International
CLINICAL TRIALS

November 2021

epc European Pharmaceutical Contractor
STARTS ON PAGE 31

QUESTION TIME

Many different regions' regulatory panels hold challenging examinations of application dossiers

DIVERSE DILEMMA

Adopting a patient-centred approach can improve enrolment and patient experience for diverse populations worldwide

TRIALLING YOUR BEST

The world has changed irreparably, and many researchers are finding that the design of their trial can make a difference

A Vision for Real-World Data Technology

To enable faster access to innovative medical interventions for patients and physicians, part of the strategic approach must be unlocking the full scope of usable, useful information from patient electronic health records (EHR) to generate productive real-world data

Douglas Drake at Clinerion Ltd

EHR Data Integration Produces Building Blocks of the Patient Journey

Drug discovery, development, clinical research, regulatory, and commercial activities are distinct efforts within the pharmaceutical lifecycle, and, therefore, contain distinctly gated process mazes with unique checkpoints, end points, approvals, or dead ends. These mazes often become data silos; data generated within the distinct process stays only within its respective maze and is no longer visible once the therapeutic effort has left that maze.

Increasingly, data integration is critical to effective working across the mazes of global pharma processes, and to providing key insights from early development through commercial considerations. To be effective, data must be fit for purpose and applicable to the specific focus and need, not just single-sourced, hence, possibly one-dimensional. Like defeating a 'Legend of Zelda' game dungeon and acquiring a prerequisite item to move forward, the knowledge gained from each distinct process maze or data silo can be passed on to the next process within the clinical development lifecycle through data integrations,
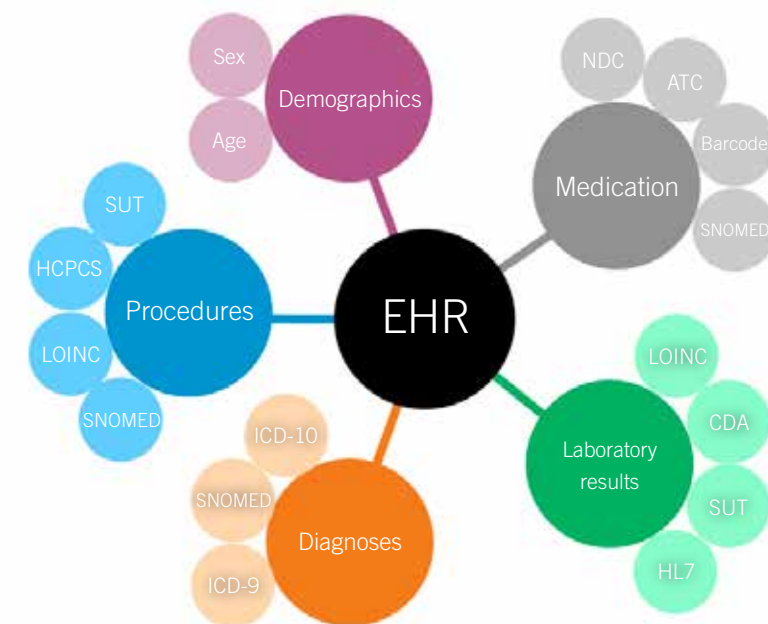
... but data needs intelligence!

Figure 1: Components of EHRs. EHR data codes are maintained longitudinally and in real-time in a hospital information system. We propose a technology that can integrate data ontologies for query design and patient cohort search

therefore, improving the conduct and potential outcome of ongoing or future clinical trials.

The patient journey is a similar case: often, patients are in contact with multiple caregivers, multifarious expertise and procedures that lend themselves to a truly multidimensional perspective and interpretative stratification and healthcare. Within this ecosystem, individual caregivers are individual data silos unless capable of being integrated into the whole patient perspective. This multidimensional aspect is important to generate clinical research data, as well as retrospective patient data. Having a

multiplicity of data from many sources with distinct overlap within the patient journey timeline, it is important to verify and confirm clinical and retrospective observations and conclusions.

