PhUSE 2017

Paper RW03

Devices, Big Data and Real World Evidence

Mike Collinson, Oracle Health Sciences, Reading, UK

ABSTRACT

The world of clinical research is experiencing a revolution with a huge range of connected devices growing in popularity, across healthcare, fitness tracking and diet. Pharmaceutical companies sponsoring trials are incorporating these devices into ever more elaborate clinical trials focusing on Real World, Healthcare and Big Data sources. These can be integrated with large scale indexed data stores, running a very different type of distributed processing query, asking unique questions: such as the simulation of a clinical hypothesis, without dosing patients.
This paper will consider how relational and indexed data stores can complement each other, combining transactional information with real world data from indexed systems: organizing, collecting, cleaning, transforming, and aggregating the data to prepare it for analysis and then using various Predictive Analytics and Machine Learning techniques to transform Real World Data into Real World Evidence.

INTRODUCTION

The world of clinical research is experiencing a revolution with a huge range of connected devices growing in popularity, with wearable and implantable devices across healthcare, fitness tracking and diet. Pharmaceutical companies sponsoring trials are incorporating these devices into ever more elaborate clinical trials, and sifting through social media streams for drug safety reaction information and to identify target populations.
In addition, sponsors, technology providers and Standards Development Organizations (SDO)s are working towards Electronic Health Record (EHR) and Electronic Data Capture (EDC) integrations. There can be substantial savings in both time and energy from EHR to EDC integration:

- There may be more than 745,000 data points collected and managed in a production trial.
- Approximately 300,000 data entry key strokes could be saved if as little as 40 percent (40%) of the EDC data could be mapped from EHR.

These numbers also produce savings on the monitoring side of the equation, as data fields in the EDC case report forms (CRFs) sourced from the EHR system don't have to be source data verified - as the data is populated automatically from EHR.

Simultaneously, we have new opportunities in data processing as we move to the cloud. Cloud platforms provide lightning responses to queries from transactional systems, and these can now also be integrated with large scale indexed data stores, running a very different type of distributed processing query, asking unique questions. It is now easier than ever before to store, manage and query ever increasing datasets.

Wearable devices are now being included in up to 50% of clinical protocols. These devices present challenges for data collection and cleaning, and provide sponsors with opportunities to:

- Improve patient trial adherence
- Improve patient engagement
- Improve the granularity of patient data
- Extend the dimensions to describe subject's characteristics

