Evidence Generation Using Innovative, Technology-Driven Data Collection

Dara Stein, MSc
Senior Research Scientist, Real-World Evidence, Evidera

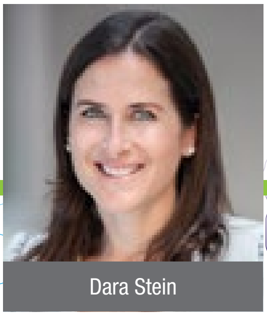
Nafeesa Dhalwani, PhD
Former Senior Research Associate, Real-World Evidence, Evidera

Don O’Hara, MSc
Senior Research Associate, Real-World Evidence, Evidera

Background

Real-world evidence (RWE) is becoming increasingly more important across the pharmaceutical product lifecycle, from advancing the understanding of disease and informing clinical guidelines to supporting regulatory and outcome-based reimbursement decisions.[1] The landscape continues to rapidly shift towards the need for richer and more comprehensive sources of health outcomes.[1] To keep up with the growing demand for RWE there is a need to devise innovative methods to access data and generate robust and reliable evidence.

Electronic medical records (EMR) are now widely implemented in healthcare organizations,[2] and health information exchange (HIE) technology has been widely used to link patient information across different electronic sources. This offers an opportunity to connect to and communicate with EMR systems to extract data in an automated, repeatable, and secure manner for research purposes. Using enhanced HIE-based technologies to extract information from hospital EMRs, researchers get the best of both worlds by ensuring direct access to a rich source of clinical data while removing manual data entry labor, reducing site burden, and maintaining scientific rigor.



Dara Stein

Nafeesa Dhalwani

Don O’Hara

In the Fall 2019 issue of *The Evidence Forum*, we discussed key operational considerations to successfully implement technology-driven solutions for hospital EMR data collection based on our experiences[3] using this approach. The focus of this article is on the advantages, guiding principles, and best practices for using enhanced HIE-based technology to systematically extract data from hospital EMR systems in light of the need for rapid, repeatable, and automated data collection.

What is Technology-Driven Data Collection?

The focus here is on the use of technology to directly identify and extract patient-level data from hospital EMR systems for research purposes. Data extraction software is securely configured to the hospital EMR systems, all the while ensuring that industry best practices for patient privacy and data security are met at all levels. Once configured, the software user interface at the sites communicates securely with the site EMRs and off-site software user interface accessed remotely by researchers (See Figure 1). This allows authorized remote researchers to query multiple hospital EMR systems simultaneously to identify potentially eligible patients for inclusion in research studies, subsequently extract data, and generate queries to clarify ambiguities for enrolled patients. This step replaces the traditional method of having a person manually review and enter data from the EMR into an electronic data capture system.

Advantages of Technology-Driven Data Collection

Technology-driven data collection offers many advantages to traditional methods for capturing data in observational studies. While the use of existing administrative claims and EMR databases is rapid and cost-effective, many databases have inherent limitations due to long time lags between data recording and availability, limited capture of inpatient prescribing and disease-specific biomarkers, and incomplete recordings of risk factors and outcomes.[4,5] The traditional methods of collecting data via manual chart review and data entry by local hospital staff into an electronic case report form (eCRF) overcome some of the limitations of EMR databases, however, this approach is very labor intensive and prone to human error.[6] Additionally, for each new chart review study, a new or updated eCRF needs to be implemented and requires manual data entry by site staff, which is time consuming. The careful selection of key outcome measures is essential to limit site burden, leading to compromises between desired versus feasible data elements to collect in a given timeframe.

> Technology-driven data collection offers many advantages to traditional methods for capturing data in observational studies.

