Challenges and Enablers in the Collection of Health Data for use in Phase II-III Clinical Trials

Sheila Kelly

A dissertation submitted to the University of Dublin, in partial fulfilment of the requirements for the degree of Master of Science in Health Informatics

2018

Declaration

I declare that the work described in this dissertation is, except where otherwise stated, entirely my own work, and has not been submitted as an exercise for a degree at this or any other university.

Signed: _____

Sheila Kelly

Date:

Permission to Lend & Copy

I agree that the Trinity College Library may lend or copy this dissertation upon request.

Signed: _____

Sheila Kelly

Date:

Summary

The title of this dissertation is Challenges and Enablers in the Collection of Health Data for use in Phase II-III Clinical Trials, and the main aim of this research is to identify the challenges and to outline the enablers that may enhance the clinical health data collection process for phase II-III clinical trials.

Janet Woodcock, of the FDA's Centre for Drug Evaluation and Research, told a recent workshop that “the clinical trials system is "broken" and there needs to be new ways to collect and utilise patient data” (Woodcock, 2017).

Clinical research is facing many challenges in the collection of health data for use in phase II-III clinical trials. The key elements that will shape health data collection in clinical research are the collection of data securely for secondary use; the development of consistent standards; adherence to applicable regulatory and ethical guidelines and legislation; collaboration across teams and networks; improved end user experience; and the convergence of patient care and clinical research practices (Embi & Payne, 2014).

The shifting clinical research landscape means that researchers are forced to review current practices and look for new ways to collect data. Reviewing current literature and qualitative research through interviews with key informants, the challenges and enablers in collection of Health Data for use in Phase II-III Clinical Trials have been explored. Several key themes emerged, and despite the desire to use data collected in health records as the primary source of clinical research data, obstacles such as regulation and data privacy remain, and no clear path has yet emerged for how the data collection process for phase II-III clinical trials will evolve.

Acknowledgements

I would like to thank my work colleagues for encouraging me to pursue this course of study; my supervisor, Dr Lucy Hederman, for her guidance and patience, and my husband, Tom, and my daughter, Lily who provided me with “A Room of One’s Own” (Woolf, 1929) to complete this work.

TABLE OF CONTENTS

1 INTRODUCTION

1.1 INTRODUCTION

This chapter outlines the dissertation title, provides context on the clinical research industry, outlines the motivation for the research and the research aims, and provides an overview of the dissertation.

1.2 CLINICAL RESEARCH INDUSTRY CONTEXT

The clinical research industry is seeking new ways to manage clinical trials by establishing networks and consortia with healthcare providers, clinicians, and patients through a range of initiatives (Wang & Motti, 2015), with the aim of streamlining the data collection process.

Clinical research is heavily regulated and data collection practices are slow to change. The last technological evolution was the adoption of electronic data capture (EDC) for the direct capture of patient data at hospital sites over 15 years ago. Since then, despite in-roads in the collection of patient reported outcomes electronically, and technological advances such as the release of Apple's Health Records app (Apple Inc, 2018), the current data collection processes have not changed greatly, and there is little interoperability across data collection systems.

The tried and tested processes for the collection of health data for secondary use in clinical research are ripe for disruption. The key elements that will shape health data collection, include the collection of data securely for secondary use; the development of consistent standards; adherence to applicable regulatory and ethical guidelines and legislation; collaboration across teams and networks; improved end user experience and the coming together of patient care and

clinical research practices to develop evidence-based research and practice, and data “contextualisation” (Embi & Payne, 2014).

1.3 CHALLENGES

Clinical research is facing many challenges. Estimated costs to bring a new drug to market are in the region of $2.6bn (Mullin, 2014) over a 12-year period (Torjesen, 2015), and only 14% (approx.) of all clinical trials are approved (Wong, et al., 2018)

Industry leaders, such as Janet Woodcock, of the FDA's Centre for Drug Evaluation and Research, told a recent workshop at the National Academies of Sciences, Engineering, and Medicine that “the clinical trials system is "broken" and there needs to be new ways to collect and utilise patient data” (Woodcock, 2017).

In a survey of over 300 industry professionals, to assess the barriers in the delivery of clinical trials, over two thirds of responders identified challenges in the lack of data visibility and the inability to interrogate related issues in real-time. Too many disparate data sources made it difficult to foresee issues, and the majority favoured the development of unified data platforms such as interactive dashboards for analytics, and the automation of key performance indicators and alerts. All responders identified data quality as their biggest challenge (Hublou, 2016).

1.4 MOTIVATION

In 2015, I was assigned to the International Consortium of Health Outcomes (ICHOM, 2018) which aims to collect health outcomes data to drive improved standards of care. I had been familiar with the collection of health data for secondary use in clinical research using systems and processes prescribed for the clinical research industry. The ICHOM project introduced new research avenues to me by suggesting that patient data could be collected once for both primary care and secondary data use. Today, like many other initiatives that are seeking to streamline the collection of health data for secondary use, the ICHOM project is still in the pilot phase and has not developed a solution that could be implemented for phase II-III trials. My belief is that the clinical research industry will not change itself, and despite efforts to transform the industry from within, it is likely to be replaced by new players in the market such as Google Health (Google, 2018).

1.5 DISSERTATION TITLE & RESEARCH AIMS

The title of this dissertation is Challenges and Enablers in the Collection of Health Data for use in Phase II-III Clinical Trials, and the main aim of this research is to identify the challenges and to outline the enablers that may enhance the clinical health data collection process for phase II-III clinical trials.

The aim of this research is to:

- Outline the current practice for data collection in phase II-III clinical trials with a focus on the collection of eCRF data
- Identify gaps or pain points in this process
- Identify new developments in the industry that are impacting the data collection process
- Review new ways of harnessing health care data for clinical trials, such as direct access to EHRs (Cowie, et al., 2017)

1.6 OVERVIEW OF THE DISSERTATION

This dissertation is presented in six chapters, as follows:

Chapter 2: Background. This chapter provides an overview of clinical research and the four phases of clinical trials in human subjects relating to sample size, trial design, data collection practices and trial duration. A definition of EHR systems, as it pertains to this research, has been provided.

Chapter 3: Literature Review. This chapter outlines the key challenges and enablers in the collection of health data in clinical trials today and how new developments may contribute or hinder the use of health data for secondary use.

Chapter 4: Research Methodology. This chapter outlines the approach to addressing the research question; the research methodology deployed, including the data collection and analysis processes; the reasons why these methodologies were chosen; and the limitations of these methodologies.

Chapter 5: Qualitative Research Results. This chapter outlines the qualitative research findings from a series of interviews with key informants and presents individual and integrated research findings which highlight the key themes emerging as a result of the research.

Chapter 6: Discussion. This chapter discusses the results of the key informants' interviews and compares these results to the current literature.

Chapter 7: Conclusion. This chapter presents the conclusions and recommendations, including suggestions for further research on this topic.

2 BACKGROUND

2.1 INTRODUCTION

2.1.1 This chapter provides an overview of clinical research and details the four phases of clinical research trials in human subjects, including an overview of the data collection processes for phase II-III clinical trials. A definition of EHR, as it pertains to this research, has been provided.

2.2 WHAT IS CLINICAL RESEARCH?

The NHS describes clinical research as “how we develop new treatments and knowledge for better health and care, building the evidence for new approaches that are safe and effective” (NHS, 2017). While Kamateri et al, refer to clinical research as targeting “new and better ways to understand, diagnose, prevent or treat a specific pathological process” (Kamateri, et al., 2014). At any one time, thousands of clinical trials are ongoing. The US government website provides details for approximately 280,000 global clinical trials, of which over 66,000 are active (US National Library of Medicine, 2017).

2.3 THE FOUR PHASES OF CLINICAL RESEARCH IN HUMANS

In humans, clinical research is conducted by performing clinical trials over four distinct phases, as follows:

- Phase I: Phase I studies are performed in a small number of healthy volunteers to test the safety of a new medicine and to refine dosing regiments (Zivin, 2000). Typically, patients attend a dedicated phase I unit that may be located in a large teaching hospital, and clinicians administer the doses and monitor the patient’s reaction to the drug, while documenting all observations

(Sundar, et al., 2018). Typically, phase I trials take several weeks to complete and 70% of drugs progress to phase II (FDA, 2018).

- Phase II: The aim of phase II trials is to assess optimal dosing regimens (Zivin, 2000). The drug is tested in larger populations - usually a few hundred patients who have the condition (ABPI, 2012). Typically there will be a control group and both the patient and doctor will be blind to what patients are receiving the active drug and to those who are in the control group, and receiving placebo (Zivin, 2000). The new drug will be compared against an existing treatment and/or placebo to monitor side effects or any adverse reactions (NHS, 2017). Phase II trials may be split into sub-phases to initially test a sub-group of patients for efficacy (typically up to 50 patients) and phase IIb, where the trial is extended to hundreds of patients (Sills & Brodie, 2009). Over 30% of drugs proceed to phase III (FDA, 2018).
- Phase III: In phase III trials, thousands of patients may be recruited and the drug is compared against an existing medicine or placebo to test for further side effects and to assess if it offers an improvement over existing medicinal products (NHS, 2017). Phase III trials have a success rate of between 25-30% (FDA, 2018).
- Phase IV: Phase IV studies or post-marketing trials are conducted when a drug has been approved by the regulators, but requires longer term follow-up. During a phase IV trial, a larger sample size is reviewed to monitor long term safety, side effects and the drug's effectiveness. Patients are monitored continuously over several years while using the drug (NHS, 2017).

All four phases are presented in Figure 2.1: Four Phases of Clinical Trials.

Phase	Number and type of subject	Questions
1	50–200 **healthy subjects** (usually) or **patients** who are not expected to benefit from the IMP	• Is the IMP safe in humans? • What does the body do to the IMP? (*pharmacokinetics*) • What does the IMP do to the body? (*pharmacodynamics*) • Might the IMP work in patients?
2	100–400 **patients with the target disease**	• Is the IMP safe in patients? • Does the IMP seem to work in patients? (*efficacy*)
3	1000–5000 **patients with the target disease**	• Is the IMP really safe in patients? • Does the IMP really work in patients?
4	many thousands or millions **patients with the target disease**	• Just how safe is the new medicine? (*pharmacovigilance*) • Does the medicine work in the real world? (*real world data collected to demonstrate value*) • How does the new medicine compare with similar medic

FIGURE 2-1: FOUR PHASES OF CLINICAL TRIALS (ABPI, 2012)

2.4 CURRENT DATA COLLECTION PROCESS

Data management for clinical research may be defined as the “collection, cleaning, and management of subject data in compliance with regulatory standards”. Organisations such as the Society for Clinical Data Management (SCDM) and the Clinical Data Interchange Standards Consortium (CDISC) publish guidelines and standards to enable best practice in data collection, processing and exchange (Krishnankutty, et al., 2012).

The case record form (CRF) is the central data collection document used to collate patient information in clinical research. The CRF charts the patient's progress through the clinical trial by using a predefined set of research criteria and assessments (Bellary, et al., 2014). The CRF is annotated in accordance with CDISC or other standards so that the variables collected can be uniquely identified in the database. CRF completion guidelines are provided to clinicians (investigators) to provide context, and an explanation of what data should be entered in each field. Typically this data is collected using electronic data capture (EDC) systems (Krishnankutty, et al., 2012).

In phase I trials, the CRF is small and usually about 10-15 pages for each patient, while for phase II and phase III trials the CRF may be hundreds of pages and data may be collected from other sources such as patient reported outcomes. In post marketing trials (phase IV), other data sources are reviewed and monitored mostly using real world data sources such as patient charts, EHR records and patient reported outcomes data (Arlett, 2016).

2.5 ELECTRONIC DATA CAPTURE (EDC)

In phase II-III trials EDC (Electronic Data Capture) is the term used to describe the system designed to collect the patient health dataset. The EDC system must be a secure, validated system and the data entered undergoes rigorous data verification steps that must be tracked in an audit trial (ABPI, 2012). Data is analysed in accordance with a statistical analysis plan that must be prepared for each phase of the trial (ICH, 1998).

The EDC system must be a validated electronic system, built in compliance with the 21 Code of Federal Regulations (CFR) Part 11 (FDA, 2018). EDC systems must comply with data security best practice to ensure data confidentiality. Any

updates must be attributable to an individual user and stored in the metadata, and detailed in the audit trail (Krishnankutty, et al., 2012).

In an average clinical trial, multiple systems are used to collect patients' health data. Often these systems are incompatible and are configured to optimise workflow for a particular operational department (Hublou, 2016). For example, in a typical laboratory system, patient data is gathered with the purpose of managing patient samples, while in an imaging system, patient data is gathered to manage image sharing and review.

Study investigators enter CRF data into the EDC system remotely. Clinical Research Associates monitor this process and may raise queries on the data to resolve data inconsistencies or to ensure data completeness. Data Managers then review or validate the data to ensure that it complies with the clinical trial protocol and may raise further discrepancies with the study investigators. Additionally, medical terms are classified using commercially available coding dictionaries such as MedDRA for adverse events and medical histories, and Who Drug for medications (Krishnankutty, et al., 2012).

Once all data has been collected and all discrepancies have been closed out and coding is complete, the database is locked. Typically, no data changes are permitted once the database is locked. If a modification is required post-database lock, it must follow the necessary approvals process (Krishnankutty, et al., 2012).

Figure 2.2: The Current Clinical Data Collection Process outlines the many steps involved in data processing, which includes protocol and CRF design, database set-up, edit check programming, data collection and review, data coding, data processing, data verification, data approval, and database lock (Syngeron Ltd, n.d.)

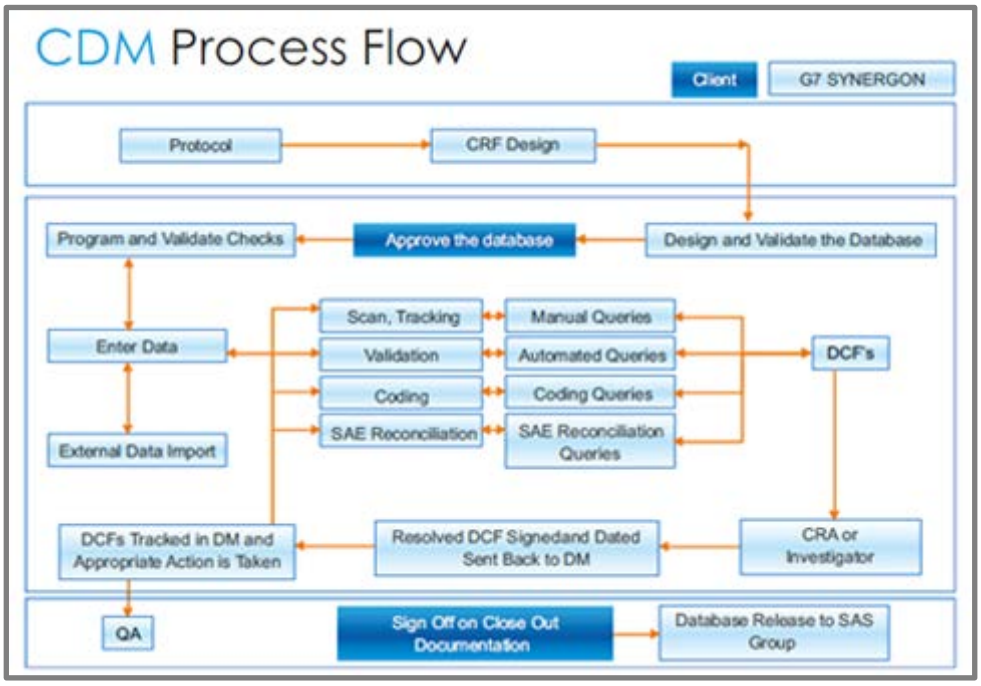


FIGURE 2-2: THE CURRENT CLINICAL DATA COLLECTION PROCESS **(SYNGERON LTD, N.D.)**

2.6 OUTSOURCED DATA COLLECTION

Typically, pharmaceutical companies outsource the data collection process to contract research organisations (CROs), and in many cases only certain aspects of the clinical trial will be outsourced. So, for example, in a phase II or III trial, one CRO may manage the laboratory data, while another CRO may collect the main patient information or case record from (CRF), and the pharmaceutical company may manage some aspects of the clinical trial itself (Wang & Motti, 2015).

2.7 DEFINITION OF EHR

There is no one universal definition of Electronic Health Record (EHR). For the purposes of this research, the term EHR is intended to mean the systems that collect the electronic patient's health record which may include demography, medical history, laboratory test results, clinical notes and other patient data that may be stored in multiple datasets or source systems across individual hospitals or institutions (Kamateri, et al., 2014).

In many cases EHR systems were originally deployed as administrative or billing systems and evolved into tracking patient data either in specific therapeutic indications or hospital-wide. Today, EHR systems contain patient-related health data such as demography, vital signs and other data related to their illness or wellbeing such as responses to quality of life instruments or home monitoring devices (Cowie, et al., 2017).

The WHO explains that the term EHR has different meanings in different regions, but can be defined as a record of a patient's health information that is recorded by healthcare providers and "extends beyond acute inpatient situations including all ambulatory care settings at which the patient receives care" (WHO, n.d.).

ISO's Integrated Care EHR (ICEHR) defines EHR as "a repository of information regarding the health status of a subject of care in a computer processable form" (WHO, 2006), while the US government refers to the EHR as the patient's electronic health record that includes their medical history and other information such as demographic data, medications, and laboratory test results (CMS.gov, 2012).

3 LITERATURE REVIEW

3.1 INTRODUCTION

This chapter outlines a review of the current literature with an emphasis on the key challenges and enablers in the collection of health data in clinical trials.

3.2 SEARCH CRITERIA

An initial search was conducted online mainly in NCBI (NCBI, 2018) which centred on the title of the dissertation - challenges and enablers in the collection of health data for use in phase II-III clinical trials. This yielded limited results so the search was adjusted to seek key themes in data collection for clinical research, and the search criteria used is detailed in Table 3.1: Literature Review – Search Criteria and Results).

