BRIEF COMMUNICATIONS
DATA HYGIENE: IMPORTANT STEP IN DECISION-MAKING WITH IMPLICATIONS FOR HEALTH IT PROFESSIONALS

GV Fant

US Department of Veterans Affairs, Washington DC, USA

ABSTRACT

Medical and health data that have been entered into an electronic data system in real-time cannot be assumed to be accurate and of high quality without verification. The adoption of the electronic health record (EHR) by many countries to the support care and treatment of patients illustrates the importance of high quality data that can be shared for efficient patient care and the operation of healthcare systems. This brief communication provides a high-level overview of an EHR system and practices related to high data quality and data hygiene that could contribute to the analysis and interpretation of EHR data for use in patient care and healthcare system administration.

KEYWORDS

Data Hygiene; EHR; Data Management Cycle; Healthcare Systems; Health IT Professionals

1. INTRODUCTION

The adoption of the electronic health record (EHR) system by many countries to the support care and treatment of patients illustrates the importance of high quality data that can be shared for efficient patient care and operation of healthcare systems. The implementation of an EHR system includes many considerations such as workflow, system testing, measuring response times for key medical system transactions, accurate medical and health data, etc. [1]. EHR data need to be accurate and of high quality because patient care and health system decisions are based on these data. The data may also be used to conduct analyses to support patient care, quality, and safety activities at the bed-side and within the healthcare system [2].

Medical and health data that have been entered into an electronic data system during the course of normal operations cannot be assumed to be accurate and of high quality without verification and data pre-processing, as needed. Health IT professionals working with clinicians, health administrators, and health informatics professionals will be involved in using the EHR data. It is also likely that Health IT and health information professionals will work together to ensure successful EHR implementation [3] that lead to accurate and high quality EHR data.

This brief communication provides a short overview of the EHR system and its basic operations. Additionally, special attention is given to issues and practices related to high data quality and data hygiene that Health IT and health information professionals could contribute to the subsequent analysis and interpretation of EHR data for use in patient care and healthcare system administration.

2. DATA HYGIENE: IMPORTANT STEP IN HEALTH SYSTEM DECISION-MAKING

Functionally, an EHR system is a type of relational database management system. The formal definition of the EHR is as follows [4]:

Electronic Health Record: An electronic record of health-related information on an individual that conforms to nationally recognized interoperability standards and that can be created, managed and consulted by authorized clinicians and staff across more than one healthcare organization.

An EHR system, in general, supports the activities or operations in five functional domains [5]:

- Non-clinical operations
- Clinical operations
- Revenue cycle/finance
- Regulatory compliance
- Reporting

Healthcare organizations use data recorded in an EHR system for financial/administrative operations, clinical reporting on patients and patient cohort groups, and healthcare quality improvement. There are different users of the EHR data who perform key activities in the previously identified domains.

An EHR system supports the population health management activities for patient cohort groups. The EHR reporting functions that deal with population health management and healthcare quality improvements are important in managing the care of a patient cohort along with the medical care services within a healthcare system. When implementing any EHR system, the specific activities that may need to be explored for population health management include the following [5]:

- Generating statistical reports for health quality improvement measures
- Compiling data from the EHR system for health system administrators and external reporting to other federal agencies
- Verifying the accuracy of generated reports prior to distribution
- Identifying reportable health quality improvement measurements
- Exploring methods to generate these reports (running queries, executing standardized reports, and creating custom reports)
- Mining data and using data science and data extraction methods
- Avoiding common reporting errors

The identification of reporting errors and the verification of the accuracy of data reports generated by an EHR system point to the importance of considering data hygiene, or data cleaning, as parts of the operations of any relational database management system. After data entry and prior to analysis, interpretation, and use of the data, essential activities, including data hygiene, must occur (see Fig 1).

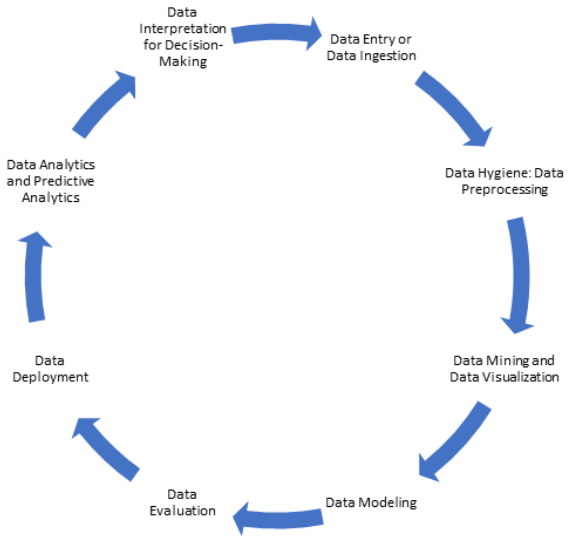

Figure 1: Data management cycle—Steps toward using data for action.
Adapted from: Healthcare Informatics training manual. Salt Lake City: AAPC, 2022

Figure 1 is a typical data management cycle. It complements a similar framework for public health data management [6]. The figure shows the many steps needed to create data that can be interpreted by decision-makers for action. The important first-step is data hygiene.