Figure 1. Technology Driven Data Collection Overview

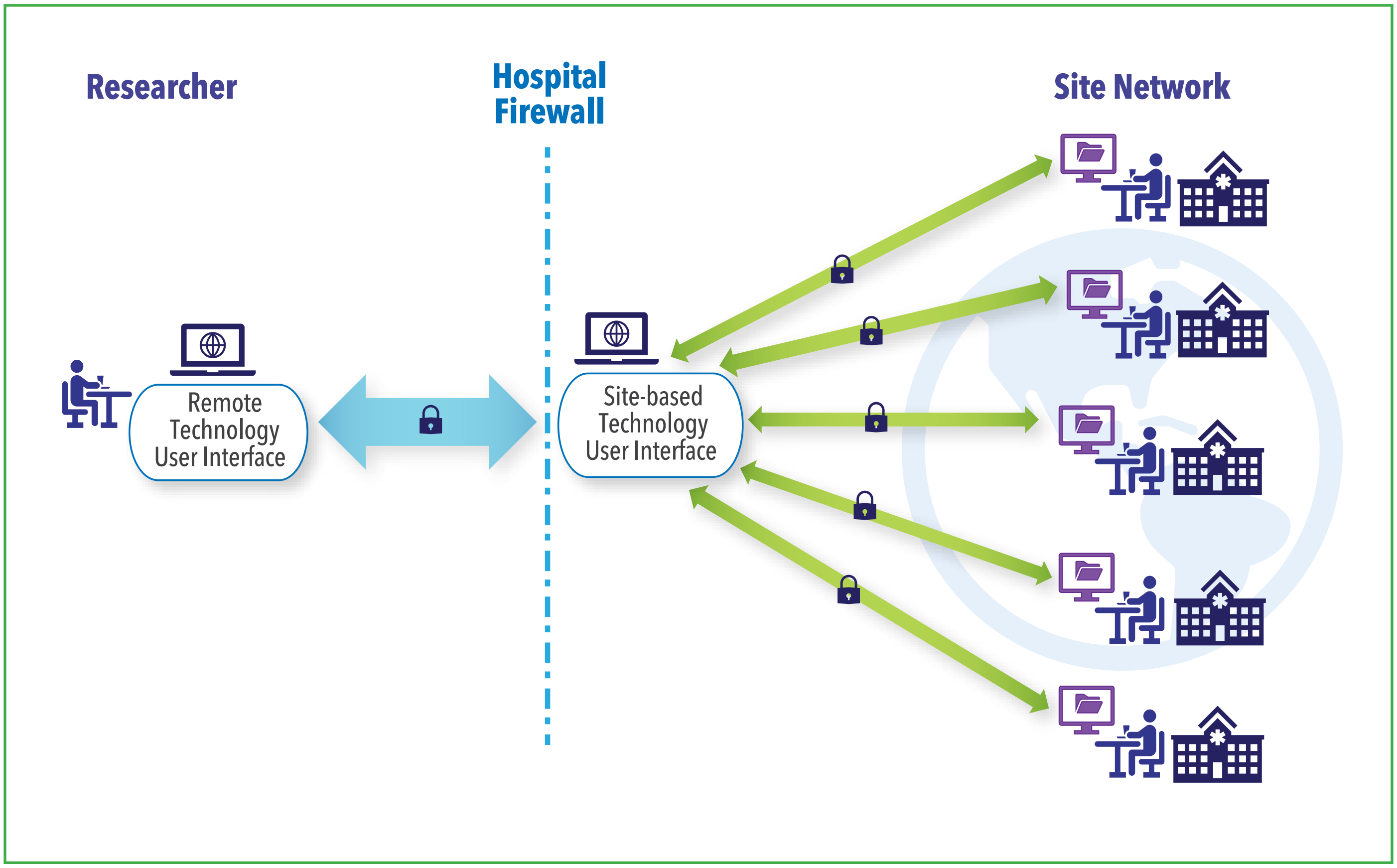
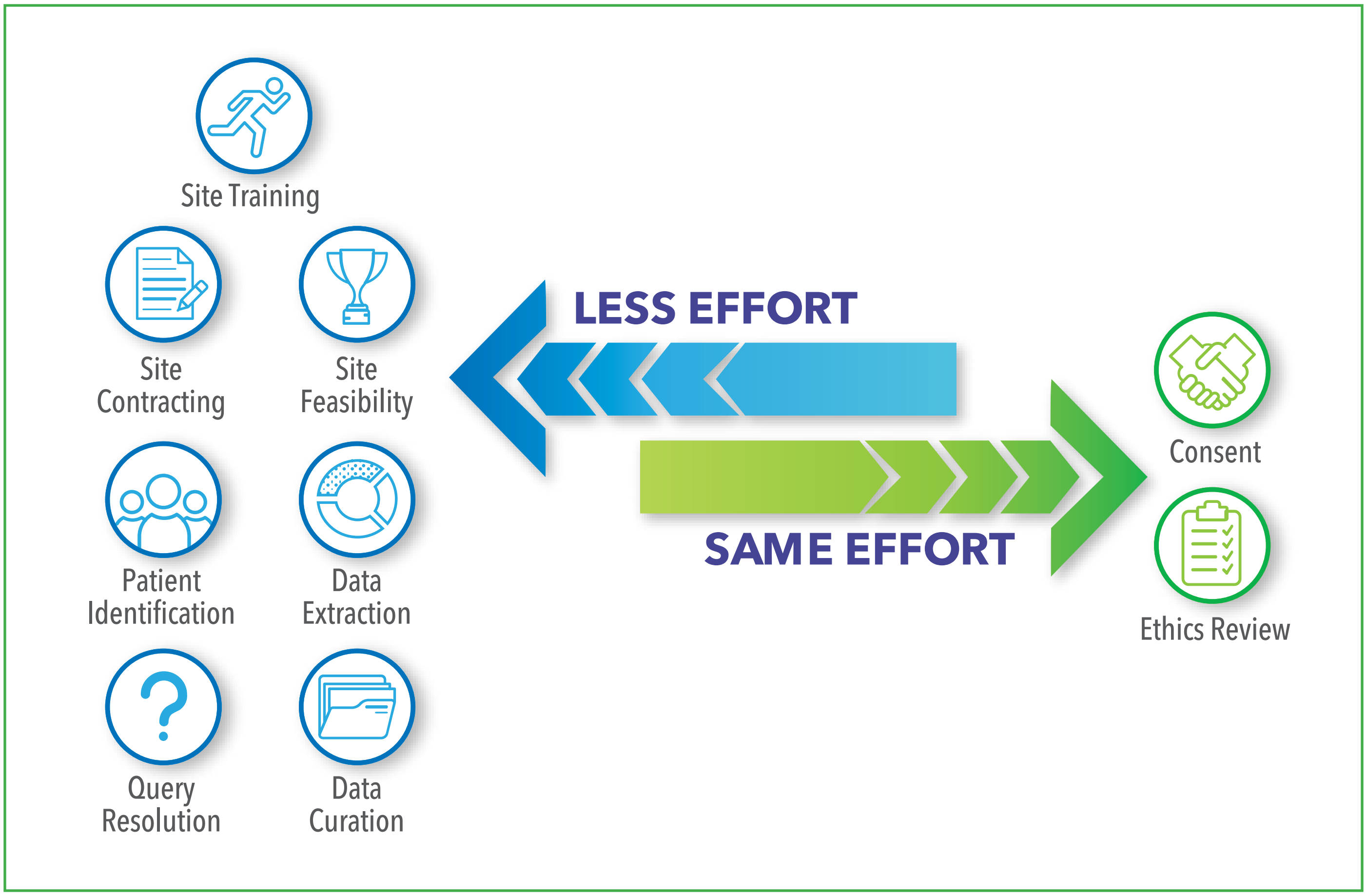