TABLE 3-1: LITERATURE REVIEW - SEARCH CRITERIA & RESULTS

Literature Review Search Criteria	Returned Search
Artificial Intelligence/Machine Learning/Blockchain	Anisingaraju, S., 2017. Genetic Engineering & Biotechnology News. [Online]
	Bookbinder, M., 2017. Clinical Informatics News. [Online] Available at: http://www.clinicalinformaticsnews.com/2017/09/29/the-intelligent-trial-ai-comes-to-clinical-trials.aspx
	Nugent, T., Upton, D. & Cimpoesu, M., 2016. Improving data transparency in clinical trials using blockchain smart contracts. *F1000 Res,* 1(5), p. 2541.
Clinical research data & results	Moore, D. et al., 2000. How generalizable are the results of large randomized controlled trials of antiretroviral therapy? HIV Medicine, 1(UNK), pp. 149-154.
	Phelan, A. et al., 2016. Exclusion of Women of Childbearing Potential in Clinical Trials of Type 2 Diabetes Medications: A Review of Protocol-based Barriers to Enrolment. *Diabetes Care,* 39(6), p. 1004–1009.
	Bourgeois, F. et al., 2017. Exclusion of Elderly People from Randomized Clinical Trials of Drugs for Ischemic Heart Disease. *Journal of the American Geriatrics Society,* 65(11), pp. 2354-2361.
	Johnson, S. B., Farach, F. J., Pelphrey, K. & Rozenblit, L., 2016. Data management in clinical research: Synthesizing stakeholder perspectives. *Journal of Biomedical Informatics,* 60(Unk), pp. 286-293.
Clinical research data standards	Richesson, R. & Nadkarni, P., 2011. Data standards for clinical research data collection forms: current status and challenges. *Jamia,* 107(UNK), p. UNK.
	Krishnankutty, B., Bellary, S., Kumar, N. & Moodahadu, L., 2012. Data management in clinical research: An overview. *India Journal of Pharmacology,* Mar-Apr(44 (2)), p. 168–172.

Literature Review Search Criteria	Returned Search
	Sinaci, A. & Erturkmen, G., 2013. A federated semantic metadata registry framework for enabling Interoperability across clinical research and care domains. *Journal of Biomedical Informatics,* 48(UNK), pp. 784-794.
Clinical research data/design	Frieden, T., 2017. Evidence for Health Decision Making — Beyond Randomized, Controlled Trials. *The New England Journal of Medicine,* 377(UNK), pp. 465-475.
	National Academy of Sciences, 2009. *Data Collection Standards and Monitoring.* Washington DC, NCBI.
Clinical research regulation	ICH, 1998. *Statistical principles for clinical trials (E9).* s.l., ICH.
	WMA, 1964; amd 1975, 1983 & 1989. *Declaration of Helsinki,* Helsinki; Tokyo; Venice; Hong Kong: World Medical Association (WMA).
	ICH, 2015. *INTERNATIONAL CONFERENCE ON HARMONISATION OF TECHNICAL REQUIREMENTS FOR REGISTRATION OF PHARMACEUTICALS FOR HUMAN USE.* UNK, ICH.
Data collection in clinical research	Lu, Z. & Su, J., 2010. Clinical data management: Current status, challenges, and future directions from industry perspectives. *Open Access Journal of Clinical Trials,* 2(UNK), pp. 93-105.
Data volumes	Bhadani, A. & Jothimani, D., 2016. Big Data: Challenges, Opportunities, and Realities. In: M. In Singh & D. Kumar, eds. *Effective Big Data Management and Opportunities for Implementation.* Pennslyvania: UNK, pp. 1-24.
EDC & data collection in clinical research	Shah, J. et al., 2010. Electronic Data Capture for Registries and Clinical Trials in Orthopaedic Surgery: Open Source versus Commercial Systems. *Clinical Orthopaedics and Related Research,* 468(10), pp. 2664-2671.
	Tufts, 2018. *Tufts Centre for the Study of Drug Development.* [Online]
EHR & clinical research	Cowie, M. et al., 2017. Electronic health records to facilitate clinical research. *Springer,* 106(Unk), pp. 1-9.
	New, J. et al., 2018. Putting patients in control of data from electronic health records. *BMJ,* 360(5554), p. UNK.
	Patel, V. & Kaelber, D., 2014. Using aggregated, de-identified electronic health record data for multivariate pharmacosurveillance: A case study of azathioprine. *Journal of Biomedical Informatics,* 52(UNK), pp. 36-42.
	EHR4CR, 2015. *EHR4CR.* [Online]
	Murphy, E., Ferris, F. & O'Donnell, R., 2007. An Electronic Medical Records System for Clinical Research and the EMR–EDC Interface. *Investigative Opthalmology & Visual Services,* 48(10), pp. 4383-89.
EHR case studies	Centre for Drug Evaluation & Research (FDA), 2017. *Drug Safety Priorities 2017,* Silver Spring, MD: FDA.
Health data sharing	EC, 2018. *eHealth News.* [Online]
Health data standards/EHR & clinical research	De Moor, G. et al., 2015. Using electronic health records for clinical research: The case of the EHR4CR project. *Journal of Biomedical Informatics,* 53(UNK), pp. 162-173.
New data sources	EUPATI, 2015. *Data Collection in Clinical Trials.* [Online]
	OECD, 2016. *OECD.* [Online]
New technologies in clinical research	Byers, C., 2017. *The Growing Importance of IT in Healthcare.* [Online]
Patient privacy	Maldoff, G., 2016. *The International Association of Privacy Professionals.* [Online]
Patient reported outcomes	Atherton, p. et al., 2016. Patient Reported Outcomes Questionnaire Compliance in Cancer Cooperative Group Trials (Alliance N0992). *Clinical Trials,* 13(6), pp. 612-620.
	Watson, C. et al., 2015. data capture via mobile electronic clinical outcome assessments (ecoa): Meeting New Demands for Patient Reported Outcomes

Literature Review Search Criteria	Returned Search
	(PRO). *Journal of mHealth,* UNK (UNK), p. UNK.
Real world data/evidence	Clinical Research Corporation, 2017. *Clinical Research Corporation.* [Online]
	FDA, 2018. *Real World Evidence.* [Online]
	Getz, K., 2017. Anticipating the Groundswell of RWD and RWE Usage. *Applied Clinical Trials*, 1 December, p. UNK.
	Siemens, T., 2016. How Will Technology Drive Global Clinical Trial Change by 2025? *Applied Clinical Trials,* 24(12), p. 42.
Secondary use of health data	Auffray, C. et al., 2016. Making sense of big data in health research: Towards an EU action plan. *Genome Medicine,* 8(71), p. UNK.
	Danciu, I. et al., 2014. Secondary use of clinical data: The Vanderbilt approach. *Journal of Biomedical Informatics,* 52(UNK), pp. 28-35.
	EMA, 2015. *European Medicines Agency.* [Online]
	EOSC, 2018. *EOSC.* [Online]
	Embi, P. & Payne, P., 2014. Advancing methodologies in Clinical Research Informatics (CRI): Foundational work for a maturing field. *Journal of Biomedical Informatics,* 52(UNK), pp. 1-3.
	Regalado, A., 2015. *Technology Review.* [Online]

3.3 Challenges in the Collection of Health Data for Secondary use In Phase II-III Clinical Trials

Current data collection practices in clinical research are facing many challenges. A survey across 35 institutions cited data quality, data standards, finance, and availability of data for researchers as the main barriers to the secondary use of data (Danciu, et al., 2014).

Concerns related to insufficient clinical trial data, complex regulations and technologies that are not meeting data collection needs, the collection of too much data or data that is never used and missing data (National Academy of Sciences, 2009), are driving researchers to seek out new ways to harness health data for secondary use in clinical research.

EHR data is seen by many as the most sufficient solution to use existing health records as the primary source of clinical research data, but issues related to

disparate data standards and a lack of interoperability abound with such an approach.

3.3.1 Current Clinical Trial Data May Not be Optimal

For a new drug to be approved, the data collected as part of the clinical research process must prove that the therapy is safe and efficacious (National Academy of Sciences, 2009). Clinical research is often criticised for providing results that are not “generalisable” in larger populations and ignoring the longer term safety or effectiveness of a treatment. There is a growing concern that the data collected in clinical trials that was once thought to be applicable to larger populations, may not be so (Moore, et al., 2000). The majority of trials exclude those over 80 years old (Bourgeois, et al., 2017), children and pregnant women (Phelan, et al., 2016), and it is suggested that randomised trials gather insufficient data to explore longer term safety, relevance or effectiveness of a treatment in larger populations (Frieden, 2017).

3.3.2 Clinical Research Regulation

Clinical trial regulation drives the data collection process today. Organisations such as the International Conference on Harmonisation’s (ICH) Good Clinical Practice (GCP) guideline outlines an international standard for the design, conduct and reporting of clinical trial data across phase II to IV trials.

Regulations are placing increased responsibilities on clinical research. The ICH E6 R2 addendum has proposed new standards for the collection of data in clinical trials and extends the principles of confidentiality and quality to electronic records. Standard operating procedures must outline, in detail, the system specification, user requirements and the system validation steps as well disaster recovery and back-up, and system decommissioning procedures. Additionally,

data integrity and contextual data must be preserved, particularly for electronic records. The investigator must have continuous access to CRF data, and this may be achieved by the centralised monitoring of investigator sites by using computerised systems to review data and site performance remotely (ICH, 2015).

3.3.3 Standards in Use in Clinical Research

One of the main challenges in the clinical data management process is the standardisation of data from different organisations while keeping up to speed with technological changes (Krishnankutty, et al., 2012). CRF design may be based on any number of standards, such as CDASH, but no one standard exists that can be applied to CRF design and the current standards will need to evolve to meet the successful collection and aggregation of CRF data (Richesson & Nadkarni, 2011).

The National Cancer Institute (NCI) and the American College of Cardiology has defined standards for domain-specific common data elements (CDEs) to measure disease-specific data. The NCI's Cancer Data Standards Repository (caDSR) attempts to curate CRF questions using ISO/IEC's standard. Other standards include CDISC's CSHARE CRF repository, and the Agency for Healthcare Research and Quality (AHRQ) records data standards in the USHIK (United States Health Information Knowledgebase) database (Richesson & Nadkarni, 2011).

The ISO/IEC standard has been criticised due to the one-to-one relationship of its data elements, and it is incompatible with universally accepted coding dictionaries such as the Systematised Nomenclature of Medicine Clinical Terms (SNOMED CT) or the Unified Medical Language System (UMLS) (Richesson & Nadkarni, 2011).

CDISC's operational data model (ODM) addresses some of the limitations of the ISO standard but does not integrate CRF generation properties within its metadata and does not allow for the validation of free text. ODM also lacks disease-specific context which is required for indication-specific use and effective data exchange or aggregation. CDISC is focused on the development of standards for clinical research, and largely ignores the secondary use of EHR data and the interrelationship between EDC and EHR data (Richesson & Nadkarni, 2011).

Initiatives such as the Health Information Technology Standards Panel (HITSP), the Federal Health Information Model (FHIM), the Clinical Data Interchange Standards Consortium (CDISC) and Mini-Sentinel, are focused on standardising data, but it is noted that they do not automatically provide interoperability across multiple domains (Sinaci & Erturkmen, 2013).

3.3.4 EDC Systems – Lack of Standardisation, Increased Data Volumes and a Lack of Interoperability

EDC systems can be costly to licence and do not readily allow for customisations or interoperability with other systems, and may be cumbersome to configure. Each deployment of an EDC system requires a team of database managers and data coordinators to install, configure and operate the system, which can involve considerable running costs (Shah, et al., 2010).

Shah et all points out that the "success of a data collection initiative depends on the quality of the data captured", and recommends aligning the EDC workflow with the in-clinic practices, but notes that EDC systems have not been widely adopted by clinicians perhaps due to the wide variety of systems available and the variation in specification across different systems (Shah, et al., 2010).

The increase in clinical research data volumes is presenting technical and workload challenges, and timelines to lock a clinical trial database have increased by approximately 10%. Surveying over 250 industry professionals, Ken Getz comments that “the volume of clinical data being collected from numerous and disparate data sources has grown dramatically in response to the rising scope and complexity of clinical trial protocols”. Additionally, there are challenges in loading third party data into EDC systems and, on average, six separate systems are used to collect data in any one clinical trial (Tufts, 2018), leading to EDC systems as being “interconnected but not interoperable” (Lu & Su, 2010).

3.3.5 EHR Data and Clinical Research

Several projects are underway to explore the use of EHR data for clinical research, and potentially as an eCRF substitute for randomised controlled clinical trials. It is recognised that this may pose challenges and that studies are required to assess the validity of EHR data for this purpose (beyond the use of critical variables) (Cowie, et al., 2017).

EHR data may be inconclusive if a patient dies outside of the institution and the data is not captured in the EHR system. A patient may attend more than one clinician or provider for treatment, so it is likely that no one EHR system will contain their complete health record, and hospitals may have to develop solutions to integrate systems. Additionally, different standards related to the terminology, the type of data captured, and how the data is managed or coded, and the lack of interoperability may hinder the development of a complete patient record (Cowie, et al., 2017).

3.3.6 Health Data Standards

While patient care records provide valuable information, there is often inconsistency in standards and data collection methods, and the data may be incomplete and lacking quality steps.

Various data standards are used by hospitals. Medics develop therapeutic-specific standards that may not be compatible across disease indications. Efforts are underway by organisations such as the OpenEHR foundation to define data archetypes to standardise all elements related to an attribute or series of attributes, while the Logical Observation Identifiers, Names, and Codes (LOINC) standard proposes a series of unlinked data concepts. Health Level 7's (HL7) Reference Information Model (RIM), used in many hospitals to exchange patient data, partly maps to standard medical dictionaries, while ODM does not recognise that coded data may have a different route across systems. Coding dictionaries such Snomed CT do not organise data into order sets and may be open to interpretation and leading to coding discrepancies (Richesson & Nadkarni, 2011).

Citing the lack of standardisation, "it is apparent that widespread incompatibility of the many data standards currently used by the clinical research and healthcare communities continues to hinder the efficient and rapid exchange of data between different electronic sources and compromises the quality of clinical trial results" (De Moor, et al., 2015).

3.3.7 Limitations of EHR

There have been some successful deployments of EHR data for clinical research. The EHR system is often used as a data source in post-marketing trials, but EHR data is not yet commonly used in phase II-III trials as the primary research data.

In one example, in England there are several EHR systems in use, but different coding dictionaries are used by general practitioners and hospitals. A subset of this data is gathered in static datasets for clinical research, but they generally they are not linked to EHR systems and are not updated frequently (New, et al., 2018).

In Europe, faced with challenges such as differing regional languages, health care practices and legislation; the EHR4CR consortium embarked on querying EHR data across 11 providers. With the primary aim of assessing protocol feasibility and patient recruitment for clinical research studies, the project worked with fragmented patient records and a myriad of proprietary health systems with few common standards. It focused on 75 EHR common data elements, and data was aggregated in accordance with demography. No individual patient data left the provider site (De Moor, et al., 2015). To date, this model is not used widely (Murphy, et al., 2007).

Other initiatives by PCORnet, the NIH and the Farr Institute of Health Informatics Research have shown how EHR data may be used for “comparative effectiveness”, but Cowie et al highlight this was because of the engagement of committed and knowledgeable stakeholders (Cowie, et al., 2017).

Challenges remain in widening the use case of EHR data for use in clinical trials, such as data quality, completeness, as patient’s records may be across different systems and providers, and in many cases EHR systems are not integrated or interoperable (New, et al., 2018).

De Moor et al, identifies the main barriers that must be overcome before EHR data is accepted within clinical research, include data privacy; regulatory requirements; institutional requirements and policies, and data quality (De Moor, et al., 2015). While Getz highlights that despite the increase in the use of real-

world data, from sources such as EHR systems, this data has yet to make in-roads in the support of key trial decisions (Getz, 2017).

In the US alone, there are over 250 EHR system providers. A lack of EHR standardisation and the inability to share EHR data across platforms has raised both security and safety concerns. Initiatives such as the US government's attempt to build a clinical data electronic exchange was abandoned in 2013, although there has been some success in the direct use of EHR data for registries (Siemens, 2016).

3.3.8 Data Privacy

The intent of the General Data Protection Regulation (GDPR) is to harmonise regulations for the processing of personal data for residents of the European Union, and this legislation impacts on how data is managed for research purposes.

Data controllers and processors must ensure that sufficient safeguards are in place to control and protect personal data from misuse and the regulation strengthens the level of "explicit" consent that must be provided by individuals for their data to be used for research purposes. Unlike the Health Insurance Portability and Accountability Act (HIPAA) which outlines a set of standards for the protection of health data (UHSS, 2013), anonymous data is considered to be out of scope of the regulation. Additionally, the transfer of personal data outside of the EU is not permitted unless these jurisdictions offer the same level of protection, although the implementation of standard contractual clauses may permit the transfer of data in certain circumstances (Maldoff, 2016).

3.4 Enablers in the Collection of Health Data for use In Phase II-III Clinical Trials

There is shift in how clinicians and legislators view the secondary use of patient data for clinical research. The informed patient is driving change, and US legislation has mandated how patient data should be collected and a greater number of institutions are engaged in research. New technology has enabled the increased availability of new data sources and several initiatives are underway to develop common data standards. There have been many projects that have successfully developed solutions for an interchange of health data for secondary use in clinical research.

3.4.1 Growing Recognition of the Value of Secondary Use

The Declaration of Helsinki outlines the principle of combining patient treatment with clinical research while ensuring that the wellbeing of the patient safety is of primary concern (WMA, 1964; amd 1975, 1983 & 1989).

Key trends across published health informatics research papers include the secondary use of patient data for clinical research; a growing awareness of data management and standards; and greater researcher support. In the US, the Hitech Act (2009) mandated that providers show “meaningful use” of certified EHRs in US, and has opened new frontiers for the secondary use of patient data (Danciu, et al., 2014).

3.4.2 Increasing data availability

As the EU Commission plans to increase the availability and sharing of health data (EC, 2018), there has been an increase in the number of public and private initiatives to explore the better use of health data for primary and secondary use (Auffray, et al., 2016). Public health projects such as European Open Science

Cloud (EOSC) (EOSC, 2018), the EU Innovation Network (EMA, 2015) and the 1m Genomes Project (Regalado, 2015) are all seeking to drive better availability, sharing and linking of health data. While the growing number of technology solutions (Byers, 2017) coupled with growing data volumes (Bhadani & Jothimani, 2016) creates opportunities in how this data is used.

Murphy et al identified a growth in the use of clinical data repositories for research purposes (Danciu, et al., 2014). The increase in the use of EHR data and the application of data standards, will improve the delivery of care. (Embi & Payne, 2014).

3.4.3 EDC Systems

Defining EDC systems "as computerised systems designed to collect clinical and laboratory data in an electronic format", but noting that they may also include patient reported outcomes systems and voice-activated systems, Shah et al identifies several benefits of these systems. EDC systems have reduced overall data processing costs by lessening the number of data errors or missing data and shortening the timeline for data collection and processing (Shah, et al., 2010).