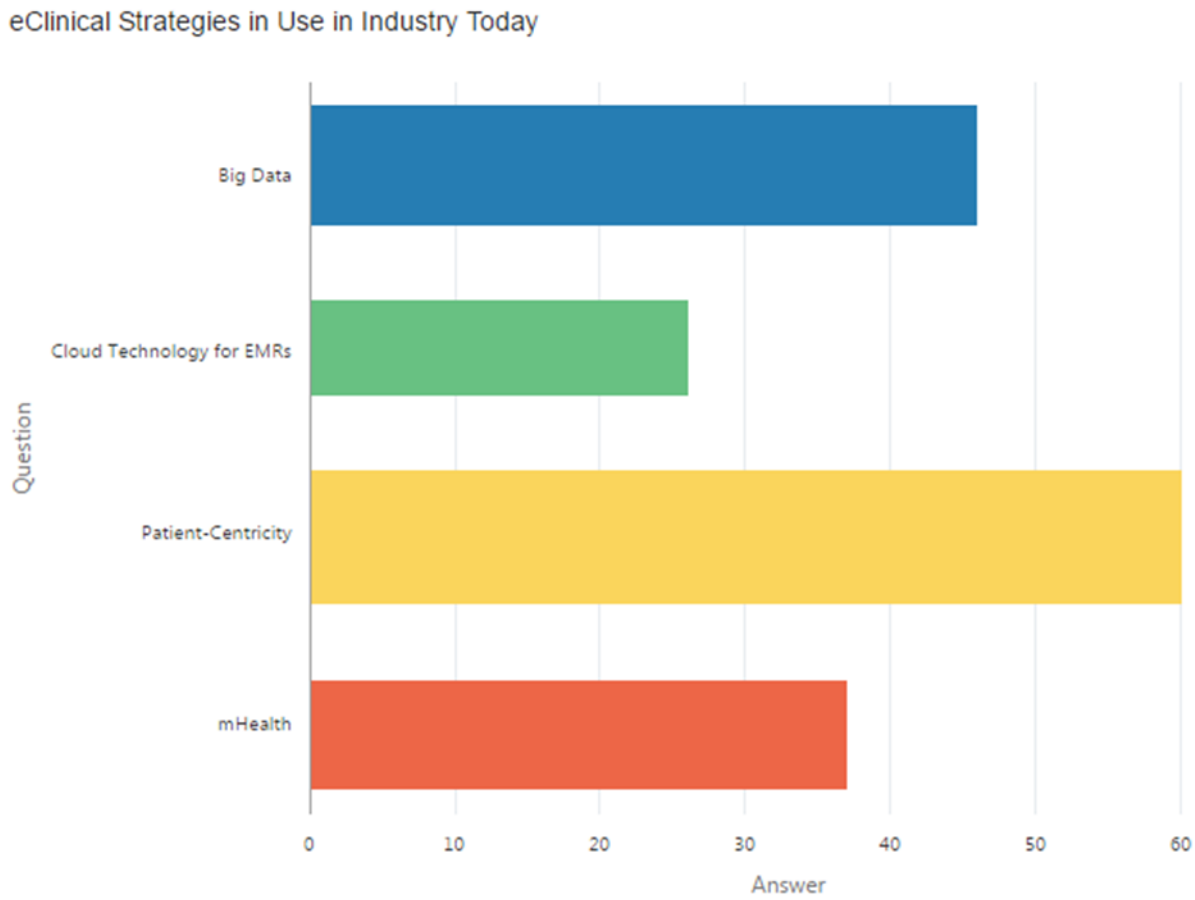

Figure 1: eClinical Strategies in use in Industry today: (Source: https://knect365.com/clinical-trials-innovation/article/e3d64458-2494-490a-b831-8f12abc467ad/eSource-clinical-trials-adoption)

eSOURCE AND DEVICES

In the digital age, our attitude to information is changing. The traditional model of data capture and supply has shifted downstream. Companies now expect a central hub for all of the information related to the trial, and are need a for a single source of the truth, for all of their internal and external stakeholders.

EDC and data management systems can also be integrated with a wide range of esource data sources ranging from wearable devices/sensors such as Pulse Oximetry, Blood Pressure and Activity monitors, to Patient ePRO questionnaires, EHR, and many more. These devices can provide near real-time data to patient and population dashboards.

Figure 2 highlights some of these novel data sources:

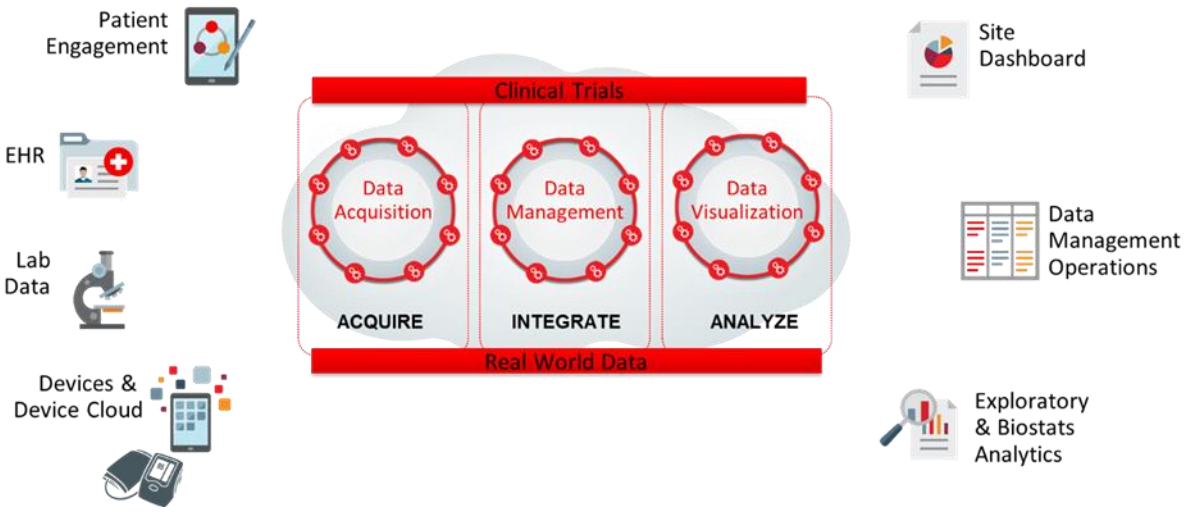

Figure 2 Integration of eSource data

EHR AND EDC INTEGRATION

Sponsors, technology providers and SDOs are working together on EHR and EDC integrations. There can be substantial savings of time and energy with EHR to EDC integration, and many biopharma organizations are looking into integration using the emerging HL7 standard, known as Fast Healthcare Interoperability Resources (FHIR).