Researcher
Hospital Firewall
Site Network
Remote Technology User Interface
Site-based Technology User Interface

Figure 2. Comparing Level of Effort between Technology-Driven Extraction via Established Site Network vs. Traditional Chart Review Outside Site Network



In comparison, technology-driven data collection facilitates extraction directly from the source to minimize data entry errors, thereby reducing the volume of data queries. It also streamlines data collection, curation, and cleaning processes, as data are extracted directly into standard formats conducive to analyses. The same ethical considerations as chart review studies apply to technology-driven data collection approaches, ensuring the same level of scientific rigor and integrity. Using technology to automate data extraction also allows for the capture of a more rich and deep set of outcome measures in larger patient populations without increasing site burden and workload. It is particularly valuable in prospective studies with the need for future data refreshes because, if the sites are already configured, the process of repeat extractions is simplified. This allows for more streamlined and efficient study set-up and roll-out periods, as well as quicker results. Figure 2 compares the level of effort for study tasks required in technology-driven data collection studies within a pre-established site network with traditional chart review methods. While studies can include extractions at one single timepoint, capturing historical data, greater value comes from the ability to automate repeated extractions at pre-specified, future time-points (e.g., every six months or more frequently). Repeated data access facilitates the evaluation of the changing treatment landscape, as well as long-term clinical and safety outcomes, which cannot always be adequately accomplished in databases with time lags or chart review studies.

While technology-driven data extraction brings several benefits, it is not without its limitations. The main hurdle is finding suitable sites for configuration that also cover large catchment areas and provide comprehensive care to avoid gaps in data on patient care and treatment patterns/outcomes. Furthermore, not all site EMRs may be compatible for setting up the extraction technology. In addition to these limitations, patient privacy and concerns over cyberattacks and the misuse of patient data have been at the forefront of several media outlets in recent months,[7,8] adding further skepticism and scrutiny as a major barrier to technology-driven data collection.

Data Security and Patient Privacy Considerations

Data security must be implemented by means of end-to-end controls embedded into all layers of an application to ensure the protection of information assets: hardware, software, people, and data.[9] All application users and system support staff are trained on information security best practices, and all users are strictly required to follow the policies, procedures, and controls put in place to ensure data security.[10,11] In addition, hardware specifications

> In addition to keeping patient data secure while the data are in motion or at rest, the application needs to be compliant with all local and regional patient privacy regulations.

and software requirements are defined to ensure data are protected, kept confidential, untampered with, and accessible to only authorized users. The application should implement the continuous monitoring of applications to detect and circumvent intrusion or data alteration attempts. Site users and support staff need to have a complete understanding of how the application components are installed and configured in the site infrastructures. Sites must be actively involved from the initial planning phases through configuration, installation, day-to-day operation, and system retirement.[12] All security concerns and mitigation steps are discussed, agreed upon, and signed off on before any solution is implemented.

In addition to keeping patient data secure while the data are in motion or at rest, the application needs to be compliant with all local and regional patient privacy regulations, such as the Health Insurance Portability and Accountability Act (HIPAA)[13] in the United States (US), and the General Data Protection Regulation (GDPR)[14] in the European Union (EU) and United Kingdom (UK). A dedicated de-identification module within the application ensures that all patient identifiable data are transformed into pseudonymized patient data before leaving the site; key data elements and identifiers that could be used to identify a patient are removed from the extracted data.[15] The pseudonymized patient data includes a key-code identifier that can be used only by authorized site users to access the patient identifiable data by means of a look-up table that remains on the site infrastructure. No patient identifiable data are transferred outside the hospital firewall.

