RESEARCH ARTICLE   **Open Access**

Methodological issues of the electronic health records' use in the context of epidemiological investigations, in light of missing data: a review of the recent literature

Thomas Tsiampalis[1] and Demosthenes Panagiotakos[1,2*]

Abstract

Background Electronic health records (EHRs) are widely accepted to enhance the health care quality, patient monitoring, and early prevention of various diseases, even when there is incomplete or missing information in them.

Aim The present review sought to investigate the impact of EHR implementation on healthcare quality and medical decision in the context of epidemiological investigations, considering missing or incomplete data.

Methods Google scholar, Medline (via PubMed) and Scopus databases were searched for studies investigating the impact of EHR implementation on healthcare quality and medical decision, as well as for studies investigating the way of dealing with missing data, and their impact on medical decision and the development process of prediction models. Electronic searches were carried out up to 2022.

Results EHRs were shown that they constitute an increasingly important tool for both physicians, decision makers and patients, which can improve national healthcare systems both for the convenience of patients and doctors, while they improve the quality of health care as well as they can also be used in order to save money. As far as the missing data handling techniques is concerned, several investigators have already tried to propose the best possible methodology, yet there is no wide consensus and acceptance in the scientific community, while there are also crucial gaps which should be addressed.

Conclusions Through the present thorough investigation, the importance of the EHRs' implementation in clinical practice was established, while at the same time the gap of knowledge regarding the missing data handling techniques was also pointed out.

Keywords Electronic health record, Healthcare quality, Medical decision, Missing data, Save cost and time

Introduction

Electronic Health Records (EHRs) constitute a challenging information system including a big, valuable collection of health information about patients' medical history and other related characteristics, both in structured and unstructured format. EHR have been implemented by an ever-increasing number of hospitals and research institutions around the world, as the mobile computing has been grown tremendously and the number of records

*Correspondence:
Demosthenes Panagiotakos
dbpanag@hua.gr
[1] Department of Nutrition and Dietetics, School of Health Sciences and Education, Harokopio University, Athens, Greece
[2] Faculty of Health, University of Canberra, Canberra, Australia

regarding personal health has been increasing exponentially [1]. According to the US Health Information Technology for Economic and Clinical Health Act (HITECH Act), in 2009, a spending exceeding $30 billion was authorized for the EHR adoption [2], with the EHR installations having been increased tremendously,between 2010 and 2014, the number of hospitals with a basic EHR system rose from 15.6% to 75.5% [3]. By 2025, the European Commission is looking to digitize all medical records throughout the 27-member bloc of European Union, to make it easier for individuals to access and share their personal data with medical professionals, particularly when they are in another country [4]. Moreover, EHR constitute a cornerstone of what is now called Real World Data, but this is a topic for another methodological review.

Several studies have already highlighted that EHRs may sufficiently improve the quality of healthcare, increase time efficiency and guideline adherence, and reduce medication errors and adverse drug effects [5–8]. At the same time, the use of EHRs in the medical decision process is rapidly growing, with an increasing number of researchers using them for the prognosis and early diagnosis of various chronic and non-chronic diseases [9]. An emerging literature has already recognized the challenges that still lay ahead in using EHRs' data in epidemiological research. The most crucial issue is the population representativeness included in EHRs (i..e, revealing the issue of selection bias), as well as the missing information in crucial clinical measurements and outcomes [10–14]. These issues are considered to be inevitable in real-world studies [15, 16], as their existence could be attributed to several reasons (e.g., refusal of patients to answer sensitive questions, lost- to follow- up, etc.). According to Bell et al., [17], as well as Little and Rubin [18], this can also lead to a substantial decrease in the efficiency and validity of the conducted data analyses and therefore, distort inferences about the referent population. Therefore, it is of crucial importance to identify the profile of the individuals with missing data, as well as to implement the right methodological approach, so as to impute the missing data and derive efficient and valid conclusions [19, 20].

The aim of the present review is to present the challenges faced during the use of the EHRs for epidemiological investigations in the context of missing data, as well as to discuss the most frequent statistical methodologies being implemented for handling such cases and confronting the obstacle of missing information to derive valid conclusions.

