



Contents lists available at ScienceDirect

Saudi Pharmaceutical Journal

journal homepage: www.sciencedirect.com



Original article

Measuring the quality and completeness of medication-related information derived from hospital electronic health records database

Monira Alwhaibi^{a,b,*}, Bander Balkhi^{a,b}, Thamir M. Alshammari^{b,c,d}, Nasser AlQahtani^c, Mansour A. Mahmoud^b, Mansour Almetwazi^{a,b}, Sondus Ata^e, Mada Basyoni^f, Tariq Alhawassi^{a,b,e}

^a Department of Clinical Pharmacy, College of Pharmacy, King Saud University, Riyadh, Saudi Arabia

^b Medication Safety Research Chair, College of Pharmacy, King Saud University, Riyadh, Saudi Arabia

^c Saudi Food and Drug Authority, Riyadh, Saudi Arabia

^d Department of Clinical Pharmacy, College of Pharmacy, University of Hail, Hail, Saudi Arabia

^e Pharmacy Care Services, King Saud University Medical City, Riyadh, Saudi Arabia

^f Saudi Patient Safety Centre, Riyadh, Saudi Arabia

ARTICLE INFO

Article history:

Received 25 May 2018

Accepted 14 January 2019

Available online 17 January 2019

Keywords:

Electronic Health Records

Medication

Completeness

Secondary data

ABSTRACT

Objective: Electronic Health Records (EHRs) database is a great source for pharmacoepidemiological research as thousands of patients' clinical and medication information is stored in the database. However, the use of EHRs database for research purposes depends greatly on the accuracy and completeness of the data being used. This study mainly aimed to assess the completeness of EHRs patients' medication-related information.

Design: A retrospective cross-sectional study using data extracted from the EHRs database was conducted.

Setting: The EHRs data was obtained from a single tertiary hospital in Saudi Arabia.

Main outcome measure(s): The completeness of data was measured considering if a patients' record contains all desired types of data (i.e., patients' demographics, clinical diagnosis, and medication-related information).

Results: A total of 23,411 unique individuals were identified after extracting the data from the EHRs. The study found that 89.9% of the patients had a complete data (i.e., age, gender, marital status, nationality, encounter type, and clinical diagnosis). Further, 83.1% of the patients had complete medication-related information. Subgroup analysis by the encounter type indicated that the data was 91.0% complete for outpatient encounter and 93.2% complete for inpatient encounter.

Conclusion: The study findings indicate that the completeness of the data varies by the desired types of data. EHRs can be a potentially great resource to conduct research to assess medication use. Further studies focusing on the content and completeness of EHRs for a specific patient population and evaluate other dimensions of EHRs data quality are needed.

© 2019 The Authors. Production and hosting by Elsevier B.V. on behalf of King Saud University. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Electronic Health Records (EHRs) are a patient-centered record containing the patient's medical and treatment information.

* Corresponding author at: Department of Clinical Pharmacy, King Saud University, School of Pharmacy, Riyadh 11149, Saudi Arabia.

E-mail address: malwhaibi@ksu.edu.sa (M. Alwhaibi).

Peer review under responsibility of King Saud University.



Recently, the use of traditional paper-based health records has been replaced by EHRs database in many hospitals in Saudi Arabia. The transition from paper-based health records to EHRs database opened up a unique opportunity for epidemiological research, medication use, practice surveillance, and quality assessment research that can improve the overall quality of care provided to patients (Coorevits et al., 2013).

Studies from different countries have reported that EHRs have been used for assessing physicians performance (Linder et al., 2009), quality of care (Cebul et al., 2011), predicting readmission (Shadmi et al., 2015), medication use (Castro et al., 2013), and inappropriate prescribing (Buck et al., 2009). EHRs can facilitate the assessment of many important patient health outcomes for

<https://doi.org/10.1016/j.jsps.2019.01.013>

1319-0164/© 2019 The Authors. Production and hosting by Elsevier B.V. on behalf of King Saud University.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

researchers, such as readmission and medication use (Ambinder, 2005). EHRs has also deemed a tool in proposing unique hypotheses for health policy researchers to start with. EHRs help researchers instantly access the data for thousands of patients with inexpensive cost (Häyrinen et al., 2008). Thus, the National Institutes of Health (NIH) emphasized the importance of conducting clinical research based on EHRs database (Weiskopf and Weng, 2013). The use of EHRs for research purpose is one of the “meaningful use” objectives of the Health Information Technology for Economic and Clinical Health Act (HITECH), which was enacted as part of the American Recovery and Reinvestment Act of 2009 to improve the healthcare delivery system (Blumenthal and Tavenner, 2010).

