



# Utilization of electronic healthcare records for advancement of medical knowledge

Zhongheng Zhang

Department of emergency medicine, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou 310016, China  
Correspondence to: Zhongheng Zhang. No. 3 East Qingchun Road, Hangzhou 310016, China. Email: zh\_zhang1984@zju.edu.cn.

Received: 09 August 2018; Accepted: 10 August 2018; Published: 06 September 2018.

doi: 10.21037/jmai.2018.08.01

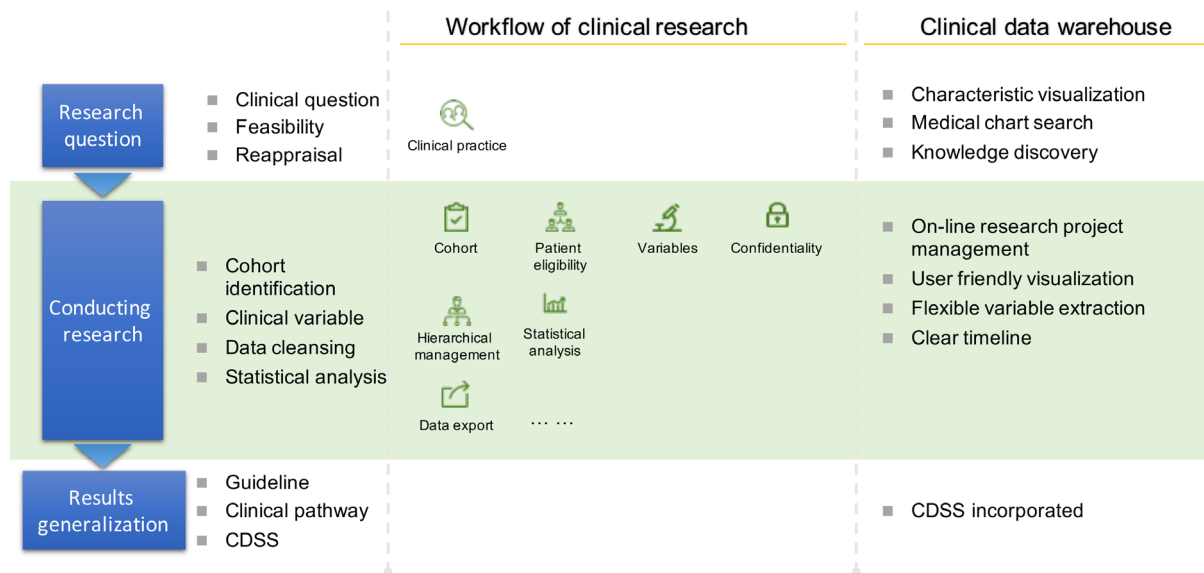
View this article at: <http://dx.doi.org/10.21037/jmai.2018.08.01>

Electronic healthcare records (EHR) has been widely used in clinical practice for decades in China (1), and now nearly all healthcare facilities have used EHR for daily clinical practice (2). The benefits of using EHR are unquestionable that it significantly enhances the flow of patient management. For instance, when a patient is suspected of stroke and presents to the emergency department, the stroke unit will be triggered via EHR system which will in turn automatically remind multidisciplinary members of the stroke unit. EHR also provide information for the past history of a patient. In older days, a patient had to bring a large amount of medical records to the clinic and the doctors had to spent a lot of time looking for the most indicative medical information. This could be time-consuming and often the patient may not provide the medical history. With the use of EHR, all past medical histories of a patient can be easily retrieved at one click of the mouse and the timeline is very clear.

While the primary goal of EHR is facilitate the management of daily clinical practice, many researchers found that the EHR can provide high-quality and real-world data for doing clinical researches. There are several potential advantages of using EHR for doing clinical researches: (I) it is convenient and cheap. As compared with randomized controlled trial, the use of EHR can be much cheaper and time-saving; (II) EHR provides real-world evidence for the effectiveness of an intervention or diagnostic tests (3). It has long been debated that evidence based on well-designed RCT may not applicable to real-world setting (4). The different scenarios in RCT and real-world practice cannot be overlooked. For example, lactate may be routinely measured at fixed time points after enrollment in a well-designed RCT, and this information may help to identify some occult sepsis that can be easily overlooked in daily busy practice; (III) EHR is able to provide data with high