Material and methods
Eligibility criteria
Type of studies
The present review has been conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; [21]). Case studies, cohort studies, cross-sectional studies, retrospective case–control, prospective cohort, and cluster-randomized controlled trials, published in English language, either conducted in a hospital setting or not, were included in the present review, while systematic reviews and meta-analyses were excluded (but assisted in retrieving articles not allocated in search process).

Information sources and search strategy
Relevant studies, without any chronological and country restriction, were identified by searching in Medline (via PubMed), Scopus, and Google scholar databases by using the search strategies presented in Table 1. After removing the duplicate studies found among the different databases, articles were manually and independently

Table 1 Search strategies in each database for retrieving the most appropriate research works

Database	Search strategy
PubMed	**Search strategy 1-** (((("electronic health records") AND (healthcare quality)) AND (save)) AND (improve)) AND (patients); **Search strategy 2**- (((electronic health record) AND ((medical OR healthcare) AND decision) AND (missing data)) NOT (impact)) NOT (systematic review); **Search strategy 3**-(((("electronic health records missing data imputation") NOT (systematic review)) NOT (meta-analysis)) NOT (review)
Scopus	**Search strategy 1-** TITLE-ABS-KEY ("electronic health records" AND ((healthcare OR medical) AND (choice OR process OR decision OR order)) AND ((lost OR missing) AND data)); **Search strategy 2-** TITLE-ABS-KEY ("electronic medical records" AND ((healthcare OR medical) AND (choice OR process OR decision OR order)) AND ((lost OR missing) AND data) AND NOT (impact OR drawback)) AND (LIMIT-TO (SUBJAREA, "MEDI")); **Search strategy 3-** TITLE-ABS-KEY ("electronic health records" AND "missing data" AND (imputation OR generation) AND NOT (systematic AND review OR review)); **Search strategy 4-** electronic health record AND healthcare quality AND save cost time AND health system AND improve AND NOT systematic review AND NOT meta-analysis
Google Scholar	**Search strategy 1-** ("benefits" OR "advantage")("electronic health records")("healthcare quality")("medical decision" OR "medical choice"") ("missing data"); **Search strategy 2-** EHRs missing data medical decision quality advantage OR benefits "electronic health records" "missing data" -impact -affect –"systematic review"; **Search strategy 3-** electronic health records missing data imputation "electronic health records" -impact -drawback –"systematic review" -review –"meta analysis"

screened by both authors (TT, DP), based on their Title and Abstract and then full text reading was conducted for the final selection decision. In the case of disagreement, another scientist was asked to comment on the eligibility of the reviewed study.

Results
Study selection
Of the 1972 references initially identified from the electronic and manual search studies (PubMed: 313; Scopus: 519; Google scholar: 1140), a total of 17 studies were included in the present narrative review, which were divided in two categories:

i) studies related to the benefits of the EHRs implementation on medical quality and health system (e.g., cost- savings, reduced medical errors, improved emergency care etc.)
ii) studies related to the methodologies being implemented for imputing missing data in the context of the EHRs.

At first, 20 duplicate records were removed, and then the remaining 1,952 records were screened based on their title and abstract. From those, 1,897 records were removed due to irrelevance to the aim of the present review. Finally, 38 records were also removed as we were not able to retrieve them from the authors after contacting them (i.e., *not available in full- text*). Thus, in category 1, 8 studies were reviewed, and in category 2, 9 studies were reviewed. In Table 2 the selection process of the studies is described.

EHRs and quality, in relation to medical decision making
In a case study published by Vuppalapati et al., [22] it was shown that selfies constitute important outpatient healthcare data which could improve the diagnosis of diseases, as well as the decision-making process. More specifically, it was reported that selfies taken for medical image purposes constitute valuable outpatient healthcare data providing new clinical insights, while they could also be used as diagnostics markers for the provision of prognosis of potential masked diseases. In addition, according to Bar-Dayan et al., [23], whose main aim was to assess the effectiveness of using the EHRs in terms of cost-savings, EHRs were shown to yield significant improvements, both to physicians, as well as to clinic practices and healthcare organizations, as they were shown to provide substantial cost- savings.