FHIR is a specification for implementation of RESTful Services, a technology approach to building APIs common to a number of SDOs. It enables access to patient data in EHR systems in support of system to system communication and interoperability. The FHIR standard is actively under development by the HL7 SDO and is implemented in several EHR vendors' systems: code can be written once to the interface standard and used with many electronic health record (EHR) systems.

Over the last couple of years, there has been a focus on using FHIR RESTful Services to integrate patient care data into the clinical research process. A key use case has been populating EDC system CRFs from EHR systems. This use case can help the biopharma company and its sites participating in a clinical trial to:

1. Reduce overall data entry volume for each clinical trial

2. Improve quality of entered data for each clinical trial, by removing redundancy of manual data transcription from one system to another

3. Reduce clinical trial costs as data populating the CRF from the EHR system does not have to be source data verified.

Each EHR system may only have a subset of the information necessary for a particular clinical trial and that subset may vary, but this approach certainly holds huge promise.

BIG DATA IN CLINICAL RESEARCH

Biopharmaceutical companies are also realizing the value of clinical trial data for secondary use, such as modelling and simulation and integration, in particular when combining and enriching with Real World Data streams:

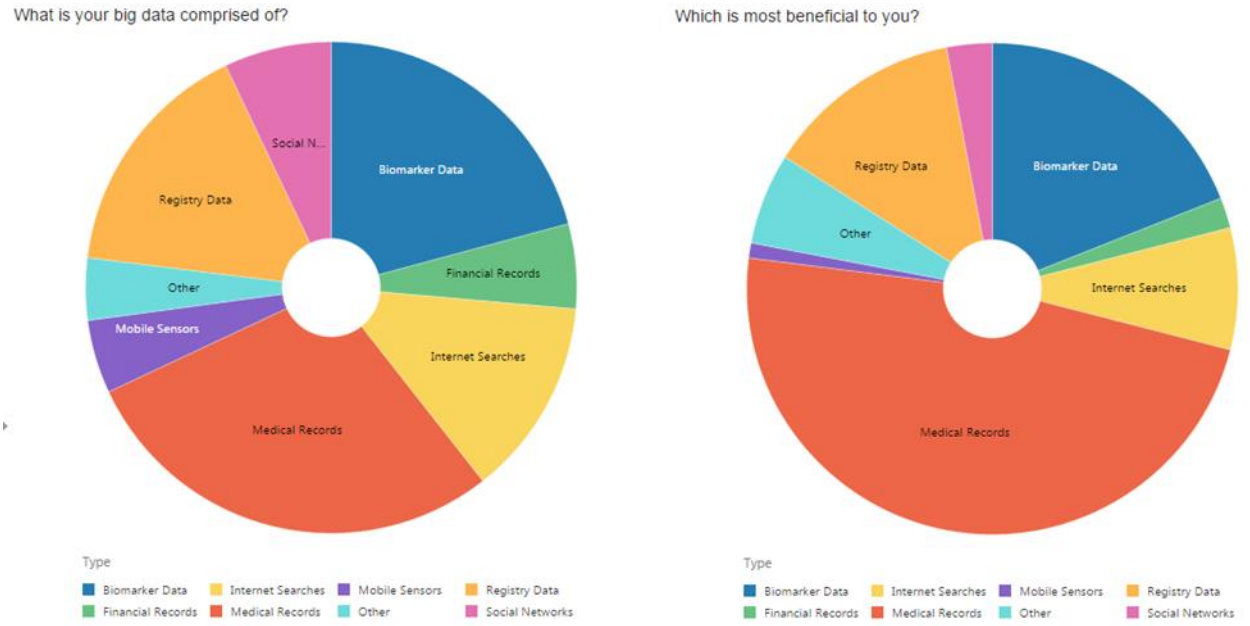

Figure 3: Big Data Survey (Source: http://www.appliedclinicaltrialsonline.com/big-data-survey-report-december-2016)

The European Medicines Agency (EMA) has established a new task force to explore how medicines regulators can use big data to support research, innovation and robust medicines development in order to benefit health, focusing on mapping sources and characteristics of big data, and exploring the potential applicability and impact of big data on medicines regulation.