EHRs provide a more reliable source than traditional paper-based medical records or claims data, as it provides full clinical structured, coded data and unstructured narrative data (Coorevits et al., 2013). For example, EHRs clinical diagnosis codes are more accurate and reliable source for identifying chronic conditions as compared to paper-based medical records (Woodfield and Sudlow, 2015). Also, EHRs database provides more comprehensive data regarding patients' diagnoses, visits, laboratory tests, prescriptions, physical examination findings, and other health services received (DeVoe et al., 2011; Häyrinen et al., 2008). However, the use of EHRs database for research purpose depend on experience with using EMR data for research, confidentiality, data security concerns, technical issues and costs the completeness of the data (van Velthoven et al., 2016).

Data completeness is a structured and documented process performed to ensure that any database is complete for its intended use (Menachemi and Collum, 2011). Different conceptualizations in defining the completeness of EHRs database exist. Weiskopf et al. in their literature review identified four definitions of EHRs database completeness (Weiskopf and Weng, 2013). Data are considered complete if a patient record contains all observations made during a clinical encounter (i.e., documentation), which rely upon the presence of a reference standard, such as paper-based health records. Also, data are considered complete if a patient record contains all desired types of data (i.e., breadth), contains a specified number or frequency of data points over time (i.e., density), or has sufficient information to predict an outcome of interest (i.e., predictive).

Further, there are different dimensions of data quality such as completeness, correctness, concordance, plausibility, and currency (Chan et al., 2010; Weiskopf and Weng, 2013). Measuring the completeness of EHRs database is an important understudied area of research (Weiskopf and Weng, 2013). In Saudi Arabia, information on the content and completeness of EHRs databases is unavailable. Therefore, our study was the first study to assess the content and completeness of the EHRs database launched in a tertiary teaching hospital in Saudi Arabia. This hospital implemented the use of EHRs in 2015 to replace the traditional paper-based health records. As completeness of the data is a typical concern about EHRs database (Weng et al., 2012), this study is conducted to justify the use of the EHRs database for the clinical researcher. This research is part of several ongoing studies that provide information for researchers about the degree of data completeness. Findings of this study can help develop reliable information using an adequate amount of records for a large number of patients. Therefore, this study aimed to assess if the database has enough sample size to conduct pharmacoepidemiological research and to evaluate the content and completeness of the data provided in the EHRs database.

2. Methods

2.1. Study design

A retrospective cross-sectional study design was used. Data were extracted from the EHRs database from January 1 to June

30, 2016. The Institutional Review Board (IRB) of a tertiary teaching hospital approved the study with protocol number (16/2109/IRB).

2.2. Data source and data extraction

The current study used six-month data retrieved from the EHRs database launched by a tertiary teaching hospital. This hospital is one of the largest tertiary teaching hospitals in Riyadh, Saudi Arabia, with an 850-bed facility and all general and subspecialty medical services. The hospital provides primary, secondary, and tertiary care services. The patient population is composed predominantly of local citizens, from the northern region in Riyadh; the hospital also serves the entire country as a referral center.

Data were extracted from the EHR database by trained researcher pharmacists and were exported to an Excel sheet into multiple files (demographics file, clinical diagnosis file, and prescription drug file). The demographics file contained information about the patients' date of birth, sex, marital status, nationality, and encounter type. The clinical diagnosis file provided information about the clinical diagnosis from inpatient and outpatient visits and the date of clinical diagnosis. Physicians reported the clinical diagnosis data using the *International Classifications of Diseases – 9th edition, Clinical Modification* (ICD-9-CM) codes, *International Classifications of Diseases – 10th edition, Clinical Modification* (ICD-10-CM) codes, or the *Systematized Nomenclature of Medicine* (SNOMED) diagnostic codes. The physicians are not required to report all the codes for a single diagnosis; rather some physicians prefer one coding system to the other. Therefore, the mapping was conducted by a data scientists using SAS software, for example, all the following codes were used to identify hypertension (ICD-9 code: 401.9, ICD-10 codes: I10, I15, SNOMED codes: 64176011, 2164904016). The prescription drug file contained information from the inpatient and outpatient pharmacy about the medications' name, strength, dose, dosage form, quantity dispensed, and dispensing date. After extracting the data for each file, multiple patient observations were converted to one observation per patient to facilitate the analysis. Then, the demographics, clinical diagnosis, and prescription drug files were merged into one file using the encrypted patient medical record number.