Electronic health records can assist in both the prevention, as well as the treatment of a disease. Lardon et al., [24] based on EHR data, developed rules to support diagnosis coding of chronic kidney disease (CKD)

Table 2 Selection process of the studies included in the review

	Number of records
Total records identified:	**1,972**
PubMed	*313*
Scopus	*519*
Google Scholar	*1,140*
Duplicate records among different databases	**20**
Records screened (Title & Abstract) after duplicate removal	**1,952**
Total records excluded:	**1,935**
Due to irrelevance	*1,897*
Due to inability to retrieve them from the authors	*38*
Total records included in the review	**17**
'EHRs and improvement of medical quality and health system'	*8*
'Missing data in the context of EHRs'	*9*

in the hospital of Saint Etienne. In another study of of Garnica et al., [25] electronic health records were shown to help in the prognosis of bacteremia, involving early diagnosis for the provision of treatments to avoid complications and death. Machine Learning (ML) techniques were applied to predict the result of blood culture for the timely administration of the correct treatment thus reducing medical costs. Furthermore, Zaballa et al., [26] presented a general framework to identify and discover the most common treatment pathways which are being exploited to treat diseases. Besides, King et al., [27] confirmed the clinical benefits of EHRs through cross-sectional data examination. EHR adopters reported benefits of EHR use in terms of clinical quality, patient safety, and efficiency, while the use of an EHR meeting *Meaningful Use* criteria was found to be significantly associated with reporting clinical benefits enabled by these functionalities. Except for that, as claimed by Huang et al., [28] EHRs constitute valuable tools which can help in the prediction of multi-type major adverse cardiovascular events. According to Linder et al., [29] it was also shown that EHR–based interventions can improve the smoking status documentation and increase the counseling assistance to smokers. In Table 3 the main findings regarding the contribution of the EHRs on medical quality and the health system, are presented.

Missing data in the context of EHRs
In the context of EHRs, lack of documentation is mainly observed in cases when the patients do not have a symptom or comorbidity. In these cases, instead of recording a negative value for each potential symptom/comorbidity, all data fields are left missing and only the positive values

Table 3 Main findings regarding the contribution of Electronic Health Records on the improvement of medical quality and health system

Study (Author, Year)	Main findings
Vuppalapati et al. [22]	- Selfies taken for medical image purposes are valuable outpatient healthcare data assets that could provide new clinical insights - Diagnostics markers that could provide prognosis of a potential masked disease and necessitate actions to avert any emergency incidence - Improve overall health outcomes of people around the globe in a cost-effective manner that epitomizes the confluence of popularity with curiosity and sharing with accountability
Bar-Dayan et al. [23]	- Positive net financial return from using an electronic medical record system - Referring patients to preferred providers from classes 1–3 was achieved without administrative staff aid. Increased efficiency by redirecting the administrative manpower engaged in this task to other goals - Using EHRs to direct referrals to preferred specialty care physicians, accompanied by a comprehensive physician education program, can play a significant role in facilitating effective utilization of healthcare providers and in lowering costs
Lardon et al. [24]	- A business rule management system could be a good basis to implement a tool to help and check diagnosis codes - Development of rules based on EHR data with the Drools rules engine in order to support diagnosis coding of chronic kidney disease (CKD)
Garnica et al. [25]	- Electronic health records were shown to help in the prognosis of bacteremia, involving early diagnosis for the provision of treatments to avoid complications and death - The three ML supervised classifiers create accurate predictive models of the blood culture outcome using hospital electronic health records
Zaballa et al. [26]	- Identification of the actions in the health system associated with a disease - Identification of those patients with a complete treatment for the disease - Discovery of common treatment pathways followed by the patients with a specific diagnosis
King et al. [27]	- Most physicians with EHRs reported EHR use enhanced patient care overall, helped them access a patient's chart remotely, and alerted them to a potential medication error and critical lab values - EHR use was associated with clinical benefits related to providing recommended care, ordering appropriate tests, and facilitating patient communication
Huang et al. [28]	- Timely and accurate prediction of major adverse cardiovascular events after acute coronary syndrome - Assist clinicians to pay more attention to high-risk patients and improve the quality and efficiency of care accordingly
Linder et al. [29]	- This electronic health record–based intervention improved smoking status documentation and increased counseling assistance to smokers - There was also a suggestion that outcomes of care may have also improved

are recorded. Therefore, lack of a symptom/comorbidity, lack of documentation of a symptom/comorbidity and lack of data collection regarding the symptom/comorbidity cannot be differentiated.