Big Data streams are characterized by the 3 Vs:

1. Volume – number of transactions / complexity
2. Velocity - number of transactions / time
3. Variety – disparate data types and formats

Machine learning now offers ready to use algorithms to detect patterns and make predictions which readily consume the large volume of data that is now available. With the easy availability of highly scalable compute it is now easier than ever to identify signals across the wide variety of data sets and data structures.

There has been a huge amount of innovation managing Big Data, which is characterized by vibrant ecosystem of open source technologies, including:

- HDFS / Map Reduce technology to store data
- Query editors including Spark, Impala, HIVE, HUE, PIG to query data
- R to analyze data
- Predictive Analytics to visualize data
- Machine Learning to interpret data

A distribution of a commercial Big Data Cloud Service or Big Data Discovery Service will typically many of these technologies packaged together as a big data platform, providing optimal performance, with minimal maintenance

Big Data processing is an agile, iterative and intuitive process:

Step 1 - Integration: Utilize distributed processing and map reduce queries to make disparate data available for processing

Step 2 - Preparation: Manage large volumes of data using intuitive recommendations, transformations and standardization

Step 3 - Discovery: Interactively explore data from a multitude of sources using powerful analytical tools

Integrating the vast volume, velocity and variety of big data will allow users to exploit machine learning and artificial intelligence. Machine learning techniques can create predictive analytical models, as in a recent example assessing risk of hospital re-admission (figure 4).

These analyses can predict the relative rate of re-admission for patients, based on their initial diagnosis. The analysis gives Healthcare Professionals quantifiable insight regarding likely re-admission, which can assist in resource planning.

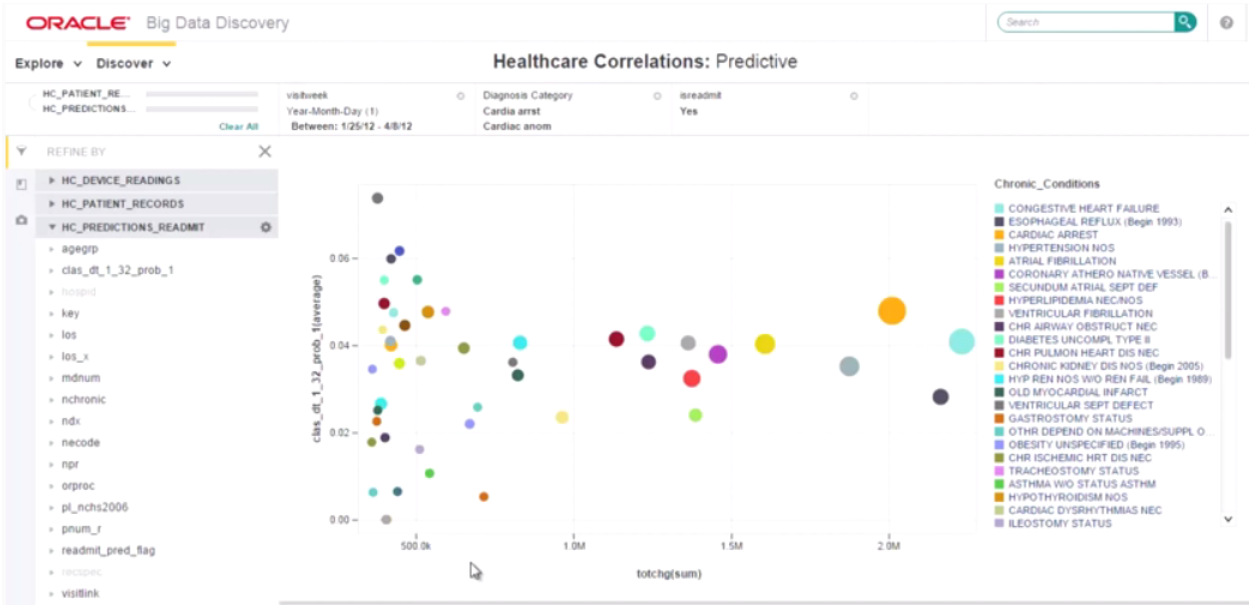

Figure 4: Healthcare Correlations: Predictive (Source: https://www.youtube.com/watch?v=6YR_YPp70cU)