2.3. Data confidentiality

Confidentiality of the data was maintained throughout the research process. Retrieved data were stored and saved as coded excel files. A customized formula was used to generate study encrypted identification numbers assigned to each participant and replaced patients' medical record number. Data extracted were stored in a secure, password protected, and limited accessed computers.

2.4. Study population

The study population composed of all patients who visited a tertiary teaching hospital between January 1 and June 30, 2016.

2.5. Measuring the completeness of EHRs data

Two independent researchers assessed the missing data. This study used Weiskopf et al. in defining the completeness. The patient record was considered complete if it contained all desired types of information needed for research purposes (i.e., breadth) (Weiskopf & Weng, 2013). Based on this definition, EHRs data were considered complete if they have the desired types of data to conduct research, such as patients' record including the date of birth (age in years), gender, marital status, nationality, encounter type, and clinical diagnosis. A patient with all six types of data present

would be considered having a complete data. In addition, completeness of medication-related information was measured using seven types of data; data were considered complete if the patients' record included all the above types of data in addition to the medication used (name, dosage form, dose, dispensed quantity, and dispensing date).

2.6. Data analysis

Descriptive analysis was performed to describe the data. Mean and standard deviation were used to describe continuous variables. Frequency and percentage were used to describe categorical variables. Percentage of missing variables and incomplete records were calculated. Subgroup analysis was conducted to assess the completeness of EHRs data by encounter type. All statistical analyses were performed using the Statistical Analysis Software (SAS[®], version 9.2).

3. Results

3.1. Description of data

A total of 183,948 patients' records were identified during the six-month period, and after removing the duplicates, the final study population was 23,411 unique individuals identified (Appendix 1.1). Table 1 shows the description of the study population. One-fourth of the study population was older adults (i.e., ≥60 years). The majority of the study population was women (67.5%), and 83% of the encounter type was outpatient visits.

3.2. Completeness of the data

Among the study population, data for patients' age, gender, and clinical diagnosis were available for 100% of the study population. However, data was missing for the marital status (7.0%), nationality (0.4%), encounter type (3.2%), and medication used (7.3%) (Table 2). It was found that 89.9% of the study population had a complete data based on the six types of data definition (i.e., age, gender, marital status, nationality, encounter type, and clinical diagnosis) (Table 3, Chart 1). When the medication-related infor-

Table 1
Description of the electronic health record data.

	N	%
Total	23,411	100.0
Age group		
<18	2497	10.7
18–29	2917	12.5
30–39	3459	14.8
40–49	3639	15.5
50–59	5131	21.9
≥60	5768	24.6
Marital status		
Single	7068	32.5
Married	14,698	67.5
Gender		
Male	8476	36.2
Female	14,934	63.8
Nationality		
Saudi	19,536	83.8
Non-Saudi	3783	16.2
Encounter type		
Outpatient	18,831	83.1
Inpatient	3837	16.9

Note: Based on 23,411 unique individuals identified during a period of six months. N: Number; %: Percentage.

Table 2
Percentage of missing data for the completeness variables.

	Available %	Missing N (%)
Age group	100.0	0 (0.0)
Marital status	93.0	1647 (7.0)
Gender	100.0	0 (0.0)
Nationality	99.6	92 (0.4)
Encounter type	96.8	745 (3.2)
Chronic conditions	100.0	0 (0.0)
Medication use	92.7	1705 (7.3)

Note: Based on 23,411 unique individuals identified during a period of six months. N: Number; %: Percentage.

Table 3
Completeness of data based on definition used.

	Complete %	Incomplete %
No missing information on six types of data*	89.9	10.1
No missing information on seven types of data**	83.1	16.9
Outpatient encounter only	91.0	6.8
Inpatient encounter only	93.2	9.1

* Including age, gender, marital status, nationality, encounter type, and clinical diagnosis.