According to the reviewed literature, there is a variety of approaches toward managing missing EHR data; Goldstein et al., [30], who conducted a systematic review regarding the challenges faced during the development of risk prediction models based on EHRs, found that only 58 of the 90 studies (64%) evaluated addressed missing data prior to analysis. Some of the simplest methodological approaches being used, involve the selection of sub-datasets that contain complete information [31, 32], as well as the stratified mean imputation [33], while others have advanced statistical methodologies which are applicable only to continuous measures and interpolate longitudinal variables with limited individual-level variability that are typically not dependent on other covariates [34]. Despite these approaches, few studies utilized "*informative observations*" where the presence of a variable is meaningful for the possibly missing values [30]. Xu et al., [35] developed a deep learning unsupervised method to impute

missing values in patient records and by comparing it with four other imputation techniques, they showed that the specific methodology could significantly reduce the imputation biases under various scenarios, and as a result it could empower physicians and researchers to better utilize the EHRs aiming at improved patient management.

In addition, Hwang, et al. [36] proposed a two-stage framework leading to more robust results for disease prediction based on EHRs with missing data. Two different imputation methods were implemented, the first of which replaced the missing values with the mean values of the attributes, while the second one used an autoencoder, which is an unsupervised ML algorithm. Furthermore, Wang et al. [37], based on the idea that among heterogeneous patient populations there exist homogeneous groups of patients, proposed a data driven approach for imputing the sparse patient EHRs by transferring relevant knowledge from patients with denser EHRs to their patients with sparse EHRs. In Fig. 1 an overview of the methodologies used for imputing missing data in the context of the EHRs, based on the research works included in the present review, is illustrated.
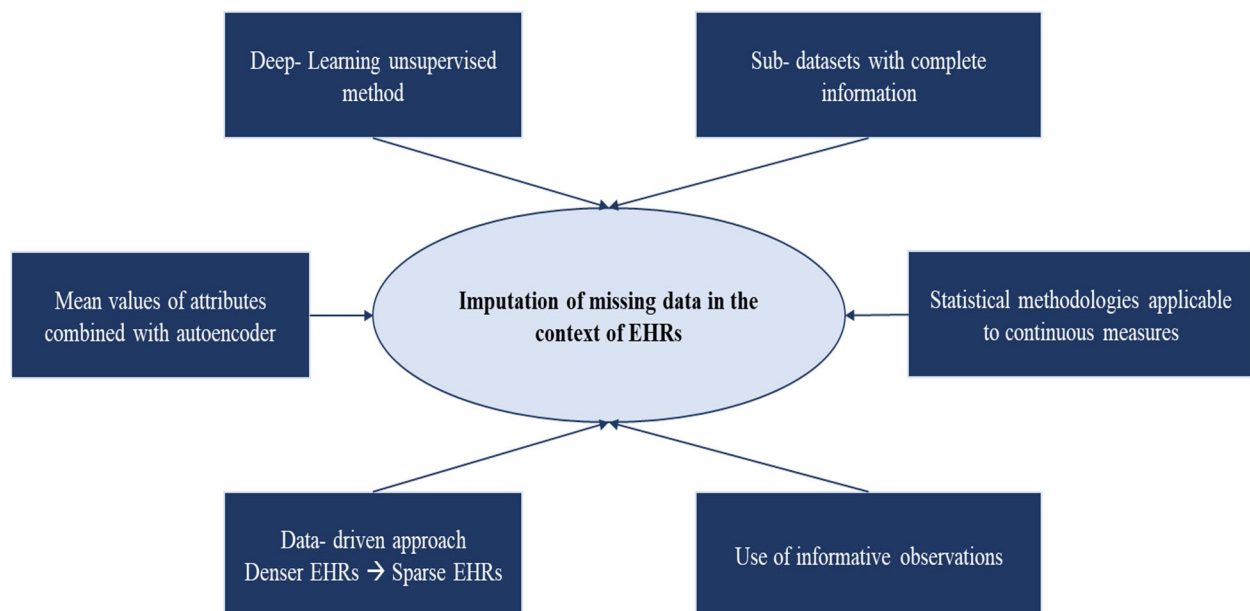
Fig. 1 Missing data imputation techniques in the context of EHRs, based on the research works included in the present review

Discussion

Based on the present review, EHRs constitute an increasingly important tool for both healthcare professionals and decision makers, which can improve national healthcare systems both for the convenience of patients and doctors, by helping on the prevention and treatment of chronic and non- chronic diseases, while regarding the statistical methodologies being implemented for imputing missing data, further steps should be conducted and new methodologies should be proposed and be tested in this context.