Commercially-available EDC systems are designed to prevent against unauthorised access by defining user roles and access rights. They offer a secure, industry-recognised workflow, which is intuitive to the user and allow for quick access to reports and metrics (Shah, et al., 2010).

Within clinical trials today only a subset of patient data is captured in EDC systems, although in a more structured way (Richesson & Nadkarni, 2011)

3.4.4 Developing Common Standards

Johnson et al's survey of clinical research professionals recommends that ideally clinical trial data should be interoperable across systems and that technology should enable this process while enhancing data quality (Johnson, et al., 2016)

It is recommended that a narrowed focus on a defined set of variables that are common to EHRs and clinical data collection systems should be conducted. Focusing on EHR variables such as mortality is seen as the first step in the re-use of EHR data for clinical research, as these variables tend to be less open to interpretation. Other recommendations include involving key stakeholders, identifying targets, and ensuring time for training and education (Cowie, et al., 2017).

For EHR data to be useful in clinical research there needs to be an understanding of what the data means. In addition, data variables or attributes need to be defined across the two domains of patient care and clinical research for the data to have real purpose in both domains (Sinaci & Erturkmen, 2013).

Richesson and Nadkarni emphasise that a "move from a mode of primarily reacting to clinical researchers' needs through service provision, to one of active leadership by suggesting directions for standardisation," is required by clinicians at clinical research sites (Richesson & Nadkarni, 2011).

Sinaci and Erturkmen propose an ISO-conformant framework where common data elements (CDEs) can be linked and re-used across domains using "federated semantically enabled metadata registries (MDR)", and where there is a unique definition of each data attribute. Using Linked Open Data (LOD) to share data, each data element is uniquely identifiable and is linked semantically. Data could be extracted using specifications that would allow for "dynamic interoperability". Citing several instances where local MDRs already exist, such as the METeOR MDR in Australia, they propose, that the EDC system queries

the MDR to search for the required patient data using SDTM (Sinaci & Erturkmen, 2013).

3.4.5 New Data Sources

Several data sources for use in phase II-III trials have emerged in recent years to complement the eCRF data, such EHR, electronic patient reported outcomes (ePRO) and data from wearables. The FDA actively encourages the use of real world data (FDA, 2018) from sources such as electronic health records (EHRs), claims and billing data, patient registries (as represented in figure 3.2: Sources of Real World Data (Clinical Research Corporation, 2017), and their Sentinel program promotes the collection of wearables data in clinical research.

3.4.5.1 EHR

New et al identifies several benefits of EHR data. Notably, EHR data is captured contemporaneously removing “recall bias” and the potential to link data captured in primary care to data for clinical research may provide a useful longitudinal patient record (New, et al., 2018). With the move towards pragmatic trial design, both the NIH and PCORI are assessing if additional data collection modules can be added to the data that already exists in the EHR. Ultimately, the aim is to reduce costs and patient burden by reducing the number of assessments and in-clinic visits within a clinical trial and centering on the collection of variables used in routine clinical care delivery (Cowie, et al., 2017)

3.4.6 ePRO

ePROs have been successfully deployed in research with good patient compliance rates (Atherton, et al., 2016). Both EUPATI and the OECD's PARIS initiative (OECD, 2016) recommend the use of electronic patient-reported outcomes (ePRO), as they result in the better data quality and provides visibility to clinicians as to the patient's condition in real-time (EUPATI, 2015).

The FDA in its guidance document for the collection of patient-reported outcomes has highlighted that electronic outcomes measurements data can improve "data quality, reliability, integrity, and traceability" (Watson, et al., 2015)

The growth of mobile technology and Bring Your Own Device has allowed for the collection of real-time, time-stamped, and personalised data directly from a diverse group of patients, in a secure way. Patients can use their own devices and access questionnaires and diaries using PIN codes, and users can be verified remotely. Control measures can be included to prevent backfilling of questions and to ensure data is contemporaneous, as well as guiding the patient to complete blank fields and to follow logical paths depending on their last response. Bluetooth technology allows for the mobile phone to collect objective data from other devices such as spirometers and activity trackers (Watson, et al., 2015).

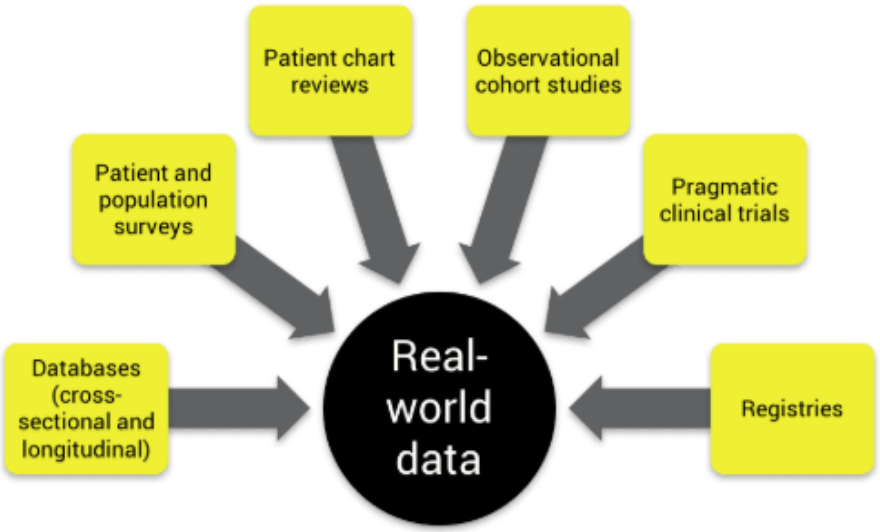


FIGURE 3-1: SOURCES OF REAL WORLD DATA **(CLINICAL RESEARCH CORPORATION, 2017)**

3.4.6.1 WEARABLES

The growth in the use of wearable devices is estimated to drive over $30bn in revenue by 2018, and this market is already providing data in real-time and helping to standardise data collection. By 2025, it is predicted that data flows will improve, but it will bring with it increased data privacy challenges (Siemens, 2016).

3.4.7 Successful Deployment of EHR for Clinical Research - Case Studies

Driven by the 21st Century Cures Act (2016) in the US, which aims to reduce the time for drug and medical device approvals, the FDA is collecting EHR data from several sources to drive post-marketing surveillance (Cowie, et al., 2017).

The Sentinel Common Data Model is based on a distributed data model, where data is maintained locally by the hospital or institution (Centre for Drug Evaluation & Research (FDA), 2017). In Europe, EHRs have been successfully been used for post-marketing trials, and the EU-ADR is linking databases in Denmark, Italy, The Netherlands, and the United Kingdom to analyse drug events in EHRs, while the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) is assessing risk using several EHRs (Cowie, et al., 2017).

There have been many successful implementations of EHR data use in clinical research. In the case of Patel and Kaelber, they implemented this process in a post-marketing surveillance study, while the Vanderbilt Institute built an extraction process to provide datasets to researchers. Further details on these case studies are outlined below:

3.4.7.1 Patel and Kaelber Case Study

In a post-marketing surveillance of hospital data using a commercially available platform, Patel and Kaelber retrospectively reviewed data for over 10 million patients with anti-rheumatologic prescriptions from EHR, billing and laboratory systems. The data was retained within the walls of the participating institutions and the Explore system was used to interrogate the data and return de-identified patient data using medication orders to signify drug administration. Testing the assumption that certain

side effects related to toxicity may not be commonly reported, the team identified a control cohort, and then compared symptoms across patients on different anti-rheumatic medications, and identified “clinically useful patterns” of potential drug side effects (Patel & Kaelber, 2014).

3.4.7.2 VANDERBILT INSTITUTE CASE STUDY

In the early 90s, the Vanderbilt University Medical Centre (VUMC) implemented new systems to move away from paper medical records and build a large data repository for 2 million patients. Their central Office of Research Informatics (ORI) now provides web-based access to this data for a fee to researchers. VUMC has built a three-layered architecture using their enterprise data warehouse as the backbone to this service, with a data layer used to organise and anonymise data, while the research layer allows for researchers to access HIPAA-compliant de-identified data for research purposes (figure 3.3: The Vanderbilt Case Study: IT Architecture) (Danciu, et al., 2014).

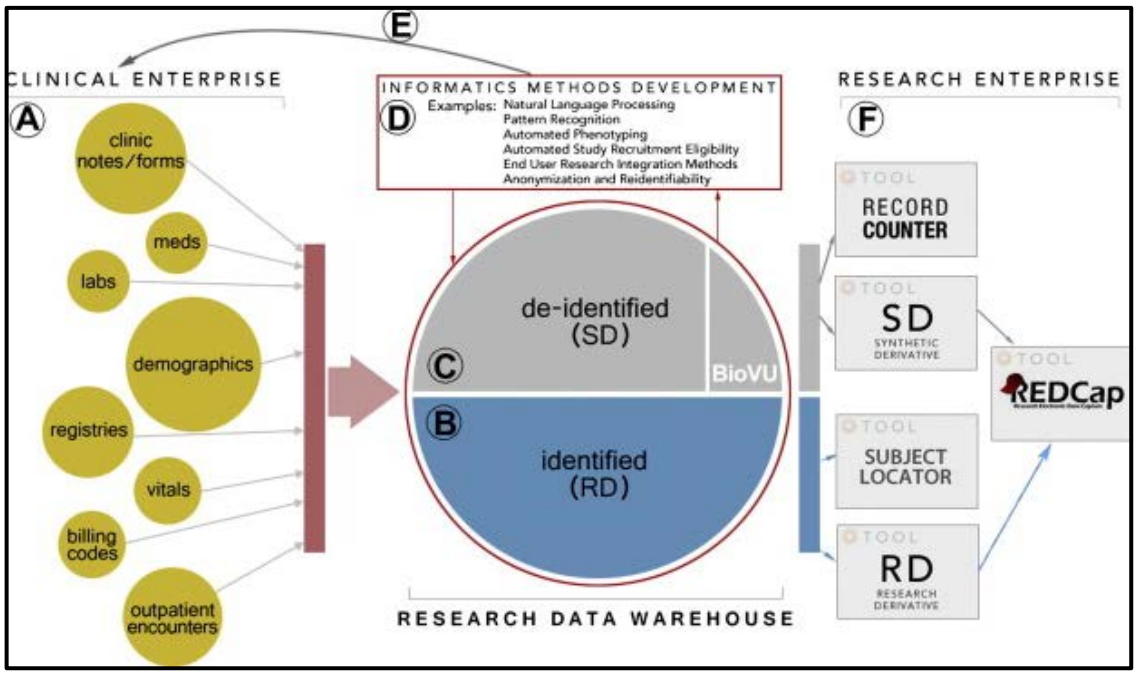


FIGURE 3-2: THE VANDERBILT CASE STUDY: IT ARCHITECTURE

The Vanderbilt case study of building “a dedicated research infrastructure” has outlined the need to understand the clinical significance of data beyond its fundamental role in the direct treatment of patients. The onus is on the institution to collect and manage data in accordance with best practice. The speed of access to reliable data, the ability to extract data safely and securely coupled with the validation of extraction techniques is of paramount importance in research, as well as ensuring that individuals with the right clinical and technical expertise are managing the data effectively (Danciu, et al., 2014).

3.4.8 New Technologies

Advances in technology can transform the clinical trial data collection process by giving researchers direct access to pertinent data and potentially reduce timelines and costs (Bhavnani, et al., 2017). The introduction of new technologies such as artificial intelligence to analyse historical data (Bookbinder, 2017), and blockchain to make data interchange more secure and result in better data quality (Nugent, et al., 2016) is promising. While applied machine learning and robotic process optimisation may improve overall trial feasibility processes by introducing efficiency in how site data is assessed and analysed (Anisingaraju, 2017).

3.4.9 The Patient is Driving Change

The voice of the patient as an educated and informed stakeholder is an important step in developing hospital data for use in clinical research. Change will be driven by the patient as a consumer, who is willing to participate in research, and the growth of large scale data enterprise initiatives in large provider institutions (Danciu, et al., 2014).

Adequate processes should be in place to protect patients and keep them adequately informed (Cowie, et al., 2017). Emphasising the benefits of an opt-in approach and providing guidance for a successful model, it is recommended that patients have full control of their data through a “health accounts” model where the patient can access their data (across providers) and update access permissions. In Salford (UK), patients were asked to opt-in to share data across providers. Less than 1% of over 20,000 patients approached declined and the exercise has been successful, with plans to implement an extended program across almost 3m patients in the greater Manchester area. In Sweden, data is collected in a standard way across healthcare providers (using Snomed)) and

this data is fed to large registries and to clinical research organisations (following ethics committees' approval). Patients or their carers have access to their EHR data and patients must opt-out if they do not want their data to be used for research purposes. In the Netherlands, most patient data is collected in electronically and shared between primary and secondary care providers (New, et al., 2018).

3.4.10 Ensuring Patient Privacy

Cowie et al recommends that distributed analyses where data remains at source may combat privacy issues, which he outlines as complex, particularly across different geographies. If EHR systems are to be used for clinical research, the patient must provide informed consent and the data must be accessed by regulatory agencies for auditing and verification purposes. Further recommendations include a "general" consent process, although this would require legislative changes as well as changes in overall hospital practices and ethics board approvals. (Cowie, et al., 2017)

3.5 CONCLUSION

This chapter explored the relevant literature with a focus on the challenges and enablers in the collection of health data for secondary use in phase II-III clinical trials. Several themes emerged which were used to as a framework for the research methodology (Chapter 4) and as the basis for the key informants' interviews (Chapter 5).

4 RESEARCH METHODOLOGY

4.1 INTRODUCTION

This chapter covers the research methodologies employed, the identification of key informants, the ethics approval process, the key informants' selection process, why these methods were chosen, and the approach to data collection and data analysis. A discussion section is included on the limitations of the research.

4.2 RESEARCH METHODS EMPLOYED

Two core methods of research are presented:

4.2.1 Literature Review

Sackett refers to using an "evidence-based" approach in clinical practice, and that best judgment is made by combining a continual review of literature (Sackett, 1981). In line with this approach, a review of the literature related to health data collection for clinical trials, with a focus on current and new initiatives as well as the challenges and enablers in these processes was conducted and is presented in chapter 3.

4.2.2 Qualitative Research

Qualitative research was used to develop a greater understanding of the research question, and to understand how the issues raised by the research are being addressed (Creswell, 2007).

Using non-probability, convenience sampling, semi-structured interviews with key informants in the clinical research industry were conducted. The key informants

were known to the researcher and were available for interview (Dudovskiy, 2017).

Using a phenomenological approach (Waters, 2017), participants were interviewed in their "natural setting" (Creswell, 2007), and were invited to provide answers based on their own experience or opinion, with the intent of providing a personalised view that may impact on future decision-making (Lester, 1999) in the collection of health data in clinical trials. The researcher's questions were used to guide the interview and additional questions were posed when new themes emerged during the course of each discussion (Creswell, 2007). The results of the key informants' interviews are outlined in Chapter 5.

4.3 INTERVIEW QUESTIONS

Three interview questions were developed based on the researcher's direct experience of the challenges faced in the collection of health data for clinical research while working in the provision of data management services to the pharmaceutical industry. The interview questions were developed to identify common issues or themes in the current state of clinical trial data collection and explore new avenues that may bring benefits to how health data is collected for clinical research. These questions are presented in Table 4.1: Key Informant Interview Questions.

TABLE 4-1: KEY INFORMANT INTERVIEW QUESTIONS

No	Interview Question
1	Janet Woodcock, Director of FDA's Centre for Drug Evaluation and Research recently remarked that the clinical trial data collection process is broken. Do you agree with this statement, and what do you see as the barriers or issues that exist today in how data is collected in phase II-III clinical trials?
2	In your opinion, what factors are changing how health data is collected in phase II-III clinical trials? What are the advantages of such changes? What are the disadvantages of such changes?
3	In the longer term, what innovations may impact how data is collected for phase II-III clinical trials? What advantages will these innovations bring? What issues may these innovations fail to address?

Identification of Key Informants

Key informants were identified using non-probability, convenience sampling (CIRT, n.d.), through the researcher's prior interactions with these individuals and/or based on their experience in the industry. Candidates such as senior managers of key clinical research data management providers, technology leaders in the clinical research industry; and other key opinion leaders from the pharmaceutical industry and/or academia were approached based on their job

title, their specialisation or area of interest, and their role in the industry (Marshall, 1996).

4.4 ETHICS APPROVAL

Ethics approval was obtained from the School of Computer Science and Statistics, Trinity College Dublin (TCD) in January 2018, using their guidelines. Participants were provided with a key informants' information sheet that outlined the purpose of the research and how their responses would be used, the interview questions and the expected duration of the interview. They were advised that they should present their personal views of the topics discussed (not their employers'), that their participation was voluntary and that they could withdraw at any time. Additionally, they were informed that the interview would be recorded and transcribed, and that they would be presented with a copy of final transcript that they could amend, if so wished. Participants were asked to sign a consent form. The ethics board approved the submission, with minor amendments, in February 2018. A copy of the ethics submission package is attached in Appendix I.

4.5 KEY INFORMANTS SELECTION PROCESS

Eleven key informants were approached on LinkedIn, and asked if they would consider participating in the research. Seven key informants agreed to proceed by return email, and the key informants' information sheet and consent form was sent to them by email for their review and consent. All key informants' agreed that their names could be published in the body of the research. A copy of the key informants' signed consent forms are attached in Appendix II.

Six interviews were conducted by phone during the month of February 2018. Marie McCarthy was interviewed face-to-face as she is located in the same office as the researcher. A full list of key informants who participated in the research is outlined Table 4.2: Key Informants' Details, below:

TABLE 4-2 : KEY INFORMANTS' DETAILS

Key Informant's Name	Area of Specialisation	Role in the Industry
Bill Byrom	Patient reported outcomes and patient engagement through gamification	Vice President of Product Strategy and Innovation, CRF Health
Jeff Lee	Data collection for clinical trials & technical solutions for patient reported outcomes	CEO, mProve Health
Louis Smith	Health data science, analytics and machine learning	Associate Director Network Contracting OptumRx United Health Group
Marie McCarthy	Use of wearables in clinical trials	Senior Director, Product Innovation, ICON Plc
Richard Young	mHealth, secondary use of health data in clinical trials and data collection technologies	Vice President, Veeva Vault EDC, Veeva
Ross Rothmeier	mHealth, secondary use of health data in clinical trials and data collection technologies	Vice President, Technology Solutions and Innovation Labs, Medidata Solutions
Tigran Arzumanov	Using patient charts and EHR data for decision support in clinical trial feasibility	Head of Sales, Clinerion

4.6 WHY THESE METHODS WERE CHOSEN

Considerable quantitative research exists today on the current and predicted state of the collection of health data for use in phase II-III clinical trials. The decision to use qualitative research was based on the researcher's interest in gathering the opinions from clinical research industry leaders. The aim of using the direct interview approach was to understand how personal experience was shaping the acceptance of new technologies and practices, and to identify both the challenges and opportunities presented by these changes. The individuals were chosen based on their standing within this industry and the belief by the researcher that they could present an in-depth analysis of the issues faced within clinical research data collection (Marshall, 1996).