Data hygiene is the process of making sure the dataset or database contain clean data. Clean data have three basic characteristics: correctly formatted and free from errors, redundancies, and inconsistencies. Healthcare organizations can take three, broad steps to minimize data errors [7]:

- Standardized methods for collecting data from the patient
- Enter the data into EHR in a standardized, consistent format
- Keep the patient's demographic data up-to-date

Data that are not clean cannot be relied upon for patient care or in healthcare system decision-making.

3. IMPLICATIONS FOR HEALTH IT AND HEALTH INFORMATION PROFESSIONALS

Data professionals have discussed the importance of data hygiene practices [8-10]. Data hygiene methods focus on correcting the errors in the data in order to meet the intended use of that data for organizational purposes. A possible health solution for a healthcare organization is the frequent and on-going examination or audit of data found in the EHR. A best practice of the data hygiene process is to conduct a data audit within an EHR system, and this audit could include examination of the following [10]:

- Accessibility – the data is available when needed;
- Accuracy – affinity with original intent, veracity as compared to an authoritative source, correlation of data elements (e.g., an insurance card and a driver's license), and measurement precision (e.g., a patient's last name is correctly spelled);
- Completeness – availability of required data attributes (e.g., a ZIP code is missing);
- Coverage – availability of required data records (e.g., some returning patient's records are missing);

- Conformity – alignment of content with required standards (e.g. birth date formatted as MMDDYYYY);
- Consistency – compliance with required patterns and uniformity rules (e.g., "Street" must be abbreviated to "ST"), supported by data entry standards, workflow management, and technical design standards;
- Integrity – accuracy of data relationships (parent and child linkage, e.g., patient is correctly represented as the mother of another patient);
- Timeliness – the currency of content (e.g., the patient's name change is recorded as soon as it is known, and automatically updated across all relevant data stores); and
- Uniqueness – each record can be unambiguously identified (e.g., patient lookup and other reports provide a unique ID, versus Last Name, First Name, and Date of Birth) – uniqueness also includes checks for redundancy of records (e.g., duplicate patient records).

Health IT and health information professionals are in a unique position to raise issues of data hygiene or data quality in the process of EHR system implementation within a healthcare organization. The connection of high quality data to various types of patient care and health system decisions could be highlighted for the benefit of data users who may be unfamiliar with the importance of data hygiene issues. These same health information professionals may perform the data audit functions and also work with clinical professionals to correct medical data entry errors in an EHR system [11].

4. CONCLUSION

Data hygiene should be considered in any data collection effort intended to inform decision-making, including EHR system implementation to support health solutions, including clinical and health system decision-making. These decisions depend on clean data that result from attention to data hygiene. Health IT and health information professionals are in a unique position to highlight important issues of data hygiene in the process of EHR system implementation.

ACKNOWLEDGEMENTS

I would like to thank Dr V Gordon, RN, DNP, and Dr L Backus, MD, for their encouragement.

REFERENCES

[1] Kilkenny M, Robinson K. Data quality: "garbage in-garbage out." Health Information Management Journal. 2018. 47(3): 103-105.
[2] Clack L. Using data analytics to predict outcomes in healthcare. J AHIMA. 2023; URL: https://journal.ahima.org/page/using-data-analytics-to-predict-outcomes-in-healthcare.
[3] Aguirre R, et al. Electronic health record implementation: a review of resources and tools. Cureus. 2019. 11(9): e5649.
[4] Hoyt R, Hersh W. Health Informatics—practical guide, seventh edition. Informatics Education, 2018.
[5] Roan L. National Healthcare Association: Electronic Health Record Specialist (CEHRS), Study Guide, 2.0. Assessment Technologies Institute, 2020.
[6] Fant GV. Technical Review: Key concepts of health database management for public health workforce development in resource-limited settings. Intl J Science and Research Archive. 2023; 09(02): 222–230.
[7] AAPC. Healthcare Informatics training manual. Salt Lake City: AAPC, 2022.
[8] Indeed Editorial Team. What is data hygiene? (And why it's important). Updated: June 24, 2022. URL: https://www.indeed.com/career-advice/career-development/data-hygiene

[9] Herold R. Best practices for data hygiene. ISACA Now Blog 2021. Date Published: March 19, 2021. URL: https://www.isaca.org/resources/news-and-trends/isaca-now-blog/2021/best-practices-for-data-hygiene
[10] HealthIT. Data Quality: Data Cleansing and Improvement. n.d. URL: https://www.healthit.gov/playbook/pddq-framework/data-quality/data-quality-assessment/
[11] Adane K, Gizachew M, Kendie S. The role of medical data in efficient patient care delivery: a review. Risk Management and Healthcare Policy. 2019; 12: 67-73.

AUTHOR

GV Fant, PhD, MSHS, CIMP-Data Science, CEHRS: Epidemiologist, Population Health Services, Veterans Health Administration, US Department of Veterans Affairs, Washington, DC, USA

Disclaimer: The views expressed in this paper are those of the author and do not represent the official position of the US Government nor the US Department of Veterans Affairs.