Benefits of EHRs

As already pointed out, some of the most important benefits related to EHRs include the easy access to computerized records, as well as the elimination of poor penmanship, which constitutes a widespread and significant obstacle in the medical world [38, 39]. Besides, EHRs provide significant cost savings, as based on the studies of Shu et al. [40] and Bar- Dayan et al. [23], it was shown that the release of EHR data to patients via smart apps can save both the hospital, as well as the patients, approximately 2 million and 1 million euros, respectively, on an annual basis. This could be attributed to the fact that, the EHR's use can substantially reduce the redundant implementation of medical tests or the need to mail hard copies of test results to different providers [41, 42]. Additionally, several studies have also shown that EHRs, compared to hard- copies, result in reduced transcription costs through point-of-care documentation and other

structured documentation procedures [43]. Furthermore, the access to electronically stored data increases the availability of data, which leads to the improvement of the ability to conduct research, as well as to the facilitation of the identification of evidence- based best health practices [44], while at the same time public health researchers by using EHRs tend to produce more beneficial for the society research outcomes. Even more, according to several studies, despite the fact that EHRs have known drawbacks when they are used solely as data sources for studies informing public health decisions [45], they contain several crucial data elements which help with a pandemic response [46, 47].

Missing data handling techniques

As far as the missing data handling techniques is concerned, several investigators have already tried to propose the best possible methodology, yet there is no wide consensus and acceptance in the scientific community, while there are also crucial gaps which should be addressed. As pointed out, missing information constitutes a widely spread phenomenon in routinely collected health data and often missingness is very informative and should be incorporated into the development process of prediction and epidemiological models [48, 49], as the absence of data in EHR records can substantially decrease our ability to create accurate predictions [49]. Besides, the majority of the hitherto developed prediction models are not able to provide a risk estimate when missing information exist in predictor variables, which delays their implementation

and may ultimately limit guideline adherence [50]. However, the correct way of handling missing values particularly in the phase of prediction model development and in the validation dataset, solely depends on the intended use of the prediction model, and more specifically, on whether the investigator intends to allow for missing data during model application in practice [51]. So far, in clinical practise and in a real clinical setting, when applying already developed prediction models in new patients arising in the medical office to predict their risk of disease onset or disease recurrence, accounting for missing values in some of their demographic or clinical characteristics is not straightforward. Ideally, when developing a prediction model the methodology regarding the handling of missing data should be integrated, however this is not a usual case in practise, as most of the developed models do not allow for missing data [51–63].

Limitations of the literature review process

However, this review paper has some limitations, such as the fact that there is not a well-established metric to evaluate the performance of the EHRs in clinical practice. Therefore, no quantitative assessment could be performed that also evaluate the cost-effectiveness of EHR in medical decision making. Moreover, no pooled analysis or quality assessment of the reviewed studies was performed, as this was out of the scope of the present work, and in many cases was not feasible.

Conclusions

Despite the limitations of the present review, the importance of the EHRs' implementation in clinical practice was highlighted, while at the same time the gap of knowledge regarding the missing data handling techniques was also pointed out. EHRs seems that they constitute an increasingly important tool for both physicians, decision makers and patients, which can improve national healthcare systems both for the convenience of patients and doctors, while they improve the quality of health care as well as they can also be used to save money.

Abbreviations
CKD   Chronic Kidney Disease
EHRs   Electronic Health Records
ML   Machine–Learning
PRISMA   Preferred Reporting Items for Systematic Reviews and
            Meta-Analyses

Authors' contributions
DBP and TT have conducted the review of the articles independently and after a thorough discussion of potential disagreements, the final selection of the studies was made.