Data Integration Is Digital Enablement

COVID-19 has highlighted the need for, and even forced, digital enablement, not only in social interactions and virtual meetings, but also in healthcare. Patient-doctor interactions have been severely disrupted and are evolving into a more mixed ecosystem of direct, indirect, and virtual interactions and metrics. There is now more reliance on digital metrics and technologies that can monitor and report on patients remotely. This directly supports the pre-existing goal of enabling personalised healthcare via measuring and collecting therapeutic outcomes directly, in real time, digitally, remotely, and effectively. Patient care will still benefit from a personal touch, but can be enhanced and better focused through digital technology and data enablement. This is also the core of the smart hospital concept using digital data and the Internet of Things (IoT).

During the COVID-19 pandemic, as most traditional patient-caregiver interactions and care resources were limited, the use of digital technologies and data driven approaches exploded, becoming, in many cases, the only link between caregivers, their patients, their families, and communities, both nationally and internationally. EHR data, traditionally used for billing, became front-line, real-world data metrics and insights to live patient care and risk factor modelling.

EHR data are the traditional metrics of patient healthcare information, e.g., demographics, diagnoses, medications, procedures, and laboratory tests that are administered in a healthcare setting. As each of these metrics has a timestamp, these records constitute an invaluable longitudinal mapping of a patient's journey from diagnosis through treatment to outcome. Hence, over time, they describe the diagnostic interpretation, the implemented therapeutic protocol, follow-up, and reassessment. Accessing these patient metrics in an anonymised fashion, therefore, becomes a critical tool to help understand disease epidemiology, incidence, severity, diagnosis, various treatment patterns, regimens, and outcomes. For epidemiologists and clinical researchers, these data have become critical to developing cohort models to understand and map patient risk factors, follow individual treatment paths or demographic groupings, and stratify patient care better.

In addition to following individual treatment patterns, EHR data are also important to understand and model time indices by segmenting patients before or after diagnosis or treatment, and look into a specific timeframe to develop and test cause-effect models and determine average responses and outcomes. This is critical to show treatment efficacy and the effectiveness of various treatment regimens. Additionally, recent research shows that the time between diagnosis and treatment and the various steps in a patient's medical journey can predict outcomes.

The application of digital data, such as EHR and other resources, has become a standard part of real-world evidence (RWE), i.e., healthcare observational data outside randomised clinical trials (RCTs). EHR data are updated in real-time, i.e., captured contemporaneously in a hospital or care settings. As in the case of the COVID-19 pandemic, this allows timely understanding of patient infections, conditions, and treatment metrics compared to a clinical trial. With the timestamps, EHR data is an RWD resource that allows more longitudinal, retrospective research studies than RCTs.

Additionally, the broader scope of EHR data going beyond a single hospital to a hospital network or global network can be used to develop inclusive and comprehensive artificial intelligence/machine learning (AI/ML) methodologies for sophisticated cohort models and predictive metrics, as well as to identify and track key statistics and time indices within target populations and patient treatment modalities. These new capabilities give better understanding and diagnostic insights to patients and their treating physicians.

What Is a Smart Hospital?

Increasing digital enablement is the name of the game in healthcare. At its simplest, smart hospitals are healthcare institutions that have enabled a digitised network and infrastructure of interconnected assets and healthcare data, and can thereby optimise and innovate new clinical processes, management, and treatment systems that enhance patient engagement, care, insights, and improved outcomes. Smart hospitals make use of not only digital connectivity, but also emerging technologies, such as AI/ML and the IoT, to create more integrated analyses and initiate patient connectivity that better addresses the needs of healthcare providers and their patients.