** Including age, gender, marital status, nationality, encounter type, clinical diagnosis, and prescription drug.

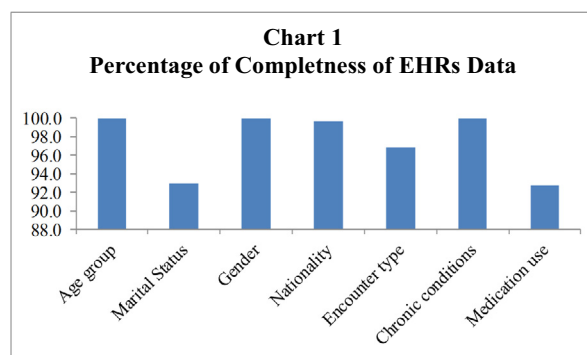


Chart 1. Percentage of completeness of EHRs data.

mation was added to define EHRs completeness (i.e., age, gender, marital status, nationality, encounter type, clinical diagnosis, and medication use), data were complete for 83.1% of the patients. Subgroup analysis by the encounter type indicated that the data was 91.0% complete for outpatient encounter and 93.2% complete for the inpatient encounter. The overall completeness rate is lower than the rate by encounter type, and this is attributed to that some individuals have missing information for the encounter type, the encounter type is missing for 743 unique individuals.

4. Discussion

This study investigated the content and completeness of the EHRs database among patients who visited a tertiary teaching hospital during a 6-month period. Our findings indicated that the EHRs database contained the desired types of data required for conducting pharmacoepidemiological research (i.e., demographics, clinical diagnosis, and medication use).

We also observed that the completeness of data varies based on the operational definition used. For example, when we defined completeness as patient records containing no missing information for seven variables, including (age, gender, marital status, nationality, encounter type, clinical diagnosis, and prescription drugs), the com-

pleteness was 83.1%. However, when we considered having patients with no missing information for six variables (age, gender, marital status, nationality, encounter type, and clinical diagnosis), the completeness was 89.9%. The percentage of uncompleted data was mainly driven by the missing information on the marital status (7.0% missing) and encounters type (3.2% missing). We can speculate that healthcare providers may not consider reporting these variables as important as the clinical diagnosis or other types of data. Also, these fields are not mandatory to report in the EHRs database. However, having a missing data on marital status and encounter type does not affect patient care or clinical research. Further, patients who received their clinical services from the inpatient had a higher complete data as compared to those who received their clinical services from the outpatient (93.2% vs. 91.0%). Healthcare providers may have more time in the inpatient setting to record patients' information as compared to the outpatient setting.

Regarding the completeness of medication-related information, only 7.3% have missing medication use data. This finding is not surprising since physicians are required to use the computerized physician order enters to complete the medication data entry. This rate is consistent with those already published in the literature among Canadian population; Hong et al. reported that 7% of EHRs have incomplete medication-related information (Hong et al., 2015). The 7.3% incompleteness rate of medication use may be attributed to the nature of healthcare services provided to certain patients. For instance, surgical patients in the one-day surgery clinics with minor procedures may not require any pharmacological interventions upon admission. Moreover, patients who may have been seen by clinical practitioners and had no ailments to be confirmed at the visit, the system allows to only document the visit as a medical note, but prescribers cannot place any medication order without a documented diagnosis. These assumptions may need a validation using the primary data type of research to assess the potential risk factors in the incompleteness of medication use data. Studies had documented a more complete data on the medications when physicians used the computerized patient order as compared to the traditional paper records (Tang et al., 1999). Studies that attempted to determine the completeness of medication lists in traditional paper-based health records, in terms of medication name, dose, route, and frequency recorded in outpatient notes, admission notes, and discharge summaries have found that 18.6% of the medication lists are missing (Owen et al., 2011).

Further, the completeness of data varies according to the types of data used, which has been reported by previous studies (Weiskopf and Weng, 2013). Weiskopf et al. found that the percentage of completeness varies based on types of data included. They defined the completeness in their study as the presence of five types of data frequently found in patients' records (i.e., laboratory results, medication orders, diagnoses, sex, and date of birth). They found that only 11.8% of patients had complete EHR data. We were unable to compare our completeness of data estimates to other EHRs database in Saudi Arabia since there are no published data from other EHRs databases in Saudi Arabia.

4.1. Strengths and limitations

The present study has several advantages. To our knowledge, it is the first study in Saudi Arabia and the Middle East that assessed the completeness of the EHRs database. A large sample was used to examine the completeness of medication-related information and its suitability for research purposes using the EHRs. However, it also has some limitations. One observation was that healthcare providers used three different clinical diagnostic coding systems to report the patients' clinical diagnosis in the EHRs database; however, we were able to handle the three different coding systems in our analyses to code patients' clinical diagnosis. Besides,

as the study was conducted in only one hospital, the findings from this study cannot be generalized to the EHRs databases from other hospitals due to the difference in the study population. This database provides information for patients mainly from the Northern region in Riyadh, and therefore the completeness rate may differ in other EHRs databases. Hospitals in Saudi Arabia use different EHRs database and sixteen different health information systems were being used across the 19 hospitals that have adopted the EHRs (Hasanain et al., 2015). Therefore, there is no consistency in the type of health information system used for electronic health record between different hospitals and there might be variation in data entry which can include structured and free text data fields (Alnuem et al., 2011).