Declarations

References
1. Katehakis DG. Electronic medical record implementation challenges for the national health system in Greece. Int J Reliable Quality E-Healthcare (IJRQEH). 2018;7(1):16–30.
2. Institute of Medicine. To Err Is Human: Building a Safer Health System. Washington, DC: National Academy Press; 2000. https://www.nap.edu/read/9728/chapter/1. Accessed 19 Feb 2017.
3. The Office of the National Coordinator for Health Information Technology. EHR Vendors Reported by Providers Participating in Federal Programs. https://dashboard.healthit.gov/datadashboard/documentation/ehr-vendors-reported-CMS-ONC-data-documentation.php. Accessed 19 Feb 2017.
4. Watson R. EU sets out plans to digitise health records across member states. 2022.
5. Institute of Medicine. Key Capabilities of Electronic Health Record. Washington, DC: National Academy Press; 2003.
6. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. N Engl J Med. 2010;363(6):501–4. https://doi.org/10.1056/NEJMp1006114.
7. Chaudhry B, Wang J, Wu S, et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. Ann Intern Med. 2006;14410:742–52.
8. Kaushal R, Shojania KG, Bates DW. Effects of computerized physician order entry and clinical decision support systems on medication safety: a systematic review. Arch Intern Med. 2003;16312:1409–16.
9. Hossain ME, Khan A, Moni MA, Uddin S. Use of electronic health data for disease prediction: A comprehensive literature review. IEEE/ACM Trans Comput Biol Bioinf. 2019;18(2):745–58.
10. Casey JA, Pollak J, Glymour MM, Mayeda ER, Hirsch AG, Schwartz BS. Measures of SES for electronic health record-based research. Am J Prev Med. 2018;54(3):430–9.
11. Gianfrancesco MA, Goldstein ND. A narrative review on the validity of electronic health record-based research in epidemiology. BMC Med Res Methodol. 2021;21(1):1–10.
12. Goldstein BA, Bhavsar NA, Phelan M, Pencina MJ. Controlling for informed presence bias due to the number of health encounters in an electronic health record. Am J Epidemiol. 2016;184(11):847–55. ISO 690.
13. Nelson A. Unequal treatment: confronting racial and ethnic disparities in health care. J Natl Med Assoc. 2002;94(8):666.
14. Polubriaginof F C, Ryan P, Salmasian H, Shapiro AW, Perotte A, Safford MM, ... Vawdrey DK. Challenges with quality of race and ethnicity data in observational databases. J Am Med Informatics Assoc. 2019;26(8–9):730–736.
15. Larkins NG, Craig JC, Teixeira-Pinto A. A guide to missing data for the pediatric nephrologist. Pediatr Nephrol. 2019;34(2):223–31.

16. Liu F, Panagiotakos D. Real-world data: a brief review of the methods, applications, challenges and opportunities. BMC Med Res Methodol. 2022;22(1):287. https://doi.org/10.1186/s12874-022-01768-6.

17. Bell ML, Kenward MG, Fairclough DL, Horton NJ. Differential dropout and bias in randomised controlled trials: when it matters and when it may not. BMJ. 2013;346:e8668. ISO 690.

18. Little RJ, Rubin DB. The analysis of social science data with missing values. Sociol Methods Res. 1989;18(2–3):292–326.

19. Tsiampalis T, Panagiotakos DB. Missing-data analysis: socio-demographic, clinical and lifestyle determinants of low response rate on self-reported psychological and nutrition related multi-item instruments in the context of the ATTICA epidemiological study. BMC Med Res Methodol. 2020;20:1–13.

20. Tsiampalis T, Vassou C, Psaltopoulou T, Panagiotakos DB. Socio-Demographic, clinical and lifestyle determinants of low response rate on a self-reported psychological multi-item instrument assessing the adults' hostility and its direction: ATTICA Epidemiological Study (2002–2012). Int J Stat Med Res. 2021;10:1–9.

21. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Moher D. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Int J Surg. 2021;88:105906.

22. Vuppalapati J, Kedari S, Vuppalapati R, Vuppalapati C, Ilapakurti A. The Role of Selfies in Creating the Next Generation Computer Vision Infused Outpatient Data Driven Electronic Health Records (EHR). In: Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018. 2019. p. 2458–2466.

23. Bar-Dayan Y, Saed H, Boaz M, Misch Y, Shahar T, Husiascky I, Blumenfeld O. Using electronic health records to save money. J Am Med Inform Assoc. 2013;20:e17-20.

24. Lardon J, Asfari H, Souvignet J, Trombert-Paviot B, Bousquet C. Improvement of diagnosis coding by analysing EHR and using rule engine: application to the chronic kidney disease. Stud Health Technol Inform. 2015;210:120–4.