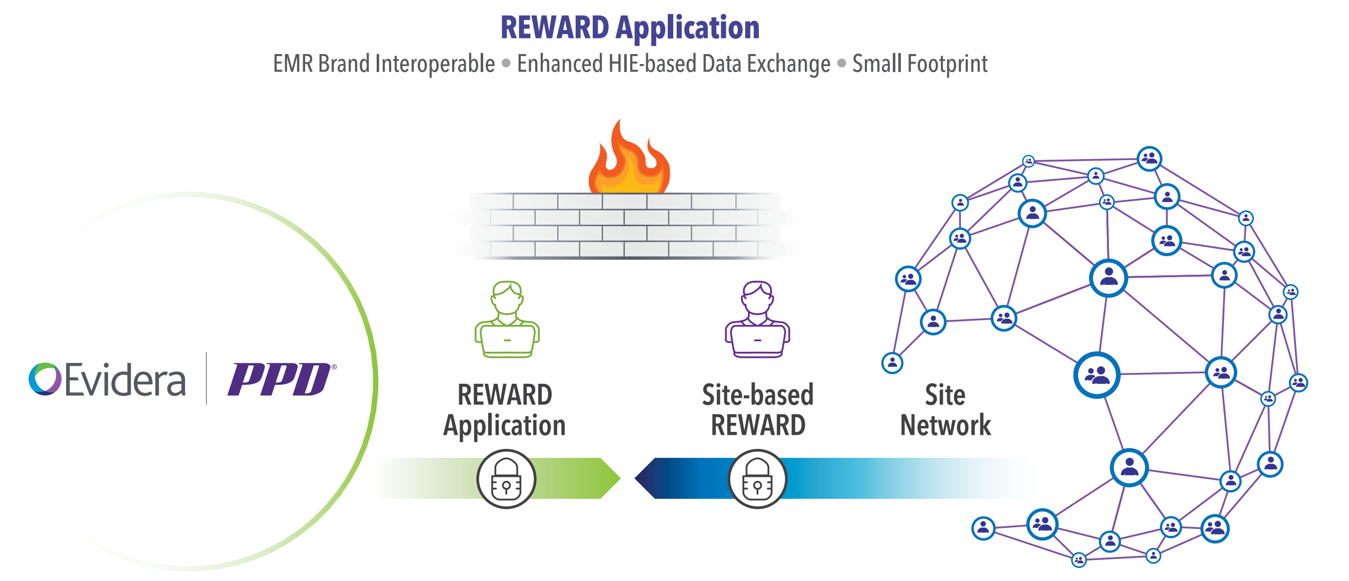
In addition to hardware and software controls used to ensure information security and patient data privacy, healthcare applications must give sites the tools to remain in control of their data. Sites should be allowed to choose the studies in which they would like to participate via an opt-in/opt-out mechanism; within a particular study, sites must have the ability to approve or deny queries from researchers asking for patient counts; and, sites need to be able to approve or deny all patient-level data being extracted. No data aggregate or individual-level pseudonymized patient data can leave the site without site permission. Patient consent should always be requested, where applicable, and study-specific ethics approval will always be sought.[16]

REWARD: Our Approach to Technology-Driven Data Collection

Real-**W**orld **A**ccess to **R**emote **D**ata (REWARD) is Evidera's solution to technology-driven data collection (See Figure 3).

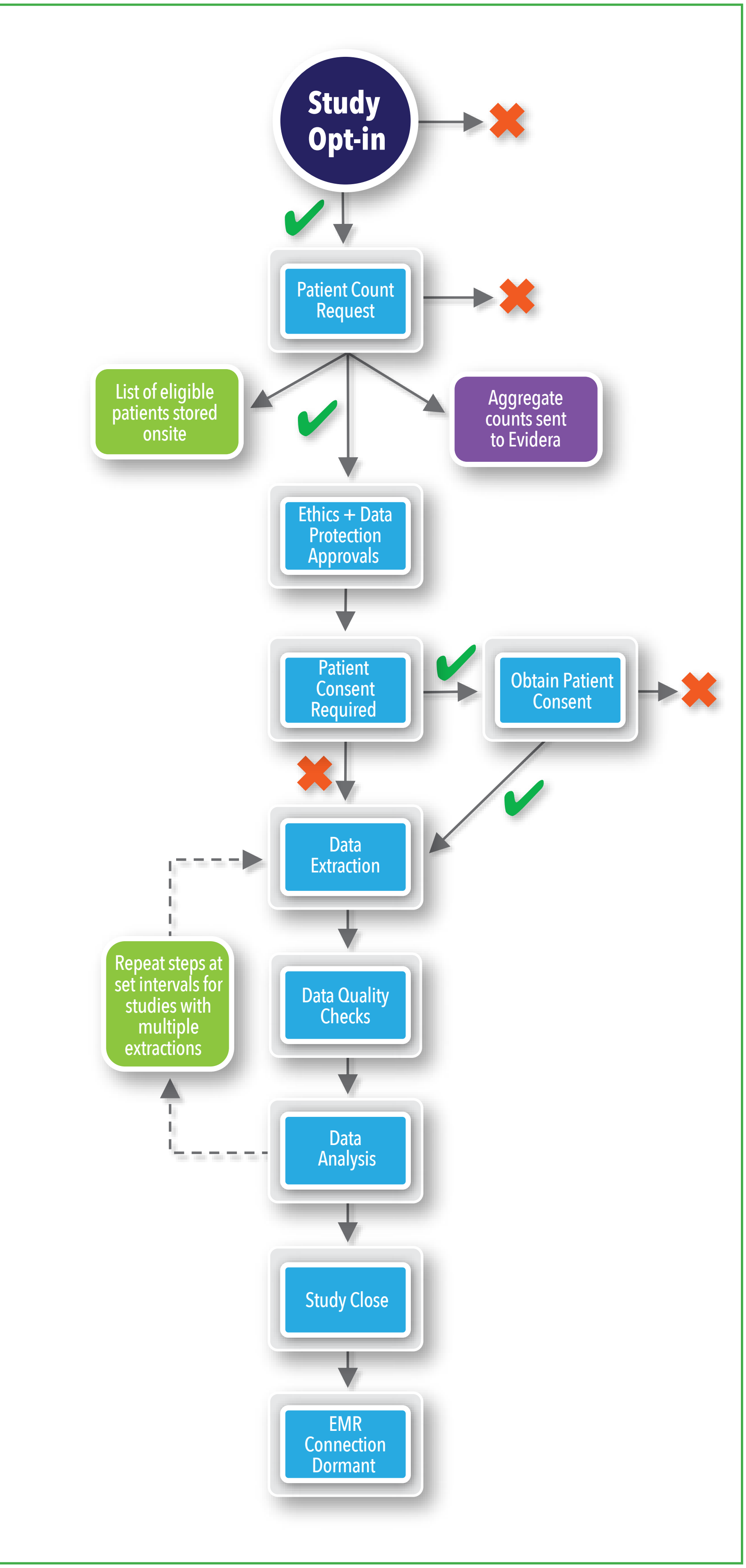
REWARD employs a systematic approach to technology-driven data collection with built-in checkpoints at each step of the study lifecycle. Through REWARD, hospitals