> *EHRs, as digitally documented records of clinical care, are helpful to identify patient groups and follow their aggregate journey, however, they do not include the deductive reasoning of the physician and interpretive aspects of their care decisions*

Electronic health records (EHRs)

- Longitudinal patient diagnoses, medications, treatments, and procedures
- Originally for claims and reimbursement
- Data can be used retrospectively for research without Ethics Committee approval when anonymised
- Limited to what data are captured in hospital information system

Randomised clinical trials (RCTs) data

- Protocol-driven data capture
- Heavily regulated and controlled
- Data cannot be used retrospectively

RCT

Pseudo control arm data

EHR

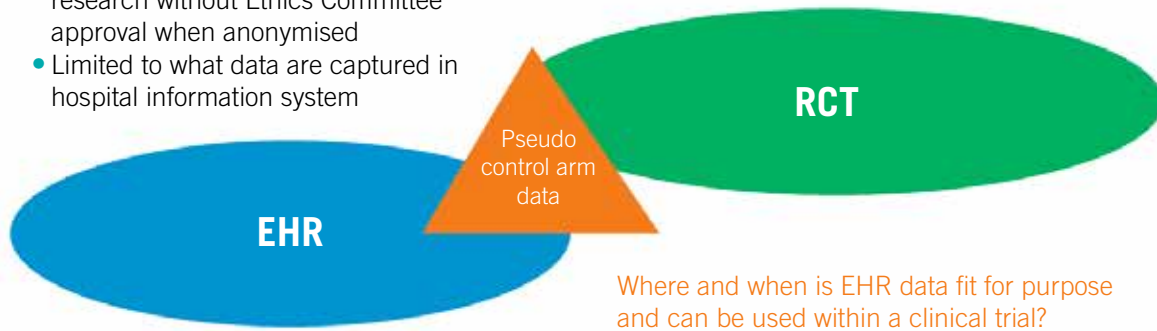Where and when is EHR data fit for purpose and can be used within a clinical trial?

Figure 2: EHRs and RCTs

A key part of smart hospital enablement is unlocking the full potential of the EHR within the hospital information system for ongoing patient care and stratification. EHRs began as an idea of recording patient information in an electronic form and have now become mandated for government healthcare reimbursement in many countries. Physicians, however, cannot always access the full digital patient records easily and may rely on their traditional, manual charts and notes for reviewing and evaluating patients. It may be possible to enable a full digital patient profile from the EHR for physician access and patient care (1).

Real-World Data Drivers and Strategic Responses

Identifying Patient Cohorts
In general, EHR data, being clinically captured data, are high-resolution, but may not match up entirely to the specific metrics needed to be readily applicable as control arm data. Exact fit depends on the trial design, drug regimen, and specific data elements needed in the design, and matching patients in the EHR longitudinal coverage match. For example, time-to-event analysis, such as overall survival and time to next treatment, are hard

to determine from EHR data because of different drug regimens used across patients, and the difficulty of measuring actual outcomes, unless the treating hospital also ensures clinical follow-up and collection of death data.

The Value Proposition of EHR to EDC
The capability to re-identify EHR would allow us to set up electronic data capture (EDC) questionnaires at specific hospitals with cohorts matching clinical trial recruitment criteria. A further update would allow the collection of the necessary fit-for-purpose EHR data elements needed for, e.g., an external control arm or surrogate control arm, or even to create *in silico* computer models of a specific patient population.

Patient Privacy and Data Security Is Paramount
One of the key drivers we see is patient privacy since increased digitalisation also enables an easy transfer of personal data. With data privacy challenged by this development, a model that can query data without moving or transferring them outside a hospital's firewall gains importance as a tool that offers analysis and visualisation of patient care metrics and RWD analytics without privacy concerns.

Privacy is increasingly critical, even when patient data is fully anonymised as re-identification is a constant risk, and some countries prohibit even anonymised data transfers.

Different Regions Have Different Privacy Regulations
A related driver is the difference in privacy regulations in various countries and regions around the world. The difference in patient data privacy regulations between the US (with the most prominent example being the *Health Insurance Portability and Accountability Act* [HIPAA]) and Europe (with their *General Data Protection Regulation* [GDPR]) creates differential data access capabilities between those two major healthcare markets, as well as with the rest of the world. HIPAA allows the use and transfer of de-identified patient data without explicit patient consent. The provisions of GDPR, however, do not allow data transfer or processing of personal data, as per article 4, without an explicit opt-in and consent (2).