From a review of the literature many opportunities and obstacles are identified to improve the collection of health data for secondary use in clinical trials. Differing views about how data collection should be improved or changed for the purpose of clinical research are presented. No one solution is evident that will resolve the challenges faced, so the decision was made by the researcher to develop a deeper meaning of the themes presented in the literature review, by conducting the qualitative research interviews with key informants (Creswell, 2007). Access to these individuals was facilitated as part of the researcher's existing role in the clinical research industry.

4.7 Data Collection Approach

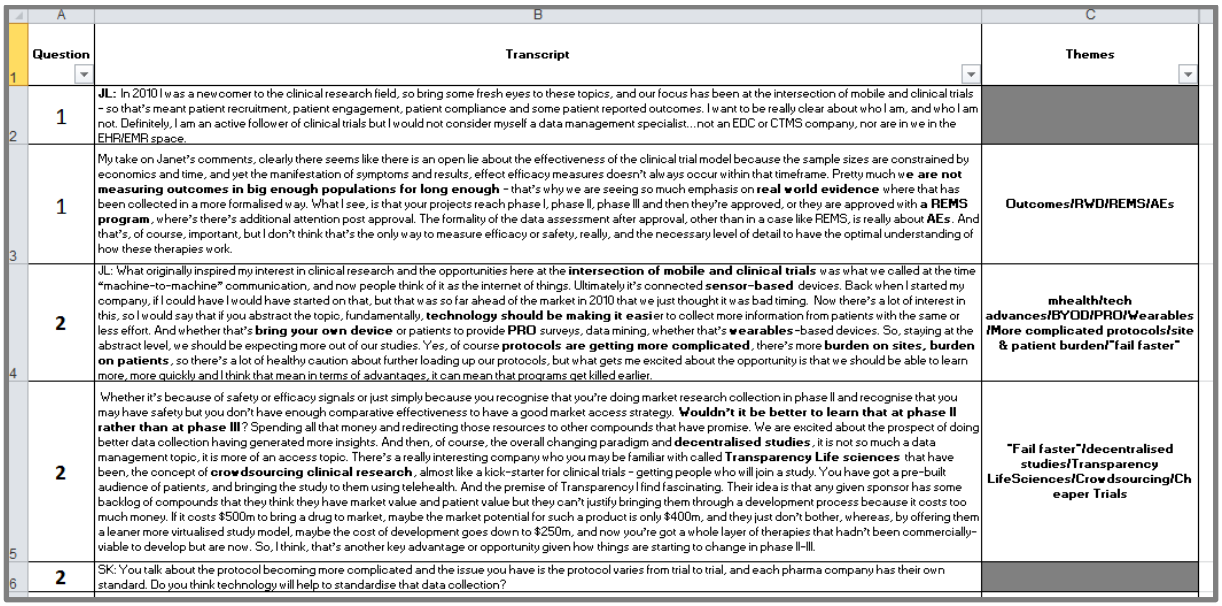
Once the participants signed the consent form, a meeting was set-up and the interviews conducted. A semi-structured interview format (Longhurst, 2016) was used to obtain the participant's personal view of the challenges and enablers in the collection of health data for secondary use in phase II-III clinical trials.

All conversations were recorded (with the exception of Marie McCarthy where the recording method deployed did not work), and all recordings were transcribed. For Maire McCarthy, the interviewer's notes were used. Participants were given an opportunity to review their final transcript and make any edits (no edits were made by the key informants).

4.8 Approach to Data Analysis

The answers provided by key informants have been used to form the basis of a data analysis section of this dissertation (chapter 5). Using an emergent strategy to check for trends in the key informant responses, the themes of each respondent's answers were extracted to form collective themes that are common to the majority of respondents. Unique themes common to one or a smaller number of participants (Waters, 2017) have been identified.

Using content analysis, transcripts were reviewed line-by-line, and themes were identified using codes created by the researcher. Where a theme was identified in answers from one key informant, this theme was sought out in the other key informants' answers. If a new theme emerged in subsequent transcripts, all transcripts were re-reviewed for this theme (Chamberlain, et al., 2015) using Microsoft Excel. Figure 4.1: provides a snapshot of how the data was collected and coded for Jeff Lee.



	A	B	C
1	Question	Transcript	Themes
2	1	**JL:** In 2010 I was a newcomer to the clinical research field, so bring some fresh eyes to these topics, and our focus has been at the intersection of mobile and clinical trials – so that's meant patient recruitment, patient engagement, patient compliance and some patient reported outcomes. I want to be really clear about who I am, and who I am not. Definitely, I am an active follower of clinical trials but I would not consider myself a data management specialist…not an EDC or CTMS company, nor are in we in the EHR/EMR space.	
3	1	My take on Janet's comments, clearly there seems like there is an open lie about the effectiveness of the clinical trial model because the sample sizes are constrained by economics and time, and yet the manifestation of symptoms and results, effect efficacy measures doesn't always occur within that timeframe. Pretty much **we are not measuring outcomes in big enough populations for long enough** – that's why we are seeing so much emphasis on **real world evidence** where that has been collected in a more formalised way. What I see, is that your projects reach phase I, phase II, phase III and then they're approved, or they are approved with **a REMS program**, where's there's additional attention post approval. The formality of the data assessment after approval, other than in a case like REMS, is really about **AEs**. And that's, of course, important, but I don't think that's the only way to measure efficacy or safety, really, and the necessary level of detail to have the optimal understanding of how these therapies work.	**Outcomes/RWD/REMS/AEs**
4	2	JL: What originally inspired my interest in clinical research and the opportunities here at the **intersection of mobile and clinical trials** was what we called at the time "machine-to-machine" communication, and now people think of it as the internet of things. Ultimately it's connected **sensor-based** devices. Back when I started my company, if I could have I would have started on that, but that was so far ahead of the market in 2010 that we just thought it was bad timing. Now there's a lot of interest in this, so I would say that if you abstract the topic, fundamentally, **technology should be making it easier** to collect more information from patients with the same or less effort. And whether that's **bring your own device** or patients to provide **PRO** surveys, data mining, whether that's **wearables**-based devices. So, staying at the abstract level, we should be expecting more out of our studies. Yes, of course **protocols are getting more complicated**, there's more **burden on sites, burden on patients**, so there's a lot of healthy caution about further loading up our protocols, but what gets me excited about the opportunity is that we should be able to learn more, more quickly and I think that mean in terms of advantages, it can mean that programs get killed earlier.	**mhealth/tech advances/BYOD/PRO/Wearables /More complicated protocols/site & patient burden/"fail faster"**
5	2	Whether it's because of safety or efficacy signals or just simply because you recognise that you're doing market research collection in phase II and recognise that you may have safety but you don't have enough comparative effectiveness to have a good market access strategy. **Wouldn't it be better to learn that at phase II rather than at phase III**? Spending all that money and redirecting those resources to other compounds that have promise. We are excited about the prospect of doing better data collection having generated more insights. And then, of course, the overall changing paradigm and **decentralised studies**, it is not so much a data management topic, it is more of an access topic. There's a really interesting company who you may be familiar with called **Transparency Life sciences** that have been, the concept of **crowdsourcing clinical research**, almost like a kick-starter for clinical trials – getting people who will join a study. You have got a pre-built audience of patients, and bringing the study to them using telehealth. And the premise of Transparency I find fascinating. Their idea is that any given sponsor has some backlog of compounds that they think they have market value and patient value but they can't justify bringing them through a development process because it costs too much money. If it costs $500m to bring a drug to market, maybe the market potential for such a product is only $400m, and they just don't bother, whereas, by offering them a leaner more virtualised study model, maybe the cost of development goes down to $250m, and now you're got a whole layer of therapies that hadn't been commercially-viable to develop but are now. So, I think, that's another key advantage or opportunity given how things are starting to change in phase II-III.	**"Fail faster"/decentralised studies/Transparency LifeSciences/Crowdsourcing/Cheaper Trials**
6	2	SK: You talk about the protocol becoming more complicated and the issue you have is the protocol varies from trial to trial, and each pharma company has their own standard. Do you think technology will help to standardise that data collection?	

FIGURE 4-1: SNAPSHOT OF HOW THE KEY INFORMANTS' DATA WAS COLLECTED & CODED

4.9 LIMITATIONS OF THE RESEARCH

There are several limitations to this research. Dudovskiy highlights that qualitative studies may lack credibility due to the presence of bias in how key informants are selected, and the key informants may be offering a biased view (Dudovskiy, 2017).

In terms of the research design, the questions used as part of the interview process and the codes use to identify themes were not peer reviewed (Spier, 2002). No software, with the exception of MS Excel, was used to identify codes or themes (Richards, n.d.), and the data analysis was based solely on the researcher's interpretation of the key informants' responses (Creswell, 2007).

All informants were selected at the same time and their selection was not as a result of any trends emerging in the data. "Snowball" sampling, where additional participants may have been included based on recommendations by the key informants, was not used due to time constraints (Richards, n.d.).

4.10 CONCLUSION

The use of key informants is a tried and tested method in qualitative research (Marshall, 1996), and this method was useful for the researcher to identify challenges and opportunities for the secondary use of health data in phase II-III trials.

Transcribing the interviews took a considerable amount of time but offered the researcher the opportunity to listen to the recordings many times and to be very familiar with each of the key informants' responses and points of view.

Although the key informants have different roles and areas of focus in the clinical research industry, several common themes emerged. These themes are presented in detail in Chapter 5.

5 QUALITATIVE RESEARCH RESULTS

5.1 INTRODUCTION

This chapter outlines the qualitative research findings from a series of interviews with key informants conducted in February 2018. Details of the key informants' individual responses, and the integrated research findings, are presented in the following sections.

5.2 INTERVIEW PROCESS

A total of seven key informants were interviewed. The key informants were selected based on their standing in the clinical research industry, and their areas of expertise ranging from patient reported outcomes to technology and mHealth. Further detail on their area of specialisation and their role in the industry is outlined in Chapter 4: Table 4.2.

Six key informants were interviewed by phone and one was interviewed face-to-face. Interviews were recorded, but in the case of Marie McCarthy's interview, the recording did not work. Transcripts of each recording were produced, and each transcript took approximately 3 hours to transcribe. For Marie McCarthy notes were taken. Both the notes and the transcripts are attached in Appendix III.

5.3 OVERVIEW OF RESPONSES

All key informants were asked to respond to Janet Woodcock's comments that the "clinical trial data collection process is broken" (Woodcock, 2017), and to identify the barriers and/or issues that exist today in how data is collected in phase II-III trials.

Of the seven key informants, two agreed, one disagreed, one partially disagreed, while the remaining three partially agreed with Woodcock's statement (Table 5.1: Key Informants' Question 1 - Overview of Responses).

TABLE 5-1 : KEY INFORMANTS' QUESTION 1 - OVERVIEW OF RESPONSES

Question 1		
Janet Woodcock, Director of the FDA's Centre for Drug Evaluation and Research recently remarked that the clinical trial data collection process is broken. Do you agree with this statement, and what do you see as the barriers or issues that exist today in how data is collected in phase II-III clinical trials?		
	Agrees	**Disagrees**
Tigran Arzumanov		√ (Partially)
Bill Byrom	√ (Partially)	
Jeff Lee	√	
Marie McCarthy	√ (Partially)	
Ross Rothmeier	√ (Partially)	
Louis Smith		√
Richard Young	√	

5.4 INDIVIDUAL RESPONSES TO QUESTION 1

All respondents sought to qualify and add context to their responses.

Tigran Arzumanov does not believe that the process is broken, but refers to it as being "*fractured*". He believes that the industry is conservative and slow to change, specifically with regard to source data verification where the original patient record must be compared to the data collected for the clinical trial.

Bill Byrom offers partial agreement to Janet Woodcock's statement and qualifies his response by stating that it is necessary to continue to collect "*controlled, randomised clinical trial data*".

Jeff Lee agrees with Janet Woodcock's statement and refers to an "*open lie*" in how data is collected today for clinical trials. He states that because of economics, the patient sample size is restricted to defined time windows and the impact of a treatment may not be measured effectively in large enough populations.

Marie McCarthy offers partial acceptance of Woodcock's statement and states that the design of a clinical trial is to "*support a hypothesis*", but if insufficient data is collected it may be impossible to prove a treatment's efficacy.

Ross Rothmeier partially agrees with Janet Woodcock's statement, and argues that greater efficiency, cost reduction and quality improvement are needed to improve the clinical trial data collection process.

Richard Young agrees unequivocally with Janet Woodcock's statement that the clinical trial data collection processed is broken, and refers to the outmoded data collection model that is in use today in clinical research.

Louis Smith does not believe that the clinical trial data collection process is broken, but identifies inefficiencies in how data is monitored today and suggests that onsite data monitoring should be targeted towards poor performing sites that may need additional oversight.

5.5 QUESTION 2 – INDIVIDUAL RESPONSES

Question 2 (Table 5.2: Key Informants' Question 2) was posed to all key informants to provide their opinion on what is changing in how health data is collected today for phase II-III clinical trials, and to outline the advantages and disadvantages of these changes.

TABLE 5-2 : KEY INFORMANTS' QUESTION 2

Question 2
In your opinion, what factors are changing how health data is collected in phase II-III clinical trials? What are the advantages of such changes? What are the disadvantages of such changes?

Tigran Arzumanov outlines that there are already links to EHR systems in near or real-time through HL7, which is allowing for new data collection methods for clinical research. He states that issues related to patient privacy have been solved by technology. To date, EHR data has only been used to assess trial feasibility rather than to replace or supplement CRF data.

Bill Byrom refers to current EDC technologies as being robust, but that they may be too complex to gather other data types such as the growing trend to gather more real world data across all clinical trial phases.

Despite the increased complexity of clinical trial protocols, Jeff Lee believes that the growth of machine-to-machine learning and the prevalence of sensor devices may change how data is collected for phase II-III clinical trials. The use of the patient's own device and the growth in the collection of patient reported outcomes data should provide more relevant data more quickly, and lead to more efficient (and faster) decision making on whether to continue with a trial.

Marie McCarthy believes that combining objective and subjective data is now possible through the use of mHealth. The use of apps in clinical trials has already made in-roads in how patients are recruited, and she cites one example with a Parkinson's disease app, where over 7,000 patients were enrolled in 6 hours, while the highest number of patients ever recruited using traditional methods was 1,700 patients. Gamification of patient apps may help to engage patients. While there are benefits in capturing the data remotely, there is a possibility of isolating patients and concerns related to data privacy, such as patient unblinding, remain.

Ross Rothmeier refers to the increase in data volumes and the challenges that new types of data from sensors, genomic testing, images, medical devices and wearable devices will pose. The old data management methods of verifying and validating clinical research data will not work on petabytes of data, and new methods of data verification will emerge through machine learning. Additionally, data that is "*esourced*" digitally cannot be verified against the original record, as is common practice today in clinical research, so the current data management methods will need to change. Ensuring that the data is used responsibly and that the patient's identity is not revealed will be challenging.

Louis Smith says the propensity of new apps, and the low cost to store data should assist in the more efficient collection of clinical research data. He proposes that data should be gathered with a longer term view in mind and not just for a particular trial or clinical study. He adds that apps and new technologies must fit into the clinician's workflow and offer some level of value.

Richard Young refers to the vast increases in the volume of data created from new mHealth technologies, and the need for tech companies to adapt to change. Traditionally, in clinical research, commercial agreements sought to lock-in sponsors and data management companies into long term technology

agreements that hinder the adoption of new technologies. The culture of balancing "*innovation and validation"* still prevails, while creating an open technology eco system for data sharing would bring benefit, Young believes.

5.6 QUESTION 3 – INDIVIDUAL RESPONSES

All key informants were asked to outline future innovations and to comment on the advantages and disadvantages of the collection of health data for use in phase II-III clinical trials. Question 3 is outlined in Table 5.3: Key Informants' Question 3

TABLE 5-3 : KEY INFORMANTS' QUESTION 3

Question 3
In the longer term, what innovations may impact how data is collected for phase II-III clinical trials? What advantages will these innovations bring? What issues may these innovations fail to address?

Tigran Arzumanov refers to how technology has overcome data currency issues by offering data in real-time directly from hospital EHR systems. "*Anonymised identification*" already removes unique identifiers from patient's data received from EHR systems and allows for this data to be used to assess clinical trial feasibility. In the future, this data could be used to complete the clinical trial CRF, and he acknowledges that the laxer data control laws in the US may make this process possible for hospitals there sooner.

Bill Byrom emphasises the conservative nature of the clinical research industry and believes it is hard to say how the data collection process will change as data will still need to be collected in a controlled way for pre-marketed drugs (in phase

II-III trials). New technologies will bring other challenges to the fore such as ensuring that the patient remains engaged.

Jeff Lee refers to new technologies such as the Apple Research Kit, which he says shows such promise by "*democratising data collection*", and gives an insight into future data collection practices. The use of sensors will pose data volume challenges and force a change in current data management practices, while EHR and EDC technologies will converge, in his view. Alleviating patient burden will be a big focus as patients will be asked to complete more electronic records directly.

Marie McCarthy believes that change is inevitable through the collection of "digitalised streams" of data which will allow for greater analytics and adaptive trial designs. She believes the use of wearables in trials is already becoming more common place. Trials will be longer and the requirement for real-world data will grow based on payers' demands for data and the need to control drug costs. Data privacy issues will continue to pose challenges.

Ross Rothmeier points to the growing use of wearables and sensors in clinical trials to collect patient reported outcomes data directly from patients. This data offers benefits as it provides a truer picture of the patient's overall health, and provides greater data granularity and patterns that are not visible in the "*two dimensions*" of EDC systems. Rothmeier believes that the data collected from sensors and devices can be used to develop synthetic control arms to simulate a patient population's reaction to a particular treatment, while relieving site burden by reducing the clinician's data collection workload.

