, 40.8% of women were insured by Medicaid. Most women were White and non-Hispanic.

The reliability of key attributes related to perinatal measures varied (→ Table 3). Kappa statistics provide a measure of agreement between data extracted from the EHR and an expert's manual extraction. Almost perfect agreement was observed for all four mode of delivery variables (vacuum assisted kappa = 0.92; 95% confidence interval [CI] = 0.88–0.95, forceps assisted kappa = 0.90; 95%CI = 0.76–1.00, cesarean delivery kappa = 0.91; 95%CI = 0.90–0.93, and spontaneous vaginal delivery kappa = 0.91; 95%CI = 0.90–0.93). Additionally, the attribute of cervical ripening demonstrated substantial agreement (kappa = 0.77; 95%CI = 0.73–0.80). Induction (kappa = 0.65; 95%CI = 0.62–0.68) and augmentation (kappa = 0.54; 95%CI = 0.49–0.58) demonstrated moderate agreement between the two data sources. Finally, vertex presentation (kappa = 0.35; 95%CI = 0.31–0.40) and postpartum hemorrhage (kappa = 0.21; 95%CI = 0.13–0.28) demonstrated fair agreement. Additional reliability statistics presented varying agreement between data extracted from the EHR and an expert's manual extraction. Specificity was generally high, except for vertex presentation at 78.8% (95% CI = 72.4–85.1). However, sensitivity varied considerably. The lowest sensitivity observed in this study was the attribute of post-partum hemorrhage at 38.2% (95%CI = 26.7–49.8).

Table 2 Characteristics of study sample (N = 3,250)

		Mean (SD)
Maternal age (y)		29.5 (5.6)
Gestational age (wk)		39.1 (2.4)
		n (%)
Maternal age (y)	16–19	75 (2.3)
	20–29	1,573 (48.4)
	30–39	1,454 (44.8)
	40+	148 (4.6)
Primary payer	Commercial	1,885 (58.0)
	Medicaid	1,327 (40.8)
	Unknown	16 (0.5)
	Medicare	22 (0.7)
Maternal race	White	2,293 (70.5)
	Black	103 (3.2)
	Asian	210 (6.5)
	American-Indian	76 (2.3)
	Hawaiian islander	34 (1.0)
	Other	470 (14.5)
	Mixed, refused, unknown	64 (2.0)
Maternal ethnicity	Not Hispanic or Latino	2,830 (87.1)
	Hispanic or Latino	414 (12.8)
	Unknown, refused	6 (0.2)
Marital status	Married	2,017 (62.1)
	Single	1,112 (34.2)
	Divorced or separated	107 (3.3)
	Significant other	11 (0.3)
	Widowed	2 (0.1)
	Unknown	1 (0.0)
Maternal BMI classification	Recommended (<25 kg/m ²)	701 (22.0)
	Overweight (25–30 kg/m ²)	1,060 (33.2)
	Obese (>30 kg/m ²)	1,429 (44.8)
Vertex presentation		3,090 (95.0)
Singleton fetus		3,174 (97.6)

Abbreviations: BMI, body mass index; CI, confidence interval.

Discussion

To perform rigorous research using EHR data, it is imperative that we are able to trust the quality of what is captured and reported through EHR mechanisms. Given the importance of assessing the reliability of the EHR data utilized in our parent

Table 3 Reliability of key attributes of labor and delivery research ($N = 3,250$)

Data element	Kappa	95% CI	Agreement %	Sensitivity %	95% CI	Specificity %	95% CI
Cervical ripening	0.77	0.73–0.80	94.4	69.3	65.3–73.3	99.1	98.8–99.5
Induction	0.65	0.62–0.68	88.0	59.7	56.4–62.9	98.2	97.7–98.8
Augmentation	0.54	0.49–0.58	90.0	52.7	48.1–57.3	96.0	95.3–96.7
Vertex presentation	0.35	0.31–0.40	88.2	88.6	87.5–89.8	78.8	72.4–85.1
Post-partum hemorrhage	0.21	0.13–0.28	94.7	38.2	26.7–49.8	95.9	95.2–96.6
Delivery method							
Vacuum assisted	0.92	0.88–0.95	99.4	94.2	90.1–98.4	99.6	99.4–99.8
Forceps assisted	0.90	0.76–1.00	99.9	100.0	100.0–100.0	99.9	99.9–100.0
Cesarean	0.91	0.90–0.93	96.8	88.2	86.1–90.4	99.8	99.7–100.0
Spontaneous vaginal	0.91	0.90–0.93	96.2	95.2	94.3–96.1	98.4	97.6–99.2

Abbreviation: CI, confidence interval.

study, which documented the interventions, resources, and costs for care by providers during labor and birth,¹³ we were able to compare extractions of automated EHR report data to manual chart abstraction completed by content experts to assess the reliability of automated reports. While mode of delivery captured by the EHR was satisfactorily reliable in our sample, other perinatal measures, such as labor induction, labor augmentation, cervical ripening, fetal presentation, and presence of postpartum hemorrhage, were less so when compared with manual chart abstraction. Such variability in reliability of variables captured in the EHR for obstetrical research implies a serious limitation in how these data can be used, as well as a need for reporting reliability and validity measures whenever perinatal, and likely other, EHR data are used for research purposes.