granularity (5). For example, the vital signs are measured in hourly basis in the operating room and intensive care unit (ICU) (6). Such high-granularity data can help to build some time-varying model with high accuracy of outcome prediction (7,8); (IV) data stored in EHR are ever-growing without much additional cost. Some modeling algorithms requiring longitudinal data can be performed with such data (9); (V) some well calibrated models can be incorporated into the EHR to assist clinical decision making. The so-called clinical decision support system (CDSS) is such kind of device. Sepsis is a heterogeneous syndrome and its identification requires information from multiple sources such as vital signs, laboratory findings and physical examination. In busy hours in the emergency department, such sepsis may be overlooked and initiation of antibiotics and fluid resuscitation can be delayed. With the use of CDSS, the machine can automatically extract information for the judgement whether a patient has sepsis and gives a probability. Clinicians can focus more on those with high probability of sepsis. There is evidence that such a CDSS can help to improve clinical outcomes.

However, the disadvantage of EHR should also be highlighted and cautions be practices in interpreting the results from EHR. (I) The healthcare processes vary substantially across different medical centers and such a variability may be responsible for discrepancy of results and conclusions for studies conducted in different places (10,11). For example, an institution tested transferrin and pre-albumin routinely for the assessment of nutritional status can show different effectiveness on the effectiveness of enteral nutritional protocol on mortality outcome, as compared with hospitals without such routine practice (12,13); (II) clinical studies based on EHR are observational studies in essence and have all inherit limitations of such



**Figure 1** Workflow of clinical research utilizing clinical data warehouse. CDSS, clinical decision support system.

kind of design (14). The causal link between a factor of interest and the outcome cannot be established with EHR. In this regard, the results obtained from EHR are hypothesis-generating at best; (III) missing value is an important issue in utilizing EHR. Unlike well-designed and scrutinized RCT, the ordering of laboratory tests is largely dependent on the judgement of the treating physician, making the processes heterogeneous among patients. fortunately, if missing data are handled properly, the estimation bias can be minimized (15,16); (IV) errors occurs frequently in EHR, making the process of data cleansing particularly important.

Despite these potential limitations and difficulties in utilizing EHR for clinical researches, more in-depth quality control and establishment of standard EHR structure can help to minimized some of these limitations. For example, the Healthcare Information and Management Systems Society (HIMSS) advocates a number of standards that impact the electronic health record. These standards require that all certified EHR products maintain data confidentiality, ensure interoperability to share information, and can execute a series of well-defined functions. Once an EHR is certified, clinicians and organizations can feel assured that their system is both effective and efficient in supporting information technology (17).

Nowadays, many hospitals have attempted to establish clinical data warehouse to facilitate clinical studies. The framework of the using clinical data warehouse for clinical

studies is shown in *Figure 1*. The research question comes from clinical practice and should focus on a point that is relevant to clinicians and patients. Typically, the importance of a research question would dependent on the extent how patient and/or their family members concern. For instance, for a patient who transfers to the ICU, he and/or his family members will ask: can I survive this critical illness? That is why most critical care trials use the mortality as the primary study endpoint. After a clinical question is well defined, the second step is to define the study cohort. The cohort is defined by inclusion/exclusion criteria and can be identified from the data warehouse by logical combinations of several criteria. The third step is to extract variables that can help to control confounding. As the confounding is the Achilles' heel of observational studies, the extraction of potential confounders is made not only by clinical judgement but also depends on the relationships between variables. The forth step is data cleansing, since the raw data may contain some errors. The fifth step is to perform statistical analysis. In this step, some sophisticated modeling strategy can be applied such as restricted cubic spline, fractional polynomials and machine learning (18-21). The final step is to apply the obtained knowledge from data mining to clinical practice. A well-known example is early warning system for clinical deterioration (22,23). Some prediction models are trained and validated with the real-world data and then the models are incorporated into the EHR for practical use.

## Acknowledgements

I thank Yiduccloud (Beijing) technology Ltd. for supporting part of idea conception and figure preparation.

*Funding:* Z.Z. received funding from the public welfare research project of Zhejiang province (LGF18H150005) and Scientific research project of Zhejiang Education Commission (Y201737841).