Louis Smith sees the future bringing greater automation in how data is collected. Data volumes will grow, and the insight into the data will increase due to its granularity and the ability to correlate objective data collected in-clinic and subjective data from wearables. Issues such as patient burden and privacy will

prevail. The ability to identify patterns in the data (through machine learning) that may not be expected or may unblind a patient, will need to be overcome.

Richard Young believes that the patient will emerge as the force of change in how data is collected for clinical trials. The informed patient will demand more feedback that may be met through patient education programs and access to patient outcomes data. mHealth will continue to drive patient compliance, but it will be important to tailor solutions to patients and find the right balance for older patients or those with more critical conditions. The development of support programs where a clinician intervenes, if an event recorded on a device requires treatment or support, will "*build the patient's confidence*".

5.7 Overview of the Themes Identified in the Responses

Seven key themes have emerged in the answers provided by the key informants. These themes were identified by pasting all responses into an Excel table and labeling each paragraph of text in accordance with the main theme identified.

For each of the three questions posed, if a key informant's answer included a reference to a particular theme, this was counted as 1, and the total number of references to that theme was tracked. For example, the theme, "Issues with Current Clinical Trial Data Collection Processes" was referenced by 6 key informants. When this exercise was complete for each of the key informants' responses, the themes were grouped and synthesised into seven key themes. The number of key informants who referenced each theme is outlined in Table 5.4: Number of Times Key Informants Referenced Key Themes.

TABLE 5-4 : NUMBER OF TIMES KEY INFORMANTS REFERENCE KEY THEMES

Themes:	1. Issues with Current Clinical Trial Data Collection Processes	2. Clinical Research Regulation	3. New Clinical Research Data Types and Sources	4. Patient Engagement in Clinical Research	5. mHealth in Clinical Research - Challenges and Opportunities	6. Data Privacy & Clinical Research	7. The New Clinical Trial Landscape	Total
Tigran Arzumanov			1			1	1	3
Bill Byrom	1		1	1	1		1	5
Jeff Lee	1		1	1	1		1	5
Marie McCarthy	1		1	1	1	1	1	6
Ross Rothmeier	1	1			1	1	1	5
Louis Smith	1	1	1	1		1	1	6
Richard Young	1	1	1	1	1		1	6
Total Key Informants (n = 7)	6	3	6	5	5	4	7	36

The seven themes were then categorised into sub-themes, and a count performed to assess how many times a key informant referenced each sub-theme. In the case of theme 1, Issues with Current Clinical Trial Data Collection Processes, there are four sub-themes: 1a: Current Data Collection Practices in Clinical research was referenced by 3 key informants; 1b. Current Clinical

Research Data Collections Systems was referenced by 3 key informants (as so on). The themes and sub-themes and the number of references are outlined in Table 5.5: Key Informants' Themes & Sub-themes.

TABLE 5-5 : KEY INFORMANTS' THEMES & SUB-THEMES

#	Key Themes (Key Informants' Responses)		# Key Informants Themes
1	**Issues with Current Clinical Trial Data Collection Processes**		
	a	Current Data Collection Practices in Clinical Research	3
	b	Current Clinical Research Data Collection Systems	3
	c	Not Using Data in the Right Way	4
	d	Insufficient Data	2
2	**Clinical Research Regulation**		
	a	Clinical Research Regulation is Stifling Change	1
	b	Managing Risk in Clinical Research Regulation	2
	c	Difficulty in Navigating Clinical Research Regulation	1
3	**New Clinical Research Data Types and Sources**		
	a	Real World Data	1
	b	EHR Data	3
	c	Patient Reported Outcomes Data	3
	d	Apps, Sensors and Wearables	3
4	**Patient Engagement in Clinical Research**		
	a	Patient Burden in Clinical Research	4
	b	New Clinical Research Technologies to Engage the Patient	3
5	**mHealth in Clinical Research - Challenges and Opportunities**		
	a	Remote Data Collection in Clinical Research	3
	b	New Clinical Trial Data and Endpoints	3
	c	Large Datasets and Increased Data Volumes in Clinical Research	2
6	**Data Privacy & Clinical Research**		
	a	New Technologies and the Risk of Patient Unblinding	4
	b	Data Sharing for Clinical Research Purposes	1
	c	Responsible Use of Data in Clinical Research	1
7	**The New Clinical Trial Landscape**		
	a	The Crossover from Healthcare to Clinical Research	2
	b	EHR Data Standards	2
	c	Changing Processes in Clinical Research	2
	d	Personalised Medicine and Pragmatic Trials	1
	e	The Voice of the Patient in Clinical Research	1
	f	Better Data Analytics for Clinical Research	3
	g	New Trial Designs	2

These themes are presented in further detail in the sections below.

5.8 ISSUES with CURRENT CLINICAL RESEARCH DATA COLLECTION PROCESSES

Several key informants present issues with the current data collection processes and systems used in clinical research today, and opposing views are presented on whether these practices will change.

5.8.1 Current Data Collection Practices in Clinical Research

Richard Young believes that current data collection practices must change. While Ross Rothmeier, on the one hand, says the clinical research must be willing to change, he emphasises that the current practice of collecting data to prove a hypothesis using rigorous and controlled methods must continue.

Bill Byrom lauds the current randomised controlled trial approach. He believes that this approach does not need to change as it is the most effective way of controlling how patients are enrolled and randomised. Managing the data in a rigorous and controlled way, as we do today, is the most effective way to assess the patient's safety profile, Byrom believes.

5.8.2 Current Clinical Research Data Collection Systems

Jeff Lee believes that "*disruption*" is inevitable as current data collection systems are designed to receive data in batches while new technology provides a continuous stream of data in large volumes. On the opposing side, Bill Byrom suggests that current clinical trial data collection systems have extensive functionality and will remain, although their configuration may be unsuitable to collect real world data.

Ross Rothmeier believes that there is a growing recognition that solely measuring the impact of treatment on a patient in-clinic is insufficient, and he suggests that current EDC systems will need to evolve to accept other data sources data for phase II-III clinical trials.

5.8.3 Not Using Data in the Right Way

Richard Young states that the clinical research industry is "*not using data in the right way*", and that traditional data collection methods are leading to futility. He identifies several challenges under the headings of "volume, variety, velocity, veracity and value" (Raghupathi & Raghupathi, 2014), where data volumes are growing exponentially; the type of data collected is changing (variety); there is an expectation that the data is available now (velocity), and the data must be verified, but perhaps not to the extent that the data is verified today - which all lead to value in terms of collecting the right data that will provide the best trial endpoint.

Citing the 6-minute stress test, Young states this is the "*least useful test in the world*", as it only tests the patient's fitness at a point in time. If the patient is having a bad day, their result (and the data) might be poor, but if they have rested sufficiently before the test, it might produce a false positive result, and the data collected would not provide a true representation of the patient's overall condition.

Louis Smith states that data captured today is not done so with the longer term value of the data in mind. In particular, he refers to operational data that is used to assess a hospital site for suitability for a trial. This data is used in the decision-making process of how to proceed to the next step in a specific trial, but the data is not re-used for the next trial nor is this data available for retrospective analysis to identify trends across trials or hospital sites. Additionally, collecting data such as how far a patient needs to travel to a hospital site may not be useful for trial management on an individual basis, but if this data was available across several trials and many patients, it may show some meaningful insights as to why a patient leaves a trial early or is non-compliant.

Citing an example in the case of pain studies, Marie McCarthy refers to the FDA's statement that certain analgesic compounds may not have been approved, as the methodology, including the data collection process, used in identifying their effectiveness was poorly designed.

Ross Rothmeier suggests that if thousands of patients are recruited for a clinical trial and the drug proves to be effective, there is nothing to say that when that drug is released onto the market and used by millions of patients that there will not be issues.

5.8.4 Insufficient Data

Both Marie McCarthy and Jeff Lee believe that we are collecting insufficient data. Marie McCarthy puts forward the argument that the patient is typically enrolled in a trial for 12-18 months and spends an average of 9,000 hours on a trial, but only 50 hours' worth of data is collected.

Jeff Lee states that often because of cost, the patient sample size is restricted, and the data collection timeframe is kept to a minimum in clinical research trials. So, the efficacy of a treatment may not be measured over long enough timescales or in large enough populations (in large enough datasets).

5.9 CLINICAL RESEARCH REGULATION

Three key informants refer to regulation in the clinical research industry. Ross Rothmeier highlights how it is often difficult to navigate or interpret the regulators' requirements and this stifles change, while Louis Smith believes that the greater risk lies with the regulator and this may influence their decisions. Richard Young is often disappointed with how the regulatory industry reacts to suggestions to engage.

5.9.1 Clinical Research Regulation is Stifling Change

Ross Rothmeier states that anyone working with clinical data cannot justify changing the current data collection processes. On one hand, the regulators are supportive by publishing new guidance documents, but simultaneously they are still deemed to be "*a bit of an adversary*". The risk of changing established practices is deemed to be too great for many, as the threat of receiving a warning letter from the authorities always looms if any side steps are taken in the data collection process. This landscape is stifling change.

5.9.2 Managing Risk in Clinical Research Regulation

The regulators are hesitant to speed up drug trials for fear that this may result in shortcuts in how a drug is approved. Louis Smith argues that if the regulator reduces clinical research cycle times, the drug companies may benefit, but the regulator carries all of the risk if the drug causes significant side effects for patients.

In the last few years, the regulators have published a series of white papers on new trends such as mHealth. Richard Young sees this as an opportunity to engage with the regulators, but he is disappointed that the regulator's responses

are often isolationist rather than in the spirit of collaboration, and one is left feeling that a change in process is too high a risk for the regulator.

5.9.3 Difficultly in Navigating Clinical Research Regulation

In many cases it is not clear what the regulators want or expect. Rothmeier highlights the disparity in the regulators' area of focus in different regions. He refers to his experience of being inspected by regulators from five different countries and states that each experience was very different with different areas of focus. He mentions the adage common in clinical research circles, "…*if you ask four regulators a question, you will get five different opinions…*", but further states that the regulators are not obliged to be consistent.

5.10 NEW CLINICAL RESEARCH DATA TYPES AND SOURCES

Six key informants refer to new data types and sources as impacting how data is collected for phase II-III clinical trials. Real world data, such as EHR data and patient reported outcomes and data from wearables and sensors are presented as supplementary data sources that could complement existing clinical research data collection methods and deepen the understanding of how the drugs are used in practice.

5.10.1 Real World Data

Currently real-world data is collected in phase IV clinical research, when the drug has already been approved for market. Bill Byrom refers to the 21st Century Act that mandates the need for more real-world data in clinical trials in the US. He believes that the current practice of collecting real world data when a drug has already been approved and is available on the market will continue, while for phase II-III studies, current trials designs will continue as data must be collected in a controlled way.

5.10.2 EHR Data

There are opposing views on how hospitals' EHR systems will shape clinical trials. Bill Byrom recognises that there are some successes with capturing EHR for clinical research, but believes that this is restricted to particular hospitals using a single technology.

Tigran Arzumanov sees EHR data as the future, and already is working on providing a continuous stream of EHR data from a site in France in near real-time. He states that the concept of recording patient data in two systems has "outlived its usefulness". Processes such as source data verification of the

clinical trial data record against the original hospital record, highlights a "*fractured*" process.

Louis Smith argues that limiting trial enrolment to recruiting sites that are willing to submit data electronically through EHR may drive efficiency.

5.10.3 Patient Reported Outcomes Data

Bill Byrom sees the advantages that the collection of more patient reported outcomes will bring. He foresees that patients will use their own mobile devices to download a questionnaire or diary app, and this data will automatically populate the clinical research data management system.

Richard Young states that it is expected that 80% of trials will use electronic patient reported outcomes, and this data will come in different formats, at different frequencies and will require different levels of processing and verification. There is a growing expectation that the data will be available at a greater velocity in real-time.

Jeff Lee believes that use of the patient's own device and the growth in the collection of patient reported outcomes data will lead to more relevant data more quickly.

5.10.4 Apps, Sensors and Wearables

Jeff Lee predicts that future trials will collect objective and subjective data using both active and passive measures, such as sensors, and, as a result, data collection will be much faster and cost effective. The collection of passive data through wearables, sensors, and technologies such as the Apple Research Kit will have a role to play in the discovery of new trial endpoints, but these technologies must be developed using new methods and principles in what he calls the application of ALCOA (accurate, attributable, legible, contemporaneous,

original) principles (Barkow, 2016). Referring to the six-minute walk test, Lee believes that wearable devices to continuously assess patient mobility and related factors, will replace this test.

Marie McCarthy believes that a continuous stream of data from wearables would provide a greater insight into a patient's condition, especially for conditions such as Parkinson's, where symptoms tend to be episodic.

Louis Smith believes that apps will lead to more efficient data collection directly from patients.

5.11 PATIENT ENGAGEMENT IN CLINICAL RESEARCH

Several key informants recognise that alleviating patient burden and providing solutions to engage the patient in the clinical trial process are necessary to ensure the successful deployment of mHealth technologies in clinical research.

5.11.1 Patient Burden in Clinical Research

Marie McCarthy says that the patient must not be over-burdened with a myriad of devices and sensors. She suggests that the pharmaceutical industry is losing sight of the burden to the patient that mHealth poses, but burden can be alleviated through gamification to encourage patients to stay on a trial.

The burden of having to take time to master the use of wearables and devices and to record and enter data remotely, is what Jeff Lee refers to as an "*imbalanced proposition*" with little benefit to the patient. Lee refers to the ePRO Consortium's commissioned study to review what research had been conducted on the patient burden in the clinical research data collection process. The study came up empty. There is an onus on the clinical research industry to make the proposition more appealing to patients and to offer features such as helping them with the scheduling of the doctor's visits, adding reminders of what items to bring with them, providing details of what to expect at each visit and providing links to reference materials.

Bill Byrom believes that engaging the patient is necessary through formal feedback, but patient burden may still remain. He refers to the collection of data directly from older patients or those with more debilitating illnesses, and suggests that this process may be too onerous for them.

Louis Smith points out that technology must be useful and must not place additional burden on the patient or the clinician.

Referring to paper diaries, Jeff Lee recalls the "*big joke*" of patients completing paper diaries in the car park before their doctor's visit; and at least mHealth is able to capture the time point for when the data is entered or picked up by the sensor; and passive collection of data should lessen the burden on the patient.

5.11.2 New Clinical Research Technologies to Engage the Patient

Bill Byrom believes that we need to consider how we can make apps useful to the patient, but without compromising the results of a trial. On the one hand we want to offer the patient assistance in managing their condition, but the app cannot "*become part of the intervention itself*". He gives an example where a patient is asked to track their steps using a smartphone for a drug that may be designed to improve their levels of activity. There is a risk the patient may use the activity tracker to set themselves targets to improve activity rates, and thus change the purpose for which the app was issued to them. This would make it impossible to separate the effects of the drug versus the influence of the app on activity levels.

Byrom points out that apps have a limited appeal. In the beginning users are excited to download and use them but over time, users disengage, and the drop-out rates are high. Unless the patients are engaged, and the app is providing value to the patient, it will not work.

Jeff Lee states that alleviating patient burden will be a factor in determining the successful deployment of new data collection technologies. Citing the example of the Apple Research Kit where patients can enter their own health records, Lee states that the uptake has been low. Thousands of people downloaded the app, but usage rates have diminished greatly. In contradiction, he also sees the value in the Research Kit, and refers to it as the "*first generation*" of technology that will shape data collection in clinical research.

Lee emphasises the need to ensure "*balance and engagement*" when designing technologies to engage the patient. We need to broaden the availability of features that are of value to the patient. Techniques like gamification, and providing education and useful instructions will engage the patient. He refers to organisations such as Patients-Like-Me (PatientsLikeMe, 2018) who are aiming to provide the patient with feedback and a network of contacts with similar conditions. He believes this concept is not prevalent in clinical trials, but needs to change to make the trial process more "*meaningful for the patient*".

Richard Young believes that the pharmaceutical industry should be viewed as a patient ally by engaging patients on trials through education programs and the tracking of outcomes beyond the clinical trial. He believes this level of engagement will drive patient compliance in trials.

5.12 mHEALTH IN CLINICAL RESEARCH – CHALLENGES & OPPORTUNITY

mHealth technologies pose opportunities as well as challenges. The use of sensors and wearables for the collection of data remotely directly from patients offers efficiency and a greater level of data accuracy. Regulation may hinder the widespread deployment of mHealth technologies in clinical research, but the promise of new endpoints presents opportunities to enhance clinical trial design. Larger datasets from mHealth technologies will require a change to current data collection processes, while issues related to ensuring patient anonymity remain.

5.12.1 Remote Data Collection in Clinical Research

Bill Byrom does not see a future where all data will be collected remotely for phase II and III trials. Remote data collection will be restricted by the regulations that require patients to travel to a hospital site to consent to their participation in a trial or for them to receive trial medications.

Referring to Pfizer's virtual trial (Pfizer, 2011) which created one super-site to enroll all patients in the US, Byrom sees what he calls a "hub and spoke" model as more likely. In this model there would be one super site managed by an experienced investigator who will oversee the safety aspects of a trial, and less experienced investigators who will manage more mundane data collection activities (as opposed to a fully decentralised model).

Jeff Lee cites an example of what he calls a "*kick-starter for clinical trials*" using crowd-sourcing (Transparency Life Sciences, 2018), to build a patient audience online. Pharmaceutical companies may use this data to assess their backlog of compounds and decide if there is sufficient interest in therapies for specific indications. This "*leaner more virtualised study model"* may reduce the costs to develop a drug and make it viable to develop a therapy that previously was thought to be too costly.

mHealth provides many opportunities for pragmatic trial models. Richard Young provides an example of a diabetes patient who could be treated at home, where their activity would be tracked on their phone and their laboratory result would be available to them immediately. This data could be reviewed remotely by a clinician who could liaise with the patient by phone to review or adjust their treatment plan. Young suggests a hybrid data collection model will evolve, where a remote data collection will work for some patients, but perhaps not for others.

5.12.2 New Clinical Trial Data and Endpoints

Jeff Lee points out that wearables and sensors are not yet used in a standard way. New measurements and endpoints will need to be developed using these technologies which will take considerable effort, although there is the possibility for great opportunity using mHealth technologies.