The variability in reliability across perinatal measures may also reflect the nature of the phenomena being captured by the variable. For example, those variables that have very clear-cut categories that have value outside of clinical decision making (such as mode of delivery) tended to be the most reliable, as compared with variables that may reflect processes of care (labor induction, augmentation, or cervical ripening), clinically important but often assumed variables (vertex presentation), and those variables requiring a level of interpretation by provider (postpartum hemorrhage). Other potential reasons for variability in reliability may be due to poor reporting and missing data in the EHR, variability in entering clinical data (i.e., multiple entry points on different forms), or variables that required interpretation by clinical providers.¹¹ Our findings support what was reported in the 2012 systematic review examining similar measures,¹² suggesting measures that are publicly reported for quality or financial initiatives have greater reliability. We believe this may reflect the organizational attention focused on the documentation and reporting of variables that are publicly reported. When facilities are invested in the reporting of clinical parameters to meet such criteria as Meaningful Use, clinicians are trained and repeatedly reminded about the

importance of documenting those variables, and reports are tested and verified to determine appropriate capture. The same rigorous level of attention and focus is not always applied to non-reportable variables, which may create problems for using these data for research purposes.

Our work supports a need for reliability assessment of EHR data, both in perinatal research as well as across other disciplines. Other areas of clinical research more frequently utilize and report a variety of statistical tests to assess the reliability of extracted EHR data. Studies commonly report (1) Cohen's kappa;^{15–19} (2) measures of test performance, such as sensitivity²⁰ and specificity,^{7,18,19,21–25} positive predictive value (PPV),^{7,19,20,23,25,26} and negative predictive value (NPV);^{19,23} (3) the area under the curve (AUC);⁷ (4) regression;^{24,27} and (5) with simple agreement indices.^{10,28–31} Due to lack of a consistent methodology for assessment of reliability and quality of EHR data in the extant literature, we chose to use a combination of techniques: sensitivity, specificity, agreement, and Cohen's kappa. The use of multiple modalities for assessing reliability and quality is a methodological strength in the current work. By internally validating the reliability of the manually abstracted data used as the gold standard, we also provided a rigorous reference for comparison. Encouraged by the call to report kappa scores as the new standard in EHR research,¹⁸ we evaluated our extracted versus abstracted data with multiple statistical analysis techniques to increase the trustworthiness of our work, to introduce the concept of reliability testing to perinatal researchers and to make our work more readily generalizable.

We do acknowledge that by the use of an EHR within a single institution, we are limited to interpretation of reliability of obstetrical data within this particular EHR; however, our methods to assess reliability could be transferable across EHR systems and within other clinical areas. Data entered into the EHR for obstetrical care may differ from other fields of inquiry given that data are used for clinical, not research, purposes. The potential for misclassification and underreporting of outcomes is also inherent in these types of data, for

which we had to include only variables that were relatively complete in reporting. A large limitation to our study is that our study is a secondary analysis, hence reliant on the variables chosen for the parent study and not necessarily examining variables of importance for research purposes. Limitations within the parent study include that the principal investigator conducted chart abstraction with a limited number of chart abstractions replicated by a second reviewer, introducing the potential for bias. Despite these limitations, our study presents a rigorous approach to assessing reliability of EHR data in perinatal research.

We have identified several implications from our study to guide future research. Such variability in reliability of variables captured in the EHR for perinatal research implied a serious limitation in how these data can be used. As such, studies that use EHR data should have an assessment of reliability as part of their findings. Until EHR reporting mechanisms are better tailored to accurately capture variables of interest, reliability and validity measures are crucial for determining the trustworthiness of the findings reported. On that note, there is a need for continued improvement of data capture mechanisms within available EHR systems, specifically for those variables that may be important for research purposes. Standardization of forms, data capture techniques, and use of advanced analytic processing tools may help with the capture of variables often found in free-text formats. Lastly, we have demonstrated the use of an innovative combination of statistical measures to comprehensively assess reliability of variables captured in the EHR; using sensitivity, specificity, and Cohen's kappa. These analyses in combination provide a rigorous assessment of reliability that can be used not only in perinatal research, but across different areas of inquiry.