25. Garnica O, Gómez D, Ramos V, Hidalgo JI, Ruiz-Giardín JM. Diagnosing hospital bacteraemia in the framework of predictive, preventive and personalised medicine using electronic health records and machine learning classifiers. EPMA J. 2021;2:365–81.

26. Zaballa O, Pérez A, Gómez Inhiesto E, Acaiturri Ayesta T, Lozano JA. Identifying common treatments from electronic health records with missing information. An application to breast cancer. PloS one. 2020;15(12):e0244004.

27. King J, Patel V, Jamoom EW, Furukawa MF. Clinical Benefits of Electronic Health Record Use: National Findings. Health Serv Res. 2014;49:392–404.

28. Huang Z, Lu Y, Dong W. Utilizing electronic health records to predict multi-type major adverse cardiovascular events after acute coronary syndrome. Knowl Inf Syst. 2019;60(3):1725–52.

29. Linder JA, Rigotti NA, Schneider LI, Kelley JH, Brawarsky P, Haas JS. An electronic health record–based intervention to improve tobacco treatment in primary care: a cluster-randomized controlled trial. Arch Intern Med. 2009;169(8):781–7.

30. Goldstein BA, Navar AM, Pencina MJ, Ioannidis J. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. J Am Med Inform Assoc. 2017;24(1):198–208.

31. Bloomfield GS, Hogan JW, Keter A, Holland TL, Sang E, Kimaiyo S, Velazquez EJ. Blood pressure level impacts risk of death among HIV seropositive adults in Kenya: a retrospective analysis of electronic health records. BMC Infect Dis. 2014;14(1):1–10.

32. Martín-Merino E, Calderón-Larrañaga A, Hawley S, Poblador-Plou B, Llorente-García A, Petersen I, Prieto-Alhambra D. The impact of different strategies to handle missing data on both precision and bias in a drug safety study: a multidatabase multinational population-based cohort study. Clin Epidemiol. 2018;10:643.

33. Dalton A, Bottle A, Soljak M, Okoro C, Majeed A, Millett C. The comparison of cardiovascular risk scores using two methods of substituting missing risk factor data in patient medical records. J Innov Health Inform. 2011;19(4):225–32.

34. Ebrahim GJ. Missing data in clinical studies molenberghs g. and Kenward M. G. J Trop Pediatr. 2007:53(4):294. https://doi.org/10.1093/tropej/fmm053.

35. Xu D, Hu PJ, Huang TS, Fang X, Hsu CC. A deep learning-based, unsupervised method to impute missing values in electronic health records for improved patient management. J Biomed Inform. 2020;111: 103576.

36. Hwang U, Choi S, Lee HB, Yoon S. Adversarial training for disease prediction from electronic health records with missing data. arXiv preprint arXiv:1711.04126. 2017.

37. Wang F, Zhou J, Hu J. DensityTransfer: a data driven approach for imputing electronic health records. In 2014 22nd International Conference on Pattern Recognition. IEEE. 2014. p.2763–68.

38. Rodriguez-Vera FJ, Marin Y, Sanchez A, et al. Illegible handwriting in medical records. J R Soc Med. 2002;95(11):545–6.

39. Winslow EH, Nestor VA, Davidoff SK, et al. Legibility and completeness of physicians' handwritten medication orders. Heart Lung. 1997;26(2):158–64.

40. Shu T, Xu F, Li H, Zhao W. Investigation of patients' access to EHR data via smart apps in Chinese Hospitals. BMC Med Inform Decis Mak. 2021;21:53.

41. Chen P, Tanasijevic MJ, Schoenenberger RA, et al. A computer-based intervention improving the appropriateness of antiepileptic drug level monitoring. Am J Clin Pathol. 2003;119(3):432–8.

42. Tierney WM, Miller ME, Overhage JM, McDonald CJ. Physician inpatient order writing on microcomputer workstations Effects on resource utilization. JAMA. 1993;269(3):379–83.

43. Agrawal A. Return on investment analysis for a computer-based patient record in the outpatient clinic setting. J Assoc Acad Minor Phys. 2002;13(3):61–5.

44. Aspden P. Patient Safety Achieving a New Standard for Care. Washington, D.C: National Academies Press; 2004.

45. Cifuentes M, Davis M, Fernald D, Gunn R, Dickinson P, Cohen DJ. Electronic health record challenges, workarounds, and solutions observed in practices integrating behavioral health and primary care. J Am Board Family Med. 2015;28(Suppl 1):S63–72.