Richard Young refers to the six-minute stress test in patients with cardiovascular disease. Results have shown that some patients build themselves up to perform well in the test at the doctor's site, while others (after the test) will spend days recovering in bed, or others will perform badly at the test, which results in inconsistency in the data collected. Using mHealth to continuously track the patient's activity before and after the test offers a solution and provides qualitative data to show a pattern in the patient's activity and thus provide a more accurate overview of their overall condition. Marie McCarthy agrees that endpoints should be linked to clinical outcomes.

5.12.3 Large Datasets and Increased Data Volumes in Clinical Research

Ross Rothmeier highlights the growth in data volumes using new technologies, while Jeff Lee says that the data from sensors is "*so granular*" that the current data cleaning processes will not apply, and the emphasis will be on identifying data patterns across large datasets.

Citing Tufts (Getz, 2018), Richard Young states that approximately 1 million data points are collected in the average phase II-III clinical trial today. Young recently conducted a pilot study using an Actigraph to gather sleep and step activity from 12 patients over a 9-month period, and this resulted in a dataset of over 12 billion patient records. He states that "*no existing technology that can handle this volume of data*".

5.13 DATA PRIVACY & CLINICAL RESEARCH

Four key informants outline data privacy challenges with the sharing of patient data. They refer to hospitals' reluctance to share EHR data, coupled with the need for the data to be used ethically and anonymously for clinical research.

5.13.1 New Technologies and the Risk of Patient Unblinding

Ross Rothmeier, Marie McCarthy and Louis Smith point out potential issues with technologies such as artificial intelligence, machine learning and sensors that collect data at a granular level and that may unintentionally identify the patient based on unique data points relevant to that particular patient. While Tigran Arzumanov states that issues related to patient privacy have been solved by technology.

5.13.2 Data Sharing for Clinical Research Purposes

Tigran Arzumanov states that certain institutions are still reluctant to share data. Both hospitals and pharmaceutical companies do not want to breach the patient's privacy, but this is at the expense of research. Some parties are very conservative and slow to change or share data, while others are more progressive and willing to take risks. Today hospitals are providing data for clinical research to assist with patient recruitment but, in his experience, there is an unwillingness to provide EHR health records. He believes that it is only a matter of time before this practice changes.

5.13.3 Responsible Use of Data in Clinical Research

Ross Rothmeier highlights the need to de-identify data, and to give patients an assurance that data will be used responsibility. He highlights his uncertainty about the patient sharing genomic data as it may be used for other purposes such as to judge insurance risk. Louis Smith points out that data privacy will continue to be a concern for patients.

5.14 THE NEW CLINICAL TRIAL LANDSCAPE

All key informants point to several key themes emerging in the new clinical trial landscape including the crossover between healthcare and clinical research, the growth of personalised or precision medicine, and the drive to make the patient's visit to the hospital more valuable to the patient.

5.14.1 The Crossover from Healthcare to Clinical Research

Marie McCarthy refers to the crossover from clinical research to healthcare and how these two worlds are colliding in terms of data collection.

Referring to the 21st Century Cures and Affordable Cares Acts which mandated and subsidised doctors' surgeries to use electronic data collection methods for patient records, Jeff Lee, states that "meaningful use" (ONC, 2017) will move to the next stage, where data must be collected electronically once for both primary care and research purposes. This will blur the lines between EDC and EHR systems, and Lee expects to see these changes impact data collection for clinical research over the next 5 to 10 years.

5.14.2 EHR Data Standards

Tigran Arzumanov refers to the requirement for data standards in EHR systems. He highlights that hospital data collection systems are different to systems used in clinical research. They use different coding conventions which he views as a greater challenge than the technology itself, as it not always easy to map one coding dictionary to another. Bill Byrom argues that EHR systems will not become the source of clinical trial data, as EHR data is considered to be insufficient and of poor quality.

5.14.3 Changing Processes in Clinical Research

Ross Rothmeier believes that the clinical trial data cleaning processes in use today will not work with future technology. The current method of programming edit checks to verify the parameters of a particular variable or across multiple variables will not work across 335 petabytes of data. Rothmeier further highlights as data becomes more digitised, it will be impossible to review and verify this data, as the data in itself is already verified as it will be captured directly from a wearable or other device. He says, "*imagine the value that we can derive if we start looking at data more effectively*". The potential change in clinical research data types is placing the emphasis on the development of programmable solutions using machine learning and artificial intelligence.

mHealth will challenge current data processing practices, as the large data volumes will make them defunct. Richard Young suggests that the concept of "*good enough could be good enough*" needs to be applied to clinical trial data, and there is a need to move away from practices such as source data verification.

5.14.4 Personalised Medicine and Pragmatic Trials

Richard Young believes that personalised or precision medicine will evolve to prove that a particular drug is an effective treatment. Programs such as the 1,000 genomes project (IGSR, 2018) will drive change in how reimbursement levels are decided, "*not just for the patient in front of you, but for the patient yet to come*."

Young recognises that a patient must be 100% compliant with the treatment process for it to work, and he states that patients, on average are only 65% compliant. The risk of reducing hospital payments may drive a change in how trials are conducted and Young predicts a rise in the number of pragmatic trials,

and even refers to the "*ultimate pragmatic trial*" where patients are provided with the study drug and their data is tracked remotely.

5.14.5 The Voice of the Patient in Clinical Research

Richard Young refers to "*the patient's demand for knowledge*" and their frustration that trials today do not necessarily offer them a solution as the drug may not prove to be effective. He believes that future trials will be tailored to the individual patient, specifically in rare diseases, as the patient's risk profile is very different. Young wants to "*give them (the patient) their visit back*", by using technology effectively and removing the time spent on transcribing or entering data. This will allow for the doctor and patient to spend more time focusing on the patient's condition and less on administrative tasks.

5.14.6 Better Data Analytics for Clinical Research

Referring to the "*network effect*", Louis Smith is interested in exploring the connection between a doctor's social network and their performance on a clinical trial. Often investigators may be dismissed because they lack clinical trial experience, but if the doctor works in a hospital where there is a network of experienced investigators with significant clinical trial experience, they could rely on this network to guide them through the clinical trial process. So, the inexperienced doctor with a network would seem like a better candidate than a doctor with no experience and no support network. Today, this data exists in Citeline (PI, 2018), but it is not aggregated nor compared during the trial feasibility process.

Ross Rothmeier highlights that historical data is useful for predictive analysis to build a more targeted approach to data capture, in what he calls "*collecting data, as expected"*.

Jeff Lee believes that the growth of machine-to-machine learning will change data collection practices over time.

5.14.7 New Trial Designs

Ross Rothmeier proposes that the development of synthesised models using genomics and historical data could be used to assess what treatments are likely to be more effective and have a higher probability of success. Payers and insurance companies will welcome new models that provide a confidence level of success as they will be based on historical evidence and genomic data that will show that the drug works. This may result in the need for data from fewer patients to assess a compound's efficacy, and the historical data can be re-used across multiple trials.

Bill Byrom suggests that phase II trials will continue to focus on assessing efficacy and dosage rates using the data collection methods we use today. In phase II-III studies, there will still be a need for randomised controlled trials but there may be a shift towards less rigorous studies with an emphasis on gathering more real-world data on the drug's efficacy. Catalysts such as the 21st Century Cures Act will drive the collection of increased volumes of supplementary data such as real-world data. Byrom states further that if data collection apps are deployed to thousands of patients, and the fall off rate is high, even if a few thousand patients enter sufficient data in the app, perhaps this data will be sufficient to build patient profiles, even if the dropout rate is high.

5.15 CONCLUSION

Seven core themes have emerged from the key informants interviews. These themes include the need to address the challenges with current clinical trial practices, systems and data; challenges with compliance with clinical regulation; new data sources and types; the increased role of the patient in clinical research; the opportunities and challenges posed by mHealth; and the shifting clinical research landscape. These themes are discussed, side by side with the results of the literature review, in Chapter 6.

6 DISCUSSION

6.1 INTRODUCTION

The purpose of this study was to explore the challenges and enablers in the collection of health data for use in Phase II-III clinical trials by identifying process gaps and new developments through a review of the current literature and through interviews with key informants. Several themes have emerged and have been classified in Table 6.1: Literature Review and Key Informants' Themes

TABLE 6-1 : LITERATURE REVIEW AND KEY INFORMANTS' THEMES

Theme By Type:	Definition
Common themes:	Common themes emerged in the literature and in the key informants' interview process
Interview-only themes:	Themes that emerged in the key informants' responses but not in the literature review

6.2 COMMON THEMES

Several common themes have emerged through the review of the literature and the key informants' interview process, which are outlined in Table 6.2: Common Themes.

TABLE 6-2 : COMMON THEMES

#	Theme	Literature	Interviews
1	Issues with the Current Clinical Trial Data Collection Processes	√	√
2	Current Data Collection Systems	√	√
3	Not Using Data in the Right Way/Current Clinical Trial Data May Not Be Optimal		
4	Insufficient Data	√	√
5	Clinical Research Regulation	√	√
6	New Clinical Research Data Types and Sources	√	√
7	Large Datasets and Increased Data Volumes in Clinical Research	√	√
8	Data Privacy/ Responsible Use of Data in Clinical Research	√	√
9	Data Sharing for Clinical Research Purposes/Growing Recognition of the Value of the Secondary Use/Increased Data Availability	√	√
10	The Crossover from Healthcare to Clinical Research	√	√
11	EHR Data Standards	√	√
12	The Voice of the Patient in Clinical Research	√	√
13	Better Data Analytics for Clinical Research/New Technologies	√	√

6.2.1 Issues with the Current Clinical Trial Data Collection Processes

Current data collection practices in clinical research are facing many challenges. Danciu et al refer to data quality, data standards, finance, and availability of data as the main barriers for the secondary use of data in clinical research (Danciu, et al., 2014). There are inefficiencies in how data is collected as well as complex regulations and technologies that are not meeting data collection needs (National Academy of Sciences, 2009).

Of the key informants, Young believes that current data collection practices must change in their entirety, while, on the opposing side, Rothmeier and Byrom recognise that the current practice of collecting data to prove a hypothesis using rigorous and controlled methods must continue.

6.2.2 Current Data Collection Systems

The literature refers to the myriad of different EDC systems in use today for the collection of health data for secondary use in clinical research. Both Shah et al (Shah, et al., 2010), and Lu & Su (Lu & Su, 2010) refer to the multitude of EDC systems in use and that lack of interoperability across systems. These systems are expensive to deploy, but offer extensive functionality such as checks for data quality and access control (Shah, et al., 2010). While Krishnankutty outlines that there is no common CRF standard used across organisations or systems (Krishnankutty, et al., 2012) nor is there interoperability across CRF design standards (Richesson & Nadkarni, 2011). There are various standards in place for clinical research data such as the National Cancer Institute (NCI), (caDSR, CDISC; and (AHRQ), but these standards are incompatible and neglect to include important contextual information (Richesson & Nadkarni, 2011). The HITSP, FHIM, and CDISC are focused on standardising clinical research data,

but they do not provide interoperability across multiple domains (Sinaci & Erturkmen, 2013).

The key informants do not mention the cost or the extensive offering of EDC systems on the market, but accept that these systems must change to accommodate new data sources and formats (Lee and Rothmeier). Only Byrom mentions the extensive functionality that EDC systems offer, and he believes they will continue to be used.

6.2.3 Not Using Data in the Right Way/Current Clinical Trial Data May Not be Optimal

In the literature, Moore et all points out that there is a concern that the data collected today in clinical trials today is not “generalisable” to larger populations (Moore, et al., 2000).

Of the key informants, Ross Rothmeier supports the concern related to generalisablility of trial data. McCarthy adds that data clinical trial data collection methodologies are poorly designed. Young identifies several challenges under the headings of “volume, variety, velocity, veracity and value” (Raghupathi & Raghupathi, 2014), while Smith states that data captured today is not re-used for retrospective analysis to identify trends across trials or hospital sites.

6.2.4 Insufficient Data

In the literature, Frieden states that insufficient data is gathered to ensure the long term safety and effectiveness of a drug (Frieden, 2017),

Two key informants (McCarthy and Lee) believe that we are collecting insufficient data in clinical trials. On average, only 50 hours of patient data is gathered on a typical trial over a period of 12-18 months (McCarthy), and restrictions on study duration and the sample size mean that insufficient data is collected in clinical trials (Lee).

6.2.5 Clinical Research Regulation

The literature emphasises the increasing demands from the regulators in how to harness and display clinical data (ICH, 2015). While the key informants emphasise that regulation is stifling change and difficult to navigate (Rothmeier), There is a risk both for the regulator and clinical researcher by moving away from tried and tested practices (Young and Smith).

6.2.6 New Clinical Research Data Types and Sources

The literature outlines several initiatives that are underway to increase the use of real world data, such as EUPATI's and the OECD's drive to increase the use of electronic patient patient-reported outcomes (ePRO). While the value of these initiatives is recognised, Cowie et al points out that EHR data is currently used, mainly, for the pre-screening of patients and may not contain the complete patient record (Cowie, et al., 2017). Both the NIH and PCORI are assessing if additional data collection modules can be added to the data that already exists in the EHR to make the dataset more suitable for clinical research.

Referring to the 21st Century Act that mandates the need for more real-world data in clinical trials, Byrom recognises that there are some successes with capturing EHR for clinical research. He adds that for phase II-III studies, data will continue to be collected in a controlled way using systems prescribed for clinical research. Tigran Arzumanov sees EHR data as the future, and the concept of recording patient data in two systems is redundant, in his view.

The faster availability of ePRO data, and wearables, sensors and technologies such as the Apple Research Kit will have a role to play in the discovery of new trial endpoints, and may replace certain assessments such as the six-minute walk test (Lee and Young). A continuous stream of data from wearables would provide a greater insight into a patient's condition, especially for conditions such

as Parkinson's, where symptoms tend to be episodic (McCarthy), and apps and the use of EHR will lead to more efficient data collection directly from patients (Smith).

6.2.7 Large Datasets and Increased Data Volumes in Clinical Research

The literature refers to the increase in the volume of clinical research data (Bhadani & Jothimani, 2016), and that it is presenting technical and workload challenges such as the timelines to lock a clinical trial database have increased by approximately 10% (Tufts, 2018).

The key informants (Rothmeier and Young) highlight the growth in data volumes, and agree that current processes and systems will be replaced or need to adapt to process new data sources from wearables and sensors data, while recognising the need to identify data patterns across large datasets (Lee).

6.2.8 Data Privacy

The literature refers to the General Data Protection Regulation (GDPR) which stipulates that personal data must not be used without consent and restricts the movement of data without pre-defined levels of protection (although anonymous data is excluded). The HIPAA act (in the US) controls data privacy through the suggested removal of defined variables that would identify the patient (Maldoff, 2016).

It is recommended that the use of distributed analyses where data remains within the hospital may combat privacy issues, but if EHR systems are used for clinical research, the patient must provide informed consent and the data must be accessed by regulatory agencies for auditing and verification purposes. (Cowie, et al., 2017)

The key informants have concerns related to data privacy and point out potential issues with technologies such as artificial intelligence, machine learning and sensors that may unintentionally identify the patient based on unique data points (Rothmeier, McCarthy and Smith). Rothmeier highlights the need to de-identify data, and to use patient data responsibility, while Smith points out that data privacy will continue to be a concern for patients.

6.2.9 Data Sharing for Clinical Research Purposes/Growing Recognition of the Value of the Secondary Use/Increased Data Availability add to table

The use of EHR as the primary data source for clinical research will accelerate clinical research (Embi & Payne, 2014). There is an increase in the number of public and private initiatives, such as EOSC (EOSC, 2018) and the 1m Genomes Project (Regalado, 2015) to explore the better use of health data for primary and secondary use (Auffray, et al., 2016), and to drive better availability, sharing and linking of health data (Danciu, et al., 2014). One key informant highlights that there is still a reluctance to share health data for fear of breaching patient privacy, but this is at the expense of research, but it is only a matter of time before this practice changes (Arzumanov).

6.2.10 The Crossover from Healthcare to Clinical Research

Both the literature and the key informants outline the crossover from healthcare to clinical research. The literature highlights the growing recognition of the value of secondary use of health data which has been driven by the Hitech Act in the US (Danciu, et al., 2014) and the Declaration of Helsinki (WMA, 1964; amd 1975, 1983 & 1989). Healthcare and clinical research are converging, and data will be collected once for both primary care and secondary use (McCarthy and Lee).

6.2.11 EHR Data Standards

In the literature, many authors converge on the lack of standards and interoperability across health data standards and systems. Johnson et al recognizes that health systems and data should be interoperable (Johnson, et al., 2016). While De Moor et al emphasise the lack of interoperability across health and clinical research data (De Moor, et al., 2015) due to the various data standards in use by hospitals (Richesson & Nadkarni, 2011).

Projects to extract EHR data have had limited success as there are too many EHR systems and a lack of standardisation (Siemens, 2016), so EHR data has yet to be used for key trial decisions (Getz, 2017).

Cowie et al proposes focusing on a small number of variables to develop common health and clinical research standards (Cowie, et al., 2017), while Richesson & Nadkami recommend that a common data definitions should be in place across EHR and EDC systems (Richesson & Nadkarni, 2011)

Two key informants refer to issues with health data standards. Tigran Arzumanov highlights that hospital data collection systems are different to systems used in clinical research, which he views as a greater challenge than the technology itself, while Bill Byrom considers that EHR data to be insufficient and of poor quality.

6.2.12 The Voice of the Patient in Clinical Research

The voice of the patient as an educated and informed stakeholder is an important step in developing hospital data for use in clinical research (Cowie, et al., 2017). Change will be driven by the patient as a consumer, who is willing to participate in research, and the growth of large scale data enterprise initiatives in large provider institutions (Danciu, et al., 2014). Various initiatives where patients were

asked to opt-in to share their data for research purposes have been successful in both Salford, UK and in Sweden (New, et al., 2018).

One key informant states that, overtime, trials will be tailored to the individual patient, specifically in rare diseases, and technology will be used more effectively to remove the time spent on transcribing or entering data, thus giving the patient more time with the doctor (Young).

6.2.13 Better Data Analytics for Clinical Research/New Technologies

Advances in technology may transform the clinical trial data collection process (Bhavnani, et al., 2017). New technologies such as artificial intelligence (Bookbinder, 2017), and blockchain can make data sharing more efficiency (Anisingaraju, 2017) and secure (Nugent, et al., 2016). The growing number of technology solutions (Byers, 2017) coupled with growing data volumes (Bhadani & Jothimani, 2016) creates opportunities in how health data is used.