Figure 3. Overview of REWARD



EMR = electronic medical record; HIE = health information exchange; REWARD = **Re**al-world **A**ccess to **R**emote **D**ata

Figure 4. Data Extraction within REWARD



Study Opt-in
✓
✗
Patient Count Request
✗
List of eligible patients stored onsite
✓
Aggregate counts sent to Evidera
Ethics + Data Protection Approvals
Patient Consent Required
✓
Obtain Patient Consent
✗
✗
✓
Data Extraction
Repeat steps at set intervals for studies with multiple extractions
Data Quality Checks
Data Analysis
Study Close
EMR Connection Dormant

keep control of data access and flow, while the application safeguards patient privacy and securely stores data. Once configured, sites are invited to participate in each study through the REWARD site application and can opt-in or opt-out of the study. If sites opt-out of the study, no further contact is made in relation to that particular study. If sites opt-in, then the Evidera user issues a patient count request within the REWARD application to identify potentially eligible patients and obtain initial patient counts. Sites have to approve the request before any active linkage with the site EMR is made and any aggregate counts can be shared with Evidera. Once the request is approved, the REWARD application links to the EMR data and returns aggregate counts to Evidera. REWARD also creates a list of potentially eligible patients and stores that information within the site application; this list, however, is not shared with Evidera. Following ethics approval, sites confirm patient enrollment and consent (when required) using this pre-stored list via REWARD. Sites then approve the data extraction request, at which point data extraction and patient data de-identification is undertaken via REWARD. If subsequent extracts are required, sites will be prompted to approve this within REWARD beforehand. Any queries regarding extracted data are sent to the site for review and comment. Once data extraction and curation is complete, the site becomes dormant until a repeat extract is requested in prospective studies, or the site opts-in to participate in another study.

Summary

Implementation of technology-driven data collection using a systematic approach to research that has checkpoints, safeguards patient privacy, and ensures data security can address a breadth of research questions pertinent to drive drug approvals and improve patient care. In order to build trust, it is integral that hospitals remain the gatekeepers to their patients' data and be in control of data access through all steps of the research study. Short-cuts should not be taken, and full transparency regarding the process is essential for success. ■

For more information, please contact Dara.Stein@evidera.com.