4.2. Future implications

The study findings provide a better understanding of missing data, which can help users improve the documentation that can support clinical care and research. Although the quality of information is particularly important in patient care, it is more important for researchers and policymakers. The use of EHRs holds promise for facilitating efficient retrospective pharmacoepidemiological research, which can be a foundation for advancing the clinical practice and therefore better quality of care. Finding from this study indicate that the EHRs data can be used to assess the quality and patient safety. The use of information from the EHRs has many potential benefits for decision makers to make the changes in the current healthcare system into one that safer, more efficient that can deliver high-quality care.

As this study was focused on the completeness of medication-related information, future studies need to identify data completeness among a specific population such as patients with cancer as cancer treatment such as surgery and radiation therapy are not in EHRs structured format and are usually available from EHRs physicians' notes (unstructured format). Therefore, researchers interested in the use of EHRs data for clinical research should consider the completeness of EHRs data quality and be aware of the task-dependence of data quality. Future studies are needed to assess other dimensions of EHRs data quality such as correctness, accuracy, currency, and validity.

5. Conclusion

The findings from this study indicate that the completeness of data varies based on the type of information required and the encounter type. As we mostly obtained a complete data on most of the demographics, clinical diagnosis, and medications use, the EHRs database can be utilized to conduct medication-related clinical research. The EHRs database also provided a large population data, which are collected over multiple time points. Further studies focusing on the content and completeness of EHRs for a specific patient population and evaluate other dimensions of EHRs data quality are needed.

6. Authors' contribution

All authors substantially contributed to the conception, design, acquisition, and interpretation of data. All authors have also participated in drafting and revising the manuscript and approved of the final version.

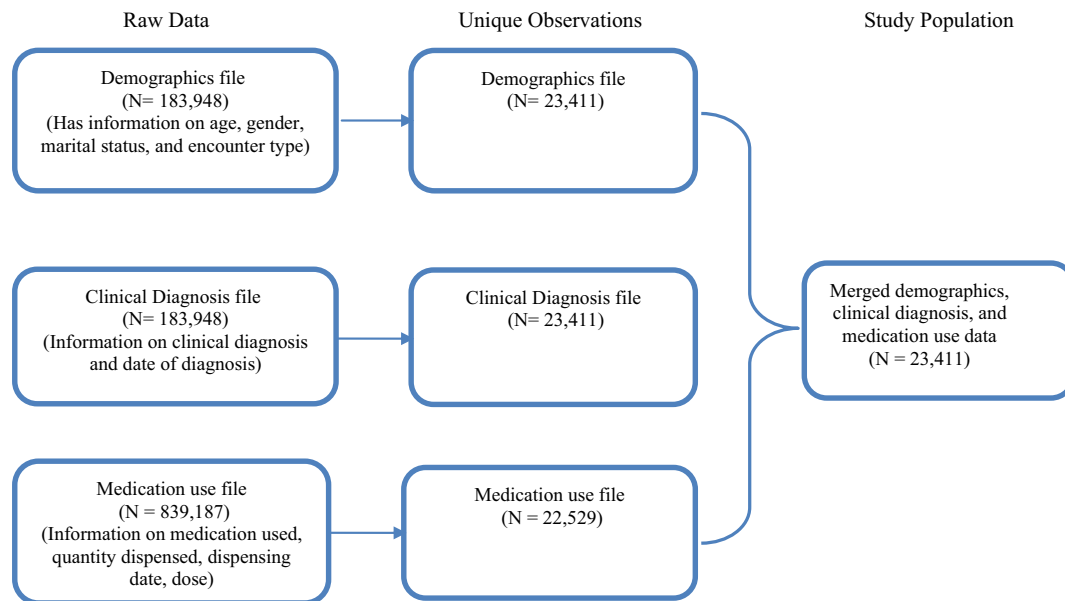
Acknowledgment

The project was fully supported financially by the Vice Deanship of Research Chairs, King Saud University Riyadh, Saudi Arabia.

Conflict of interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Appendix 1.1. Development of flow diagram for study population to assess the completeness of data.