Louis Smith is interested in exploring links between data across trials. He refers to Citeline (PI, 2018), which contains vast quantities of site-related data, but this data is not aggregated nor compared during the trial feasibility process. Ross Rothmeier highlights that historical data is useful for predictive analysis and to help with building a more targeted approach to data capture, while Jeff Lee believes that the growth of machine-to-machine learning will change data collection practices.

6.3 INTERVIEW-ONLY THEMES

Several themes emerged in the key informants' interviews but not in the literature review, and are outlined in Table 6.3: Interview-Only Themes.

TABLE 6-3 : INTERVIEW-ONLY THEMES

#	Theme	Literature	Interviews
1	Patient Burden & New Technology to Engage the Patient	-	√
2	mHeath	-	√
3	The New Clinical Trial Landscape - New Trial Designs	-	√

6.3.1 Patient Burden & New Technology to Engage the Patient

Several key informants refer to patient burden when participating in a trial and that there must be a drive to engage the patient, particularly as mHealth involves greater patient involvement (Byrom, McCarthy Smith and Lee)

The pharmaceutical industry is losing sight of the burden to the patient that mHealth poses. Alleviating patient burden will be a factor in determining the successful deployment of new data collection technologies and technology must be useful for the patient (Lee, Byrom, McCarthy & Young). Byrom emphasises that new technology should not "become part of the intervention itself".

6.3.2 mHealth

mHealth will challenge current data processing practices. The change in the types of clinical research data is placing the emphasis on the use of machine learning and artificial intelligence, as current clinical trial data cleaning processes will not work with future technology (Rothmeier). There is a need to move from outmoded practices such as source data verification (Young).

6.3.2.1 Remote Data Collection in Clinical Research

Remote data collection will be restricted by the regulations that require patients to travel to a hospital site to consent to their participation in a trial or for them to receive trial medications. A "*leaner more virtualised study model"* will evolve (Lee). A "hub and spoke" model is more likely, where one super site will oversee the safety aspects of a trial, and less experienced investigators/sites will manage more mundane data collection activities (Byrom). Young suggests a hybrid data collection model will evolve, where a remote data collection will work for some patients, but perhaps not for others.

6.3.2.2 New Clinical Trial Data and Endpoints

New standard measurements and endpoints will need to be developed for wearables and sensors (Lee, McCarthy and Young)

6.3.2.3 Personalised Medicine and Pragmatic Trials

Personalised or precision medicine will evolve as result of programs such as the 1,000 genomes project (IGSR, 2018), and it is predicated that there will be a rise in the number of pragmatic trials using a real-world data (Young).

6.3.3 The New Clinical Trial Landscape - New Trial Designs

Historical and genomic data may be re-used to develop data models resulting in the need for data from fewer patients (Rothmeier).

The randomised controlled trial model will prevail to ensure rigorous research, while the 21st Century Cures Act will drive the collection of increased volumes of supplementary real-world data. There may be a need to deploy apps to larger patient populations, realising that the drop-out rate may be higher (Byrom).

6.4 CONCLUSION

This chapter presented a comparison of the results of the literature review and the key informants' interviews. Thirteen themes were represented in the literature and the key informants' interviews, while a further three themes was raised by the key informants (but not in the literature). In conclusion, the themes raised by the key informants are broadly in line with the themes in the literature, although it could be argued that they had a different emphasis. The literature is concentrated on the development of standards and interoperability, while the key informants place greater emphasis on obtaining access to data and implementing innovative technologies for the harnessing of health records.

7 RESEARCH CONCLUSIONS

Clinical research is ripe for disruption. There is a broad recognition that issues prevail with the current clinical research data collection processes and systems. The lack of generalisability of clinical research results to broader populations is a concern, and the belief prevails that the data collected is insufficient and not being used in the right way.

The volume of health technologies available today will drive change in the collection of health data for secondary use in clinical research, but the conservative nature of the industry will mean this change will be by stealth. There will still be a need to collect data in a prescribed and controlled way for clinical research, and in the intermediate term, it is unlikely that wearables, sensors and EHR data will replace existing EDC systems, but it is likely that EHR and mHealth data will complement and begin to supplement certain variables used in clinical research.

In the longer term, the convergence of EHR and EDC systems is likely. Whether this means that EDC consume EHR technologies (or vice versa) is yet unclear.

The clinical landscape is shifting, and the crossover from healthcare to clinical research will drive change. The patient is emerging as a force of change, but their privacy must be respected and they should be equal partners in clinical research initiatives. Institutions are reluctant to share EHR data but the influence of the patient may change this practice, as their growing awareness will put the patient in charge of their own health records.

7.1 FUTURE WORK

It is evident that changes are happening in how health data is collected for secondary use in clinical research. Influences such as advances in technology and the growing awareness of the patient in how their data can be used will drive change, but may be stifled by regulation, and a reluctance to change clinical trial data collection methods for fear of falling foul of the regulator.

7.1.1 Data Anonymity & Patient Privacy

Currently, data is collected anonymously for phase II-III clinical trials, and researchers receive patient data second-hand from clinicians, without having direct contact with the patient. The use of mHealth technologies will require researchers to have direct patient contact to allow for the deployment of devices and to provide support services such as technical advice. Removing patient anonymity from the phase II-III trial process will force a rethink of clinical trial design particularly related to randomisation practices.

7.1.2 New Extraction Techniques

The use of EHR data as a supplement or replacement for EDC data, will potentially remove the patient's anonymity. By virtue of validated systems, all codes or de-identification steps will be captured in the audit trail by research teams. There are suggested solutions to de-identify the data by removing variables such as date of birth in clinical research datasets. This would require the need for additional extraction technologies to be deployed at hospital sites which may prove to be costly and not scalable across the myriad of EHR systems.

7.1.3 Common Standards

There are many successful examples of how EHR data can be used for research. Under closer scrutiny, EHR data is largely being used for in-house hospital research or to assist with the administration of clinical trials. Admittedly, EHRs are a valuable source of metrics on patient populations and assist in the clinical trial feasibility, but, few, if any, examples exist where EDC data is being used as the source data to replace the data collected within EDC systems (eCRF). There are too many EDC systems and standards vary greatly across systems. Additionally different coding standards and definitions of data make interoperability across EHR systems and with EDC systems problematic. Ideally, new standards, that can apply to both data collected from primary care and secondary use are required, and may be achieved through machine learning techniques.

7.1.4 How will mHealth Work?

Although there is a proliferation of new technologies to capture health data for secondary use, there is no clear path to mHealth use in clinical research. Trials are already using wearable devices and sensors to capture data directly from patients but solutions for how this data will be harnessed and processed remain unclear, and it is suggested that wearable devices may be suited to certain therapeutic indications or patient populations. Patient burden will be a factor in driving the success or failure of these initiatives. As patients become more informed, mHealth technologies and the associated processes must provide some value to them or their families.

7.1.5 Conclusion

In 1969, Greenes et al stated that “increasing activity in the use of computers for acquisition, storage, and retrieval of medical information has been stimulated by the growing complexity of medical care, and the need for standardisation, quality control, and retrievability of clinical data” (Greenes, et al., 1969).

Almost 50 years later, this statement is still applicable, and it points to the slow pace at which change happens in how health data is organised. It is likely that the pace of digitisation will continue to speed up, but the clinical research industry will continue to evolve cautiously over the next 5-10 years.

8 TABLE OF ABBREVIATIONS

Term	Definition
21 Code of Federal Regulations (CFR) part 11	Title 21 CFR Part 11 is the part of Title 21 of the Code of Federal Regulations that establishes the United States Food and Drug Administration (FDA) regulations on electronic records and electronic signatures (ERES), which mandates the need for an audit trail and other controls in clinical trial data collection systems when the data will form of a regulatory authority submission.
21st Century Cures Act	The 21st Century Cures Act (Cures Act), signed into law on December 13, 2016, is designed to help accelerate medical product development and bring new innovations and advances to patients who need them faster and more efficiently.
Actigraph or Actigraphy	Actigraphy is a non-invasive method of monitoring human rest/activity cycles. A small actigraph unit, also called an actimetry sensor, is worn for a week or more to measure gross motor activity. The unit is usually, in a wrist-watch-like package, worn on the wrist.
Agency for Healthcare Research and Quality (AHRQ)	The Agency for Healthcare Research and Quality (AHRQ) is a US agency with responsibility for improving the safety and quality of US's health care system.
ALCOA	Clinical trial data should meet certain fundamental elements of quality. Whether they're recorded on paper or electronically, source data should be attributable, legible, contemporaneous, original, and accurate (ALCOA). If these best practices for source documentation aren't followed, there is no valid evidence that the test article is safe and effective.
American College of Cardiology (ACC)	The American College of Cardiology, based in Washington, D.C., is a non-profit medical association established in 1949. The ACC is the professional home for the entire cardiovascular care team, and its mission to transform cardiovascular care and improve heart health.
Applied Machine Learning (AML)	The design of systems that can learn from and make decisions and predictions based on data. Machine learning enables computers to act and make data-driven decisions rather than being explicitly programmed to carry out a certain task.
Artificial Intelligence (AI)	The theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages.
Blinding	Blinding is used in the design of some clinical trials and other research studies to try to eliminate the bias of expectation influencing the research findings. A blind trial is a trial where the participants do not know which treatment/intervention they have been allocated.
Blockchain	A **blockchain** is a decentralized, distributed and public digital ledger that is used to record transactions across many computers so that the record cannot be altered retroactively without the alteration of all subsequent blocks and the consensus of the network.
BYOD	BYOD, or bring your own device, in a clinical context, refers to patients, who use their own computing devices – such as smartphones, laptops and tablets – in the management of their health and to track their health data.
Cancer Data Standards Repository (caDSR)	caDSR (Cancer Data Standards Registry and Repository) is the National Cancer Institute's (US) database and a set of APIs and tools used to create, edit, control, deploy and find common data elements (CDEs) for metadata consumers and for UML model development.

Term	Definition
Case Record Form (CRF) & Annotated CRF	A case record form (or CRF/Cerf) is a paper or electronic questionnaire used collect clinical trial data from each participating patient. All patient data is documented in the CRF. The annotated CRF is a blank CRF that maps each item on the CRF to the corresponding variables in the database. The annotated CRF provides the variable names and coding for each CRF item included in the data tabulation datasets.
CDASH	CDASH (see CDISC) establishes a standard way to collect data in a similar way across studies and sponsors so that data collection formats and structures provide clear traceability of submission data into the Study Data Tabulation Model (SDTM), delivering more transparency to regulators and others who conduct data review.
CiteLine	CiteLine is an analytics company that provides pharmaceutical companies and clinical research organisations with site-related intelligence (using a subscription model) to assist with assessing trial feasibility.
Clinical Study/Study	A clinical study involves research using human volunteers (also called participants) that is intended to add to medical knowledge.
Clinical Trial Protocol	The protocol is a document that describes how a clinical trial will be conducted and includes details such as the objective(s), design, methodology, statistical considerations, and organization of a clinical trial, and ensures the safety of the trial subjects and the integrity of the data collected.
Common data elements (CDEs)	Developed by the NIH, a CDE is a data element that is common to multiple data sets across different studies.
CRF completion guidelines	A CRF completion guideline is a document to assist the investigator to complete the CRF/eCRF.
Declaration of Helsinki	The Declaration of Helsinki (DoH) is a set of ethical principles regarding human experimentation developed for the medical community by the World Medical Association (WMA). It is widely regarded as the cornerstone document on human research ethics.
De-identify/De-identification	De-identification is the process used to prevent a person's identity from being connected with information. Common uses of de-identification include human subject research for the sake of privacy for research participants.
Demography data	Demographic data refers to data that is statistically socio-economic in nature such as population, race, income, education, and employment, which represent specific geographic locations and are often associated with time.
Duchenne's/Duchenne Muscular Dystrophy (DMD)	Duchenne muscular dystrophy (DMD) is a genetic disorder characterized by progressive muscle degeneration and weakness. It is a rare and fatal degenerative disease for which there is no cure. The symptom of muscle weakness usually begins around the age of four in boys and worsens quickly. Typically muscle loss occurs first in the upper legs and pelvis followed by those of the upper arms.
EDC	Electronic Data Capture (EDC) - the system used to collect patient health data electronically through an eCRF
EHR4CR	Electronic Health Records for Clinical Research
Endpoints	In a **clinical** research **trial**, a **clinical endpoint** generally refers to occurrence of a disease, symptom, sign, or laboratory abnormality that constitutes one of the target outcomes of the **trial.**

Term	Definition
ePRO/electronic outcomes measurements	An electronic patient-reported outcome (ePRO) is a patient-reported outcome that is collected by electronic methods. ePRO methods are most commonly used in clinical trials, but they are also used elsewhere in health care.
Ethics Board	Hospital ethics committee must convene to decide a complex matter such as approvals of clinical trials within its organisation. These bodies are composed primarily of healthcare professionals, but may also include philosophers, lay people, and clergy – indeed, in many parts of the world their presence is considered mandatory to provide balance. U.S. recommendations suggest that Research and Ethical Boards (REBs) should have five or more members, including at least one scientist, one non-scientist, and one person not affiliated with the institution. The REB should include people knowledgeable in the law and standards of practice and professional conduct. Special memberships are advocated for handicapped or disabled concerns, if required by the protocol under review
EU-ADR	The EU-ADR project aims to develop a computerized system to detect adverse drug reactions (ADRs)from electronic healthcare records (EHRs) of over 30 million patients from several European countries
EUPATI	EUPATI stands for 'European Patients Academy on Therapeutic Innovation', and develops educational material, training courses and a public Internet library to educate patient representatives and the lay public about all processes involved in how medicines are developed.
European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP)	The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP®) is a network coordinated by the European Medicines Agency (EMA)
Farr Institute of Health Informatics Research	The Farr Institute is a UK-based collaboration involving academic institutions and health partners focused on research in health informatics to advance the health and care of patients and the public.
FDA	The Food and Drug Administration is a federal agency of the United States Department of Health and Human Services, one of the United States federal executive departments.
Gamification	The application of typical elements of game playing (e.g. point scoring, competition with others, rules of play) to other areas of activity, typically as an online technique to encourage engagement with a product or service.
General Data Protection Regulation (GDPR	The General Data Protection Regulation (GDPR) of the European Parliament, the Council of the European Union and the European Commission and is intended to strengthen and unify data protection for all individuals within the European Union (EU).
Health Information Technology Standards Panel (HITSP)	The American National Standards Institute (ANSI) Healthcare Information Technology Standards Panel (HITSP) is part of the United States Department of Health and Human Services) to promote interoperability in health care by harmonizing health information technology standards.
Health Insurance Portability and Accountability Act (HIPAA)	The Health Insurance Portability and Accountability Act of 1996 (HIPAA) (US) protects health insurance coverage. Title II of HIPAA, known as the Administrative Simplification (AS) provisions, requires the establishment of national standards for electronic health care transactions and national identifiers for providers, health insurance plans, and employers.

Term	Definition
Health Level 7 (HL7) Reference Information Model (RIM)	The Reference Information Model (RIM) is the cornerstone of the HL7 Version 3 development process. An object model created as part of the Version 3 methodology, the RIM is a large, pictorial representation of the HL7 clinical data (domains) and identifies the life cycle that a message or groups of related messages will carry. It is a shared model between all domains and, as such, is the model from which all domains create their messages.
Hitech Act (2009)	The Health Information Technology for Economic and Clinical Health Act (HITECH Act) sets forth a federal standard for security breach notifications relating to the unauthorized dissemination of protected health information (PHI).
Home monitoring devices	Home monitoring devices or remote patient monitoring (RPM), also called homecare telehealth, is a type of ambulatory healthcare that allows a patient to use a mobile medical device to perform a routine test and send the test data to a healthcare professional in real-time.
IMP	Investigational medicinal product – the drug or compound under investigation within a clinical trial
International Conference on Harmonisation's (ICH) Good Clinical Practice (GCP)	The International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) is unique in bringing together the regulatory authorities and pharmaceutical industry to discuss scientific and technical aspects of drug registration.
Interoperability	The ability of computer systems or software to exchange and make use of information.
Investigator(s)	An investigator involved in a clinical trial is responsible for ensuring that an investigation is conducted according to the signed investigator statement, the investigational plan, and applicable regulations; for protecting the rights, safety, and welfare of subjects under the investigator's care; and for the control of drugs under investigation. The investigator must also meet requirements set forth by the FDA, EMA or other regulatory body. The qualifications must be outlined in a current resume and readily available for auditors.
ISO	The International Organisation for Standardization (ISO) is an international standard-setting body composed of representatives from various national standards organizations
ISO/IEC	ISO/IEC JTC 1 is the standards development environment where experts come together to develop worldwide standards for business and consumer applications for integrating diverse and complex IT technologies.
KPIs	Key Performance Indicators IKPIs) help businesses and employees define and achieve their goals by using metrics to assess performance
Linked Open Data (LOD)	Linked Open data is data that is open if anyone is free to use, reuse, and redistribute and linked using a method of publishing structured data so that it can be interlinked and read automatically by computers.
Logical Observation Identifiers, Names, and Codes (LOINC)	The universal standard for identifying health measurements, observations, and documents.
MedDRA	The Medical Dictionary for Regulatory Activities (MedDRA) was created to manage clinical information about pharmaceuticals, biologics, vaccines, and drug-device combinations for the entire lifespan of products.
Metadata registries (MDRs)	A metadata registry is a central location in an organization where metadata definitions are stored and maintained in a controlled method.
METeOR	METeOR is Australia's repository for national metadata standards for health, housing and community services statistics and information.