REGULATORY IMPACT

On June 20, 2017, the U.S. Food and Drug Administration (FDA) published a Question & Answers draft document on the use of electronic records and electronic signatures in clinical investigations, focusing on new technologies including devices and EHR. It includes a definition of a data originator - 'wearer', 'technology' or 'EHR', and applies to:

- Electronic systems, including COTS and customized electronic systems owned or managed by sponsors and other regulated entities
- Electronic services, outsourced by the sponsor or other regulated entities
- Electronic systems primarily used in the provision of medical care
- Mobile technology
- Telecommunication systems

The document outlines the considerations for sponsors when processing data for FDA-regulated clinical investigations, stating that Sponsors and other regulated entities should consider whether there are adequate controls in place to ensure the reliability and confidentiality of the data:

- Validation documentation
- Ability to generate accurate and complete copies of records
- Availability and retention of records for FDA inspection for as long as the records are required by applicable regulations
- Archiving capabilities
- Access controls and authorization checks for users' actions
- Secure, computer-generated, time-stamped audit trails of users' actions and changes
- Encryption of data at rest and in transit
- Electronic signature controls
- Performance record of the electronic service vendor and the electronic service provided
- Ability to monitor the electronic service vendor's compliance

The document acknowledges cloud computing:

- If appropriate controls are in place, there are no limitations regarding the geographic location of cloud computing services.
- However, it is critical for sponsors and other regulated entities to understand the data flow and know the location of the cloud computing service's hardware in order to conduct a meaningful risk assessment regarding data access, integrity, and security.

The guidance also makes a critical distinction as to when data is classed as permanent source data, or merely transitional, specific to the device:

- When mobile technology is used in a clinical investigation to capture, record, and transmit study-related data directly from study participants, the data are collected and stored, perhaps for very short periods of time on the mobile technology before being transmitted to the sponsor's EDC system. In some cases, the data may pass temporarily through various electronic hubs or gateways before reaching the sponsor's EDC system.
- FDA considers source data as data that are first recorded in a permanent manner. In general, for data collected directly from study participants through mobile technology, the first permanent record is located in the sponsor's EDC system or the EHR, and not in the mobile technology.

CONCLUSION

- New technologies will allow patients and sponsors to more closely monitor relevant data during a clinical trial.
- These technologies provide direct insight into a remote patient's condition, transforming Real World Data into Real World Evidence.
- Cloud technologies (IaaS, PaaS, SaaS) have transformed Sponsor's access to technologies and processing types.
- Databases and indexed data stores complement each other, bringing transactional clinical data together with real world data.
- Disparate systems will come closer together through use of interoperability standards and messaging protocols, bringing huge value to Clinical Research.
- We will use Big Data tools to organize, collect, clean, transform, and aggregate the data to prepare it for analysis, leverage distributed processing and engage Predictive Analytics and Machine Learning.
- Regulators and SDOs are keen to identify synergies and opportunities and further define how we react to the evolution of patient data.

REFERENCES

SDTM in Business Intelligence, Collinson, PhUSE 2014
Clinical Data in Business Intelligence, Collinson, PhUSE 2016
http://www.appliedclinicaltrialsonline.com/big-data-survey-report-december-2016
https://knect365.com/clinical-trials-innovation/article/e3d64458-2494-490a-b831-8f12abc467ad/eSource-clinical-trials-adoption
https://blogs.oracle.com/health-sciences/entry/real_world_data_vs_real, Jones, 2016
https://blogs.oracle.com/health-sciences/entry/interpreting_big_real_world_data, Streeter, 2017
https://www.oracle.com/industries/health-sciences/products/real-world-data/index.html
https://www.youtube.com/watch?v=6YR_YPp70cU

ACKNOWLEDGMENTS

Thanks to Jim Streeter, Greg Jones, Casey Costley, Srinivas Karri, Jonathan Palmer and Paul Boyd, Oracle Health Sciences.

RECOMMENDED READING

https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM563785.pdf

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Mike Collinson
Oracle Health Sciences
Oracle Parkway
Thames Valley Park
Reading
RG6 1RA
Mike.Collinson@oracle.com