Conclusion

With increasing use of EHR data for perinatal research purposes, there is a need to assess the quality of the data retrieved from EHRs. We support the call for more rigorous, quantitative techniques for the routine assessment of data extracted from the EHR. In our assessment reliability of commonly used perinatal variables, we have found variability in specificity, sensitivity, and Cohen's kappa scores indicating a need for better capture of variables within the standardized reporting EHR tools, as well as more rigorous assessments of reliability while using standardized EHR data in perinatal research.

Future work on assessing and improving the quality of data from the EHR can take many forms. Quality improvement projects can address documentation issues, but more advanced data science approaches are likely to be more helpful in the long term. Hospital mergers and acquisitions of smaller centers and practices are continuous, as is the routine replacement and upgrading of EHR, clinical, and administrative software used by healthcare facilities. This necessitates the consolidation of disparate data sources in data lakes and managing traditional reporting functions from a single source in innovative ways.

Clinical Relevance Statement

Our study reports variability in reliability for variables captured by EHR report mechanisms in a perinatal hospital setting, limiting the ability to use these data for research and other purposes. Efforts to improve reliability of EHR data should start in the clinical setting, with standardized data entry systems for providers, improved variable capture using automation or advanced data analytics, and informatics support for continuous reliability and validity assessments of clinical data collected for research.

Multiple Choice Questions

1. The following variable types have been found to be consistently reliable as documented in the EHR in obstetrical research:

- Mode of delivery
- Labor interventions
- Maternal outcome variables
- None of the above

Correct Answer: The correct answer is a, mode of delivery.

2. Given the lack of standards in assessing reliability of EHR data, our study used a combination of the following methods:

- Content analysis, area under the curve
- Sensitivity, specificity, and Cohen's kappa
- Negative predictive value, positive predictive value, area under the curve
- Multiple regressions, Cohen's kappa

Correct Answer: The correct answer is b, sensitivity, specificity, and Cohen's kappa.

Protection of Human and Animal Subjects

This study has been approved by the academic research institution's Institutional Review Board as well as by the study healthcare institution.

Conflict of Interest

None.

References

- The American Presidency Project. Executive Order 13335 Incentives for the Use of Health Information Technology and Establishing the Position of the National Health Information Technology Coordinator. 2004. Available at: <https://www.gpo.gov/fdsys/pkg/WCPD-2004-05-03/pdf/WCPD-2004-05-03-Pg702.pdf>. Accessed July 25, 2017
- Henry J, Pylypchuk Y, Searcy T, Patel V. Adoption of Electronic Health Record Systems among US Non-Federal Acute Care Hospitals: 2008–2015. The Office of National Coordinator for Health Information Technology; 2016
- Moen H, Peltonen L-M, Heimonen J, et al. Comparison of automatic summarisation methods for clinical free text notes. *Artif Intell Med* 2016;67:25–37
- Turner M, Barber M, Dodds H, Dennis M, Langhorne P, Macleod M-J; Scottish Stroke Care Audit. Agreement between routine electronic hospital discharge and Scottish Stroke Care Audit (SSCA) data in identifying stroke in the Scottish population. *BMC Health Serv Res* 2015;15(01):583