Real-World Data Have Value (Even Without Personal Identifiers)

Another driver is understanding that RWD and intelligence on the

patient journey through healthcare can be obtained by an effective use of anonymised patient data. The challenge is doing this in a fashion that ensures patient data privacy and security. Enabling access to EHRs and running on anonymised patient data can generate information on patient journey and outcomes and offer metrics for patient care and stratification, digitally democratising care data directly within care centres and hospitals.

Data Integration, Not Data Harmonisation

Just as in the story of the Tower of Babel, where different languages prevented the completion of the construction, we cannot leverage RWD without a unified data standard. One challenge, as in language, is that the translation of all words into a single ontology or language would result in much subtle, but often important cultural meanings or meta data being lost. Rather than harmonisation, the key is to create linkages between data sources that match the relationships in the data (3).

One example is tiered data relationships between data types. Let us think about an oncology patient for whom a recent MRI is a critical piece of information made up of several levels of data. First, there are the images as the primary data, which are often not meaningful to an untrained eye. Then there is the secondary relevant detail not contained in the image, such as patient information, referring physician, orientation of the scan, use or no use of a contrast agent, treating location, and time of scan. At the third level are the summary details, which include interpretation by the radiologist and physician of the images, which will include exact image interpretation of where and to what extent the tumour grows, from which the physician may determine tumour grade, stage, progression, and response to treatment. The interpretive summary-level data are, therefore, the most important aspect for understanding

the patient journey and treatment decisions the doctor with the patient will make.

Data integration into the full patient journey must link the various summary level elements while also preserving the underlying raw or primary and reference level elements in a traceable manner. Rather than shifting data, the strategy should increasingly link data through relationship and ontology maps that preserve the original data structure and integrity, while enabling insights across the entire spectrum of a patient experience and healthcare journey.

AI/ML Can Identify Diseases That Have Been Hard to Diagnose

Enabling better care models, as well as better identification and mitigation of patient risk and complication factors with specific diseases and treatments requires a further step: the enablement of further data access and analytics directly within the hospital host environment. The goal is to bring insights and metrics out and leave the data in the hospital. The enablement of AI/ML capabilities within the hospital environment will allow not only real-time data surveying, trends, and temporal analysis, but also a platform where learning algorithms can be implemented to iteratively recognise data patterns and targeted searches.

The mechanism of federated learning allows algorithms to be deployed across different sites within the network, and through the cloud connectivity, enable the tools for sharing learning and insights on the respective data available at each node. It allows the original data to always stay in place and only the insights and learnings generated come out of the hospital firewall (4).

Therefore, in addition to the query capabilities of the federated network, we also propose a federated learning AI/ML strategy that will enable model development and iterative learning across disparate hospital nodes. The

common data elements across target sites, by disease, treatment, medication, procedure, and laboratory tests can then be used by embedded algorithms within the network to better understand the overall patient journey.

There Is a Wealth of Information in Unstructured EHR Data

A major limitation of EHRs is the amount of data in them that are unstructured, e.g., in a doctor's free text notes. Useful for them to remember a patient's situation, less useful for data analysis. An important strategy for creation of a fully useful EHR is the development of multilingual natural language processing (NLP) capabilities that will allow us to index data from physician charts and notes as well. These often contain critical, subjective interpretations of the patient's status and response to treatment and our aim is to ensure these also become part of the anonymised record within the hospital, and, therefore, usable for cohort and care modelling. In addition, we foresee these NLP capabilities being important to automating EDC efforts as well, wherein additional data, from anonymised healthcare data or from patient records (in this case specifically under ethics committee approval) can be electronically collected for clinical research and analysis (5-6).

More Unstructured Data: Genomics and Medical Imagery

EHRs, as digitally documented records of clinical care, are helpful to identify patient groups and follow their aggregate journey, however, they do not include the deductive reasoning of the physician and interpretive aspects of their care decisions. For example, EHR data often do not include genetic testing or medical imaging reports, as these inherently need interpretation by the physician and context with the overall patient condition. Standardised performance testing and stage metrics, such as mobility, disease progression, vision, or aptitude, also do not appear in the EHR as these tests are only a subjective metric to facilitate the doctor's interpretation

and do not trigger a diagnosis, treatment, or medication choice in themselves.