## Footnote

*Conflicts of Interest:* The author has no conflicts of interest to declare.

## References

- Liang J, Wei K, Meng Q, et al. Development of medical informatics in China over the past 30 years from a conference perspective and a Sino-American comparison. *PeerJ* 2017;5:e4082.
- Zhang Z. Big data and clinical research: focusing on the area of critical care medicine in mainland China. *Quant Imaging Med Surg* 2014;4:426-9.
- Oude Rengerink K, Kalkman S, Collier S, et al. Series: Pragmatic trials and real world evidence: Paper 3. Patient selection challenges and consequences. *J Clin Epidemiol* 2017;89:173-80.
- Nallamothu BK, Hayward RA, Bates ER. Beyond the randomized clinical trial: the role of effectiveness studies in evaluating cardiovascular therapies. *Circulation* 2008;118:1294-303.
- Marco-Ruiz L, Moner D, Maldonado JA, et al. Archetype-based data warehouse environment to enable the reuse of electronic health record data. *Int J Med Inform* 2015;84:702-14.
- Johnson AE, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. *Sci Data* 2016;3:160035.
- Zhang Z. Accessing critical care big data: a step by step approach. *J Thorac Dis* 2015;7:238-42.
- Zhang Z, Reinikainen J, Adeleke KA, et al. Time-varying covariates and coefficients in Cox regression models. *Ann Transl Med* 2018;6:121.
- Nemati S, Holder A, Razmi F, et al. An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU. *Crit Care Med* 2018;46:547-53.
- Agniel D, Kohane IS, Weber GM. Biases in electronic health record data due to processes within the healthcare system: retrospective observational study. *BMJ* 2018;361:k1479.
- Goss FR, Lai KH, Topaz M, et al. A value set for documenting adverse reactions in electronic health records. *J Am Med Inform Assoc* 2018;25:661-9.
- Zhang Z, Li Q, Jiang L, et al. Effectiveness of enteral feeding protocol on clinical outcomes in critically ill patients: a study protocol for before-and-after design. *Ann Transl Med* 2016;4:308.
- Li Q, Zhang Z, Xie B, et al. Effectiveness of enteral feeding protocol on clinical outcomes in critically ill patients: A before and after study. *PLoS ONE* 2017;12:e0182393.
- Uddin MJ, Groenwold RH, de Boer T, et al. Instrumental Variable Analysis in Epidemiologic Studies: An Overview of the Estimation Methods. *Pharm Anal Acta* 2015;6:353.
- Zhang Z. Missing data imputation: focusing on single imputation. *Ann Transl Med* 2016;4:9.
- Zhang Z. Multiple imputation with multivariate imputation by chained equation (MICE) package. *Ann Transl Med* 2016;4:30.
- Diana ML, Kazley AS, Menachemi N. An assessment of Health Care Information and Management Systems Society and Leapfrog data on computerized provider order entry. *Health Serv Res* 2011;46:1575-91.
- Zhang Z. Multivariable fractional polynomial method for regression model. *Ann Transl Med* 2016;4:174.
- Zhang Z, Chen K, Ni H, et al. Predictive value of lactate in unselected critically ill patients: an analysis using fractional polynomials. *J Thorac Dis* 2014;6:995-1003.
- Herndon JE 2nd, Harrell FE Jr. The restricted cubic spline as baseline hazard in the proportional hazards model with step function time-dependent covariables. *Stat Med* 1995;14:2119-29.
- Chen JH, Asch SM. Machine Learning and Prediction in Medicine - Beyond the Peak of Inflated Expectations. *N Engl J Med* 2017;376:2507-9.
- Hu SB, Wong DJ, Correa A, et al. Prediction of Clinical Deterioration in Hospitalized Adult Patients with Hematologic Malignancies Using a Neural Network Model. *PLoS ONE* 2016;11:e0161401.
- Benthin C, Pannu S, Khan A, et al. The Nature and Variability of Automated Practice Alerts Derived from Electronic Health Records in a U.S. Nationwide Critical Care Research Network. *Ann Am Thorac Soc* 2016;13:1784-8.

doi: 10.21037/jmai.2018.08.01

**Cite this article as:** Zhang Z. Utilization of electronic healthcare records for advancement of medical knowledge. *J Med Artif Intell* 2018;1:9.