References

- Alnuem, M., Samir, E.-M., Youssef, A., Emam, 2011. Towards integrating national electronic care records in Saudi Arabia. In: International conference on bioinformatics and computational biology.
- Ambinder, E.P., 2005. Electronic health records. *J. Oncol. Practice* 1 (2), 57–63.
- Blumenthal, D., Tavenner, M., 2010. The “meaningful use” regulation for electronic health records. *N Engl. J. Med.* 2010 (363), 501–504.
- Buck, M.D., Atreja, A., Brunker, C.P., Jain, A., Suh, T.T., Palmer, R.M., Wilcox, A.B., 2009. Potentially inappropriate medication prescribing in outpatient practices: prevalence and patient characteristics based on electronic health records. *Am. J. Geriatric Pharmacother.* 7 (2), 84–92.
- Castro, V.M., Clements, C.C., Murphy, S.N., Gainer, V.S., Fava, M., Weilburg, J.B., Iosifescu, D.V., 2013. QT interval and antidepressant use: a cross sectional study of electronic health records. *BMJ* 346, f288.
- Cebul, R.D., Love, T.E., Jain, A.K., Hebert, C.J., 2011. Electronic health records and quality of diabetes care. *N Engl. J. Med.* 365 (9), 825–833.
- Chan, K.S., Fowles, J.B., Weiner, J.P., 2010. Electronic health records and the reliability and validity of quality measures: a review of the literature. *Med. Care Res. Rev.* 67 (5), 503–527.
- Coorevits, P., Sundgren, M., Klein, G.O., Bahr, A., Claerhout, B., Daniel, C., Singleton, P., 2013. Electronic health records: new opportunities for clinical research. *J. Intern. Med.* 274 (6), 547–560.
- DeVoe, J.E., Gold, R., McIntire, P., Puro, J., Chauvie, S., Gallia, C.A., 2011. Electronic health records vs Medicaid claims: completeness of diabetes preventive care data in community health centers. *Ann. Family Med.* 9 (4), 351–358.
- Hasanain, R.A., Vallmuur, K., Clark, M., 2015. Electronic medical record systems in Saudi Arabia: knowledge and preferences of healthcare professionals. *J. Health Inform. Devel. Countries* 9 (1).
- Häyrynen, K., Saranto, K., Nykänen, P., 2008. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. *Int. J. Med. Inf.* 77 (5), 291–304.
- Hong, C.J., Kaur, M.N., Farrokhyar, F., Thoma, A., 2015. Accuracy and completeness of electronic medical records obtained from referring physicians in a Hamilton, Ontario, plastic surgery practice: a prospective feasibility study. *Plastic Surgery* 23 (1), 48–50.
- Linder, J.A., Kaleba, E.O., Kmetik, K.S., 2009. Using electronic health records to measure physician performance for acute conditions in primary care: empirical evaluation of the community-acquired pneumonia clinical quality measure set. *Med. Care* 47 (2), 208–216.
- Menachemi, N., Collum, T.H., 2011. Benefits and drawbacks of electronic health record systems. *Risk Manage. Healthc. Policy* 4, 47–55.
- Owen, M.C., Chang, N.M., Chong, D.H., Vawdrey, D.K., 2011. Evaluation of medication list completeness, safety, and annotations. Paper presented at the AMIA Annual Symposium Proceedings.
- Shadmi, E., Flaks-Manov, N., Hoshen, M., Goldman, O., Bitterman, H., Balicer, R.D., 2015. Predicting 30-day readmissions with preadmission electronic health record data. *Med. Care* 53 (3), 283–289.
- Tang, P.C., LaRosa, M.P., Gorden, S.M., 1999. Use of computer-based records, completeness of documentation, and appropriateness of documented clinical decisions. *J. Am. Med. Assoc.* 6 (3), 245–251.
- van Velthoven, M.H., Mastellos, N., Majeed, A., O’Donoghue, J., Car, J., 2016. Feasibility of extracting data from electronic medical records for research: an international comparative study. *BMC Med. Inf. Decis. Making* 16 (1), 90.
- Weiskopf, N.G., Weng, C., 2013. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *J. Am. Med. Assoc.* 20 (1), 144–151.
- Weng, C., Appelbaum, P., Hripcsak, G., Kronish, I., Busacca, L., Davidson, K.W., Bigger, J.T., 2012. Using EHRs to integrate research with patient care: promises and challenges. *J. Am. Med. Assoc.* 19 (5), 684–687.
- Woodfield, R., Sudlow, C.L., 2015. Accuracy of patient self-report of stroke: a systematic review from the UK biobank stroke outcomes group. *PLoS One* 10 (9), e0137538.