- 5 Urech TH, Woodard LD, Virani SS, Dudley RA, Lutschg MZ, Petersen LA. Calculations of financial incentives for providers in a pay-for-performance program: Manual review versus data from structured fields in electronic health records. *Med Care* 2015;53(10):901–907
- 6 Futoma J, Morris J, Lucas J. A comparison of models for predicting early hospital readmissions. *J Biomed Inform* 2015;56:229–238
- 7 Branch-Elliman W, Strymish J, Kudesia V, Rosen AK, Gupta K. Natural language processing for real-time catheter-associated urinary tract infection surveillance: results of a pilot implementation trial. *Infect Control Hosp Epidemiol* 2015;36(09):1004–1010
- 8 Funderburk A, Nawab U, Abraham S, et al. Temporal association between reflux-like behaviors and gastroesophageal reflux in preterm and term infants. *J Pediatr Gastroenterol Nutr* 2016;62(04):556–561
- 9 Hure AJ, Chojenta CL, Powers JR, Byles JE, Loxton D. Validity and reliability of stillbirth data using linked self-reported and administrative datasets. *J Epidemiol* 2015;25(01):30–37
- 10 Knake LA, Ahuja M, McDonald EL, et al. Quality of EHR data extractions for studies of preterm birth in a tertiary care center: guidelines for obtaining reliable data. *BMC Pediatr* 2016;16(01):59
- 11 Flood M, Small R. Researching labour and birth events using health information records: methodological challenges. *Midwifery* 2009;25(06):701–710
- 12 Lain SJ, Hadfield RM, Raynes-Greenow CH, et al. Quality of data in perinatal population health databases: a systematic review. *Med Care* 2012;50(04):e7–e20
- 13 Altman MR, Murphy SM, Fitzgerald CE, Andersen HF, Daratha KB. The cost of nurse-midwifery care: use of interventions, resources, and associated costs in the hospital setting. *Womens Health Issues* 2017;27(04):434–440
- 14 Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med* 2005;37(05):360–363
- 15 Alsara A, Warner DO, Li G, Herasevich V, Gajic O, Kor DJ. Derivation and validation of automated electronic search strategies to identify pertinent risk factors for postoperative acute lung injury. *Mayo Clin Proc* 2011;86(05):382–388
- 16 Smischney NJ, Velagapudi VM, Onigkeit JA, Pickering BW, Herasevich V, Kashyap R. Derivation and validation of a search algorithm to retrospectively identify mechanical ventilation initiation in the intensive care unit. *BMC Med Inform Decis Mak* 2014;14:55
- 17 Chandra S, Agarwal D, Hanson A, et al. The use of an electronic medical record based automatic calculation tool to quantify risk of unplanned readmission to the intensive care unit: a validation study. *J Crit Care* 2011;26(06):634.e9–634.e15
- 18 Kahn MG, Eliason BB, Bathurst J. Quantifying clinical data quality using relative gold standards. *AMIA Annual Symposium Proceedings* 2010;2010:356–360
- 19 Makam AN, Nguyen OK, Moore B, Ma Y, Amarasingham R. Identifying patients with diabetes and the earliest date of diagnosis in real time: an electronic health record case-finding algorithm. *BMC Med Inform Decis Mak* 2013;13(01):81
- 20 Penman-Aguilar A, Tucker MJ, Groom AV, et al. Validation of algorithm to identify American Indian/Alaska Native pregnant women at risk from pandemic H1N1 influenza. *Am J Obstet Gynecol* 2011;204(06, Suppl 1):S46–S53
- 21 Singh B, Singh A, Ahmed A, et al. Derivation and validation of automated electronic search strategies to extract Charlson comorbidities from electronic medical records. *Mayo Clin Proc* 2012;87(09):817–824
- 22 Tien M, Kashyap R, Wilson GA, et al. Retrospective derivation and validation of an automated electronic search algorithm to identify post operative cardiovascular and thromboembolic complications. *Appl Clin Inform* 2015;6(03):565–576
- 23 Carroll RJ, Thompson WK, Eyler AE, et al. Portability of an algorithm to identify rheumatoid arthritis in electronic health records. *J Am Med Inform Assoc* 2012;19(e1):e162–e169
- 24 Lapham GT, Rubinsky AD, Shortreed SM, et al. Comparison of provider-documented and patient-reported brief intervention for unhealthy alcohol use in VA outpatients. *Drug Alcohol Depend* 2015;153:159–166
- 25 Luef BM, Andersen LB, Renåult KM, Nohr EA, Jørgensen JS, Christesen HT. Validation of hospital discharge diagnoses for hypertensive disorders of pregnancy. *Acta Obstet Gynecol Scand* 2016;95(11):1288–1294
- 26 Herrett E, Gallagher AM, Bhaskaran K, et al. Data resource profile: clinical practice research datalink (CPRD). *Int J Epidemiol* 2015;44(03):827–836
- 27 Selby NM, Crowley L, Fluck RJ, et al. Use of electronic results reporting to diagnose and monitor AKI in hospitalized patients. *Clin J Am Soc Nephrol* 2012;7(04):533–540
- 28 Ndira SP, Rosenberger KD, Wetter T. Assessment of data quality of and staff satisfaction with an electronic health record system in a developing country (Uganda): a qualitative and quantitative comparative study. *Methods Inf Med* 2008;47(06):489–498
- 29 Aronsky D, Haug PJ. Assessing the quality of clinical data in a computer-based record for calculating the pneumonia severity index. *J Am Med Inform Assoc* 2000;7(01):55–65
- 30 Jiggins K. A content analysis of the meaningful use clinical summary: do clinical summaries promote patient engagement? *Prim Health Care Res Dev* 2016;17(03):238–251
- 31 Sammon CJ, Miller A, Mahtani KR, et al. Missing laboratory test data in electronic general practice records: analysis of rheumatoid factor recording in the clinical practice research datalink. *Pharmacoepidemiol Drug Saf* 2015;24(05):504–509