46. Atreja A, Gordon SM, Pollock DA, Olmsted RN, Brennan PJ, Healthcare Infection Control Practices Advisory Committee. Opportunities and challenges in utilizing electronic health records for infection surveillance, prevention, and control. Am J Infect Control. 2008;36(3):S37-46.

47. Kukafka R, Ancker JS, Chan C, et al. Redesigning electronic health record systems to support public health. J Biomed Inform. 2007;40(4):398–409.

48. Madden JM, Lakoma MD, Rusinak D, Lu CY, Soumerai SB. Missing clinical and behavioral health data in a large electronic health record (EHR) system. J Am Med Inform Assoc. 2016;23(6):1143–9.

49. Wells BJ, Chagin KM, Nowacki AS, Kattan MW. Strategies for handling missing data in electronic health record derived data. EGEMS. 2013;1(3):1035.

50. Kotseva K, Wood D, De Bacquer D, De Backer G, Rydén L, Jennings C, ... EUROASPIRE Investigators. EUROASPIRE IV: A European Society of Cardiology survey on the lifestyle, risk factor and therapeutic management of coronary patients from 24 European countries. Eur J Prev Cardiol. 2016;23(6):636–648.

51. Hoogland J, van Barreveld M, Debray TP, Reitsma JB, Verstraelen TE, Dijkgraaf MG, Zwinderman AH. Handling missing predictor values when validating and applying a prediction model to new patients. Stat Med. 2020;39(25):3591–607. https://doi.org/10.1002/sim.8682.

52. Austin PC, White IR, Lee DS, van Buuren S. Missing data in clinical research: a tutorial on multiple imputation. Can J Cardiol. 2021;37(9):1322–31.

53. Beaulieu-Jones BK, Lavage DR, Snyder JW, Moore JH, Pendergrass SA, Bauer CR. Characterizing and managing missing structured data in electronic health records: data analysis. JMIR Med Inform. 2018;6(1):e11.

54. Buntin MB, Jain SH, Blumenthal D. Health information technology: laying the infrastructure for national health reform. Health Aff (Millwood). 2010;296:1214–9.

55. Gopalakrishna G, Mustafa RA, Davenport C, Scholten RJ, Hyde C, Brozek J, Schünemann HJ, Bossuyt PM, Leeflang MM, Langendam MW. Applying Grading of Recommendations Assessment, Development and Evaluation (GRADE) to diagnostic tests was challenging but doable. J Clin Epidemiol. 2014;67(7):760–8.

56. Hu Z, Melton GB, Arsoniadis EG, Wang Y, Kwaan MR, Simon GJ. Strategies for handling missing clinical data for automated surgical site infection detection from the electronic health record. J Biomed Inform. 2017;68:112–20.

57. Institute of Medicine. Key Capabilities of Electronic Health Record. Washington, DC: National Academy Press; 2003.
58. Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academy Press; 2001.
59. Nijman SW, Groenhof TK, Hoogland J, Bots ML, Brandjes M, Jacobs JJ, ... Debray TP. Real-time imputation of missing predictor values improved the application of prediction models in daily practice. J Clin Epidemiol. 2021;134:22-34.
60. Li J, Yan XS, Chaudhary D, Avula V, Mudiganti S, Husby H, Shahjouei S, Afshar A, Stewart WF, Yeasin M, Zand R, Abedi V. Imputation of missing values for electronic health record laboratory data. NPJ digital medicine. 2021;4(1):147.
61. Liu L, Li H, Hu Z, Shi H, Wang Z, Tang J, Zhang M. Learning hierarchical representations of electronic health records for clinical outcome prediction. In AMIA Annual Symposium Proceedings. Am Med Inform Assoc. 2019;2019:597.
62. Pedersen AB, Mikkelsen EM, Cronin-Fenton D, et al. Missing data and multiple imputation in clinical epidemiological research. Clin Epidemiol. 2017;9:157–66.
63. Zhang X, Xiao J, Gong Y, Yu N, Zhang W, Jang S, Gu F. Handling the missing data problem in electronic health records for cancer prediction. In 2020 Spring Simulation Conference (SpringSim). IEEE. 2020. p. 1–9.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.