These interpretative metrics, or summary-level data, are often in the physician's notes, clinical evaluation, and medical reports, and need to be retrospectively collected either manually or through EDC. NLP is a mechanism whereby context can be derived from digitised documentation. The goal of NLP is to accurately understand the context of language semantics to derive meaning as a human would. The largest initiatives currently are in English, but the need for global healthcare capabilities are efforts with multilingual expertise.

Combining multilingual NLP with multilingual ontology harmonisation will create additional interpretative data insights from non-structured patient data within the physician's notes and additional reports, all still securely within the hospital environment. The objective is to integrate these additional data elements with the patient EHR data to create a complete patient journey with insights into treatment regimen efficacy, while maintaining the original data in the original format at their original source. By combining data maps to understand data relationships and hierarchy within the original data sources, such as genomics and medical images, the data securely stay in context to the patient's journey and their healthcare.

Our goal, by increasing the additional patient record data elements and ontologies, is to enable primarily better healthcare insights, but ultimately better healthcare outcome metrics. We anticipate this enhanced ontology

and data mapping will also facilitate accurate, complete, and automatic EDC functionality from anonymised healthcare field indices. This will enable specific types of data collection as needed for specific geographic or global retrospective data analysis without the need to move data or screen personal and private data within the original records.

Anonymous healthcare data are becoming a cornerstone of RWD to enable RWE claims necessary to validate the efficacy of medical interventions and thereby improve healthcare. However, data security and the mechanism of access for this type of research within the hospital firewall are increasingly critical to protect patient privacy, especially with regard to regional requirements.

In summary:

- Improving patient care and outcomes should be the overarching goal of any technology and digital engagement in healthcare and life science
- The common goal of improving patient care and outcomes is a unifying theme that connects life science research and therapeutic development with direct patient care and treatment, as well as care value and reimbursement
- Digital capabilities present significant enablement throughout the life science and healthcare ecosystem toward better disease and therapy discovery, development, patient treatment, experience, and overall journey, as well as outcomes including both healthcare metrics and cost
- Our common goal with the life science and healthcare industries must be to improve patient care and

clinical effectiveness via enabling and effecting new digital capabilities, connectivity, and experience

References
1. Visit: www.researchsquare.com/article/rs-258174/v1
2. Yuan B and Li J, *The Policy Effect of the General Data Protection Regulation (GDPR) on the Digital Public Health Sector in the European Union: An Empirical Investigation,* Int J Environ Res Public Health *16(6): p1,070, 2019*
3. Zhang, H et al, *An ontology-guided semantic data integration framework to support integrative data analysis of cancer survival,* BMC Med Inform Decis Mak *18: p41, 2018*
4. Sarma KV et al, *Federated learning improves site performance in multicenter deep learning without data sharing,* J Am Med Inform Assoc *28(6): pp1,259-64, 2021*
5. Visit: aijourn.com/doctors-notes-are-critical-for-ai-to-improve-patient-outcomes/?_lrsc=fee8ee3b-33c3-4708-8615-99993c29dee7&cid=other-soc-lke
6. Névéol A et al, *Clinical natural language processing in languages other than English: opportunities and challenges,* J Biomed Semantics *9(1): p12, 2018*

Douglas Drake MS, MBA, Senior Director, Client Solutions at **Clinerion**, is originally a life science researcher with a passion for digital enablement of better patient care. With over 30 years of experience working in various aspects of diagnostics, therapeutic research, and drug discovery, Douglas has broad experience in transformative technologies, data sciences, global business development, and applying these to improving patient engagement and the patient journey.

douglas.drake@clinerion.com

> *Anonymous healthcare data are becoming a cornerstone of RWD to enable RWE claims necessary to validate the efficacy of medical interventions*