REFERENCES

1. McKinsey & Company. Pharmaceuticals and Medical Products. Cavlan O, Chilukuri S, Evers M, Westra A. Real-World Evidence: From Activity to Impact. Available at: https://www.mckinsey.com/industries/pharmaceuticals-and-medical-products/our-insights/real-world-evidence-from-activity-to-impact-in-healthcare-decision-making. Accessed March 3, 2020.
2. Lambooij MS, Drewes HW, Koster F. Use of Electronic Medical Records and Quality of Patient Data: Different Reaction Patterns of Doctors and Nurses to the Hospital Organization. *BMC Med Inform Decis Mak.* 2017 Feb 10;17(1):17. doi: 10.1186/s12911-017-0412-x.
3. Sawalhi-Leckenby N, Fernandes S, Schabert V. Research Operations for Secondary Use of Clinical Sites' EMR. Fall issue. *The Evidence Forum.* 2019. Available at: https://www.evidera.com/research-operations-for-secondary-use-of-clinical-sites-emr/. Accessed March 3, 2020.
4. Pacurariu A, Plueschke K, McGettigan P, et al. Electronic Healthcare Databases in Europe: Descriptive Analysis of Characteristics and Potential for Use in Medicines Regulation. *BMJ Open.* 2018 Sep 5;8(9):e023090. doi: 10.1136/bmjopen-2018-023090.
5. Schneeweiss S. Understanding Secondary Databases: A Commentary on "Sources of Bias for Health State Characteristics in Secondary Databases." *J Clin Epidemiol.* 2007 Jul;60(7):648-50. Epub 2007 Feb 26.
6. Panacek EA. Performing Chart Review Studies. *Air Med J.* 2007 Sep-Oct;26(5):206-10.
7. McCoy MS, Joffe S, Emanuel EJ. Sharing Patient Data Without Exploiting Patients. *JAMA.* 2020 Jan 16. doi: 10.1001/jama.2019.22354. [Epub ahead of print].
8. Wachter RM, Cassel CK. Sharing Health Care Data with Digital Giants: Overcoming Obstacles and Reaping Benefits While Protecting Patients. *JAMA.* 2020 Jan 16. doi: 10.1001/jama.2019.21215. [Epub ahead of print].
9. INFOSEC. Technical References. IT Security Standards and Best Practices. Available at: https://www.infosec.gov.hk/english/technical/standards.html. Accessed March 3, 2020.
10. Goldstein ND, Sarwate AD. Privacy, Security, and the Public Health Researcher in the Era of Electronic Health Record Research. *Online J Public Health Inform.* 2016 Dec 28;8(3):e207. doi: 10.5210/ojphi.v8i3.7251. eCollection 2016.
11. Uwizeyemungu S, Poba-Nzaou P, Cantinotti M. European Hospitals' Transition Toward Fully Electronic-Based Systems: Do Information Technology Security and Privacy Practices Follow? *JMIR Med Inform.* 2019 Mar 25;7(1):e11211. doi: 10.2196/11211.
12. Jalali MS, Kaiser JP. Cybersecurity in Hospitals: A Systematic, Organizational Perspective. *J Med Internet Res.* 2018 May 28;20(5):e10059. doi: 10.2196/10059.
13. HHS.gov. U.S. Department of Health & Human Services. Health Information Privacy. Available at: https://www.hhs.gov/hipaa/index.html. Accessed March 3, 2020.
14. General Data Protection Regulation (GDPR). Available at: https://gdpr.eu/tag/gdpr/. Accessed March 3, 2020.
15. Neubauer T, Heurix J. A Methodology for the Pseudonymization of Medical Data. *Int J Med Inform.* 2011 Mar;80(3):190-204. doi: 10.1016/j.ijmedinf.2010.10.016. Epub 2010 Nov 13.
16. Caine K, Kohn S, Lawrence C, Hanania R, Meslin EM, Tierney WM. Designing a Patient-Centered User Interface for Access Decisions about EHR Data: Implications from Patient Interviews. *J Gen Intern Med.* 2015 Jan;30 Suppl 1:S7-16. doi: 10.1007/s11606-014-3049-9.