Term	Definition
mHealth	mHealth (mobile health) is a general term for the use of mobile phones and other wireless technology in medical care. The most common application of mHealth is the use of mobile phones and communication devices to educate consumers about preventive health care services.
Mini-Sentinel	The FDA's "Mini-Sentinel" pilot program is a US-based rapid-response electronic safety surveillance system for drugs and other medical products.
National Cancer Institute (NCI)	The National Cancer Institute (NCI) is part of the National Institutes of Health (NIH), which is one of eleven agencies that are part of the U.S. Department of Health.
NHS	The National Health Service (NHS) is the name used for each of the four public health services in the United Kingdom
NIH	National Institutes of Health (US) - seeks fundamental knowledge about the nature and behaviour of living systems and the application of that knowledge to enhance health, lengthen life, and reduce illness and disability.
OpenEHR	OpenEHR is a virtual community working to transform health data into electronic formats while ensuring universal interoperability.
Open-source	Open-source software (OSS) is computer software with its source code made available with a license in which the copyright holder provides the rights to study, change, and distribute the software to anyone and for any purpose.
Patient reported outcomes (PRO)	PRO is directly reported by the patient without interpretation of the patient's response by a clinician or anyone else and pertains to the patient's health, quality of life, or functional status associated with health care or treatment.
PCORI/PCORnet	PCORnet, the National Patient-Centred Clinical Research Network, is an innovative initiative of the Patient-Centred Outcomes Research Institute (PCORI). It is designed to make it faster, easier, and less costly to conduct clinical research than is now possible by harnessing the power of large amounts of health data and patient partnerships. In the process, it is transforming the culture of clinical research from one directed by researchers to one driven by the needs of patients and those who care for them.
Petabyte	A petabyte is 10^{15} or 1,000,000,000,000,000 bytes.
Post-marketing trials/surveillance	Post marketing study studies are conducted after the regulator has approved a product for marketing. FDA uses post marketing study commitments to gather additional information about a product's safety, efficacy, or optimal use.
Post-marketing trials/surveillance	Post-marketing monitoring includes the identification and monitoring of new additional adverse drug events from doctors or other health professionals. Unlike previous stages, these will be observational studies where the long-term effectiveness will be evaluated; these are conducted right after the commercialization of the drug to the "real world".
Pragmatic trial design	Pragmatic trials measure effectiveness-the benefit the treatment produces in routine clinical practice.
Pre-marketing Trials	Pre-marketing clinical trials are studies conducted to evaluate, first, the safety and, second, the efficacy of the new compound in humans.
Provider	Healthcare provider, usually a hospital
Randomised controlled clinical (RCT) trials	A randomized controlled trial (RCT) is a type of experiment which aims to reduce bias when testing a new treatment. The people participating in the trial are randomly allocated to either the group receiving the treatment under investigation or to a group receiving standard treatment (or placebo treatment) as the control. The RCT is often considered the gold standard for a clinical trial. RCTs are often used to test the efficacy or effectiveness of various types of

Term	Definition
	medical intervention and may provide information about adverse effects, such as drug reactions.
Real-world Evidence (RWE)	Real world evidence (RWE) or real-world data (RWD) is data used for decision-making that are not collected in conventional randomized controlled trials (RCTs), includes clinical and economic data reported by patient registries, claims databases, electronic health records, patient-reported outcomes, and literature review. Real-world evidence = organized information informing a conclusion or judgment based on real-world data.
Regulatory authority	A regulatory agency or authority is a public authority or government agency responsible for exercising autonomous authority over some area of human activity in a regulatory or supervisory capacity.
Robotic Process Optimisation (RPA)	Robotic process automation (**RPA**) is the use of software with artificial intelligence (AI) and machine learning capabilities to handle high-volume, repeatable tasks that previously required humans to perform. These tasks can include queries, calculations and maintenance of records and transactions.
SDTM	**SDTM** (Study Data Tabulation Model) defines a standard structure for human clinical trial (study) data tabulations and for nonclinical study data tabulations that are to be submitted as part of a product application to a regulatory authority such as the United States Food and Drug Administration (FDA)
Semantic web technology	Semantic technology defines and links data on the web or within enterprise systems by developing languages to express rich, self-describing interrelations of data in a form that machines can process.
Sponsor	A person, company, institution, group, or organization that oversees or pays for a clinical trial and collects and analyses the data. Also called trial sponsor.
Static datasets	Static data is data that does not change after being recorded, usually held in a fixed data set.
Study investigators/Investigators	A clinical investigator (usually a hospital doctor) is involved in a clinical trial is responsible for ensuring that an investigation is conducted according to the signed investigator statement, the investigational plan, and applicable regulations; for protecting the rights, safety, and welfare of subjects under the investigator's care; and for the control of drugs under investigation. The Clinical Investigator must also meet requirements set forth by the FDA, EMA or other regulatory body.
Synthetic Control Arm	Clinical trials of experimental treatments require control arms. However, a randomized control arm may be difficult or impossible for reasons including ethical concerns about assigning a placebo and patient unwillingness to be randomized, possibly to a placebo. Historical control groups from one or a few previous clinical trials have often been used, but this approach introduces biases due to differences in baseline covariates, sites, and other factors. We minimize these problems by constructing a synthetic control arm (SCA) from Medidata's archive of >3000 trials with data rights for anonymized aggregated analyses. For a specified single-arm trial, we create an SCA containing patients from recent trials with similar eligibility criteria. We use several approaches to select patients for the SCA to match the patients in the trial, including all available patients and patient-level matching on key baseline covariates. The SCA can provide a superior alternative to using a single arm or historical controls from literature, where covariates cannot be matched (JSM, 2017).
Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT)	SNOMED is a systematically organized computer processable collection of medical terms providing codes, terms, synonyms and definitions used in clinical documentation and reporting.

Term	Definition
The Clinical Data Interchange Standards Consortium (CDISC)	The Clinical Data Interchange Standards Consortium (CDISC) is an open, multidisciplinary, non-profit standard developing organization (SDO) that is working to develop global standards and innovations to streamline medical research and ensure a link with healthcare.
The Federal Health Information Model (FHIM)	The FHIM is an information model of healthcare data.
The Society for Clinical Data Management (SCDM)	The Society for Clinical Data Management (SCDM) is a non-profit global association for Clinical Data Management professionals worldwide set-up to advance the discipline of Clinical Data Management in all regions of the world.
Therapeutic indications	Therapeutic indications are a description of the disease to be treated with a medicine, and the population for which the medicine is intended
Unblinding	Unblinding is the disclosure to the participant and/or study team of which treatment the participant received during the trial. The process of unblinding is planned and included in the study protocol. Unblinding a trial is a necessary process to protect participants in the event of medical or safety reasons. There is also a defined process to 'break the blind' of a single participant when required.
Unified Medical Language System (UMLS)	Unified Medical Language System (UMLS) integrates and distributes key terminology, classification and coding standards, and associated resources to promote creation of more effective and interoperable biomedical information systems and services, including electronic health records.
USHIK (United States Health Information Knowledgebase)	The United States Health Information Knowledgebase (USHIK) is an on-line, publicly accessible registry and repository of healthcare-related metadata, specifications, and standards. USHIK is funded and directed by the Agency for Healthcare Research and Quality (AHRQ) with management support and engagement from numerous public and private partners.
Vital signs (data)	Vital signs are a group of the 4 to 6 most important signs that indicate the status of the body's vital (life-sustaining) functions. These measurements are taken to help assess the general physical health of a person, give clues to possible diseases, and show progress toward recovery.
Voice-activated system	A system that responds to or begins to operate in response to a person's voice
Wearables	The terms "**wearable** technology", "**wearable** devices", and "**wearables**" all refer to electronic technologies or computers that are incorporated into items of clothing and accessories which can comfortably be worn on the body
WHOdrug	The WHO Drug Dictionary is an international classification of medicines created by the WHO Programme for International Drug Monitoring and managed by the Uppsala Monitoring Centre and is used for identifying drug names in clinical trials.

9 WORKS CITED & BIBLIOGRAPHY

ABPI, 2012. *Guidelines for Phase 1 Clinical Trials,* London: ABPI.

Anisingaraju, S., 2017. *Genetic Engineering & Biotechnology News.* [Online] Available at: https://www.genengnews.com/gen-exclusives/optimizing-clinical-trials/77900933 [Accessed 21 June 2018].

Apple Inc, 2018. *Empower your patients with Health Records on iPhone.* [Online] Available at: https://www.apple.com/healthcare/health-records/ [Accessed 4 May 2018].

Arlett, P., 2016. *Update on Real World Evidence Data.* UNK, EMA.

Atherton, p. et al., 2016. atient Reported Outcomes Questionnaire Compliance in Cancer Cooperative Group Trials (Alliance N0992). *Clinical Trials,* 13(6), pp. 612-620.

Auffray, C. et al., 2016. Making sense of big data in health research: Towards an EU action plan. *Genome Medicine,* 8(71), p. UNK.

Barkow, S., 2016. *Data Integrity Workshop.* Bethesda, MD, FDA.

Bellary, S., Krishnankutty, B. & Latha, M., 2014. Basics of case report form designing in clinical research. *Perspectives in Clinical Research,* 5(4), pp. 159-166.

Bhadani, A. & Jothimani, D., 2016. Big Data: Challenges, Opportunities, and Realities. In: M. In Singh & D. Kumar, eds. *Effective Big Data Management and Opportunities for Implementation.* Pennslyvania: UNK, pp. 1-24.

Bhavnani, S. et al., 2017. 2017 Roadmap for Innovation—ACC Health Policy Statement on Healthcare Transformation in the Era of Digital Health, Big Data, and Precision Health. *Journal of the American College of Cardiology,* 10(18), p. Unk.

Bookbinder, M., 2017. *Clinical Informatics News.* [Online] Available at: http://www.clinicalinformaticsnews.com/2017/09/29/the-intelligent-trial-ai-comes-to-clinical-trials.aspx [Accessed 21 June 2018].

Bourgeois, F. et al., 2017. Exclusion of Elderly People from Randomized Clinical Trials of Drugs for Ischemic Heart Disease. *Journal of the American Geriatrics Society,* 65(11), pp. 2354-2361.

Byers, C., 2017. *The Growing Importance of IT in Healthcare.* [Online] Available at: https://mytechdecisions.com/it-infrastructure/growing-importance-healthcare/ [Accessed 09 June 2018].

Centre for Drug Evaluation & Research (FDA), 2017. *Drug Safety Priorities 2017,* Silver Spring, MD: FDA.

Chamberlain, M., Grasso, J., Molitch, M. E. & Toledo, P., 2015. Qualitative Analysis of Key Informant Interviews to Establish Barriers and Solutions for Diabetes Care in a Rural Bolivian Clinic. *The Journal of Global Health Care Systems,* 5(3), p. UNK.

CIRT, n.d.. *Grand Canyon University: Centre for Innovation in Research and Teaching.* [Online] Available at: https://cirt.gcu.edu/research/developmentresources/research_ready/quantresearch/sample_meth [Accessed 05 February 2018].

Clinical Research Corporation, 2017. *Clinical Research Corporation.* [Online] Available at: http://crcaustralia.com/media-releases/real-world-data/ [Accessed 31 May 2018].

CMS.gov, 2012. *Centres for Medicare and Medicaid Services.* [Online] Available at: https://www.cms.gov/Medicare/E-Health/EHealthRecords/index.html [Accessed 17 June 2018].

Cowie, M. et al., 2017. Electronic health records to facilitate clinical research. *Springer,* 106(Unk), pp. 1-9.

Creswell, J. W., 2007. Designing a Qualitative Study. In: UNK, ed. *Qualitative Inquiry and Research Design: Choosing Among Five Approaches (2nd Ed).* Thousand Oaks: SAGE, pp. 35-41.

Danciu, I. et al., 2014. Secondary use of clinical data: The Vanderbilt approach. *Journal of Biomedical Informatics,* 52(UNK), pp. 28-35.

De Moor, G. et al., 2015. Using electronic health records for clinical research: The case of the EHR4CR project. *Journal of Biomedical Informatics,* 53(UNK), pp. 162-173.

Dudovskiy, J., 2017. *Research Methodology.* [Online] Available at: https://research-methodology.net/sampling-in-primary-data-collection/convenience-sampling/ [Accessed 9 March 2018].

EC, 2018. *eHealth News.* [Online]
Available at: http://www.ehealthnews.eu/development/5537-data-in-the-eu-commission-steps-up-efforts-to-increase-availability-and-boost-healthcare-data-sharing
[Accessed 06 June 2018].

EHR4CR, 2015. *EHR4CR.* [Online]
Available at: http://www.ehr4cr.eu/
[Accessed 23 November 2017].

EMA, 2015. *European Medicines Agency.* [Online]
Available at: http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000334.jsp&mid=WC0b01ac05800ba1d9
[Accessed 07 June 2018].

Embi, P. & Payne, P., 2014. Advancing methodologies in Clinical Research Informatics (CRI): Foundational work for a maturing field. *Journal of Biomedical Informatics,* 52(UNK), pp. 1-3.

EOSC, 2018. *EOSC.* [Online]
Available at: https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud
[Accessed 07 June 2018].

EUPATI, 2015. *Data Collection in Clinical Trials.* [Online]
Available at: https://www.eupati.eu/clinical-development-and-trials/data-collection-clinical-trials/
[Accessed 9 January 2018].

FDA, 2018. *Clinical Research.* [Online]
Available at: https://www.fda.gov/ForPatients/Approvals/Drugs/ucm405622.htm
[Accessed 05 June 2018].

FDA, 2018. *FDA.* [Online]
Available at: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=11
[Accessed 2 July 2018].

FDA, 2018. *Real World Evidence.* [Online]
Available at: https://www.fda.gov/ScienceResearch/SpecialTopics/RealWorldEvidence/default.htm
[Accessed 31 May 2018].

Frieden, T., 2017. Evidence for Health Decision Making — Beyond Randomized, Controlled Trials. *The New England Journal of Medicine,* 377(UNK), pp. 465-475.

Getz, K., 2017. Anticipating the Groundswell of RWD and RWE Usage. *Applied Clinical Trials*, 1 December, p. UNK.

Getz, K., 2018. Surmounting eClinical Data Volume and Diversity. *Applied Clinical Trials,* 27(3), p. UNK.

Google, 2018. *Google rolls out a new tool to help health providers solve the medical record mess.* [Online]

Available at: https://www.cnbc.com/2018/03/05/google-cloud-healthcare-api-to-address-medical-reord-interoperability.html

[Accessed 4 May 2018].

Greenes, R. A., Pappalardoab, N., Marble, C. W. & Barnett, G. O., 1969. Design and implementation of a clinical data management system. *Computers and Biomedical Research,* 2(5), pp. 469-485.

Hublou, R., 2016. *Pharmaceutical Processing.* [Online]

Available at: https://www.pharmpro.com/article/2016/09/biggest-challenges-delivering-clinical-trials-time-within-budget

[Accessed 22 November 2017].

ICH, 1998. *Statistical principles for clinical trials (E9).* s.l., ICH.

ICH, 2015. *INTERNATIONAL CONFERENCE ON HARMONISATION OF TECHNICAL REQUIREMENTS FOR REGISTRATION OF PHARMACEUTICALS FOR HUMAN USE.* UNK, ICH.

ICHOM, 2018. *ICHOM.* [Online]

Available at: http://www.ichom.org/who-we-are/

[Accessed 24 April 2018].

IGSR, 2018. *IGSR: The 1,000 Genome Sample Resource.* [Online]

Available at: http://www.internationalgenome.org/

[Accessed 25 June 2018].

IQVIA, 2017. *IQVIA Analyst and Investor Conference.* New York, IQVIA, p. 82.

Johnson, S. B., Farach, F. J., Pelphrey, K. & Rozenblit, L., 2016. Data management in clinical research: Synthesizing stakeholder perspectives. *Journal of Biomedical Informatics,* 60(Unk), pp. 286-293.

JSM, 2017. *Constructing a Synthetic Control Arm from Previous Clinical Trials, with Application to Cancer Trials.* Baltimore, JSM.

Kamateri, E., Kalampokis, E., Tambouris, E. & Tarabanis, K., 2014. The linked medical data access control framework. *Journal of Biomedical Informatics,* 50(UNK), pp. 213-225.

Kesselheim, A. & Avorn, J., 2017. New “21st Century Cures” Legislation. *The Jama Network,* 317(6), pp. 581-2.

Krishnankutty, B., Bellary, S., Kumar, N. & Moodahadu, L., 2012. Data management in clinical research: An overview. *India Journal of Pharmacology,* Mar-Apr(44 (2)), p. 168–172.

Lester, S., 1999. *Royal Geographical Society: An introduction to phenomenological research.* [Online] Available at: https://www.rgs.org/NR/rdonlyres/F50603E0-41AF-4B15-9C84-BA7E4DE8CB4F/0/Seaweedphenomenologyresearch.pdf
[Accessed 05 February 2018].

Longhurst, R., 2016. Semi Structured Interviews & Focus Groups. In: R. Rojek, et al. eds. *Key Methods in Geography.* Glasgow: SAGE, pp. 145-153.

Luveaux, S., 2016. *Big data and the new EU data protection legislation.* London, European Medicines Agency.

Lu, Z. & Su, J., 2010. Clinical data management: Current status, challenges, and future directions from industry perspectives. *Open Access Journal of Clinical Trials,* 2(UNK), pp. 93-105.

Maldoff, G., 2016. *The International Association of Privacy Professionals.* [Online] Available at: https://iapp.org/news/a/how-gdpr-changes-the-rules-for-research/
[Accessed 8 January 2018].

Marshall, M. N., 1996. The Key Informant Technique. *Family Practice,* 13(1), pp. 92-97.

Moore, D. et al., 2000. How generalizable are the results of large randomized controlled trials of antiretroviral therapy?. *HIV Medicine,* 1(UNK), pp. 149-154.

Mullin, R., 2014. *Scientific American.* [Online] Available at: https://www.scientificamerican.com/article/cost-to-develop-new-pharmaceutical-drug-now-exceeds-2-5b/
[Accessed 22 November 2017].

Murphy, E., Ferris, F. & O'Donnell, R., 2007. An Electronic Medical Records System for Clinical Research and the EMR–EDC Interface. *Investigative Opthalmology & Visual Services,* 48(10), pp. 4383-89.

National Academy of Sciences, 2009. *Data Collection Standards and Monitoring.* Washington DC, NCBI.

NCBI, 2018. *NCBI.* [Online] Available at: https://www.ncbi.nlm.nih.gov/ [Accessed 18 June 2018].

New, J. et al., 2018. Putting patients in control of data from electronic health records. *BMJ,* 360(5554), p. UNK.

NHS, 2017. *University Hospital Southampton.* [Online] Available at: http://www.uhs.nhs.uk/Research/For-public-and-patients/What-is-clinical-research.aspx [Accessed 24 January 2018].

NIH, 2016. *National Institutes of Health.* [Online] Available at: https://www.nih.gov/health-information/nih-clinical-research-trials-you/need-awareness-clinical-research [Accessed 22 November 2017].

Nugent, T., Upton, D. & Cimpoesu, M., 2016. Improving data transparency in clinical trials using blockchain smart contracts. *F1000 Res,* 1(5), p. 2541.

OECD, 2016. *OECD.* [Online] Available at: http://www.oecd.org/health/health-systems/PaRIS-Booklet.pdf [Accessed 9 January 2018].

ONC, 2017. *fficial Website of The Office of the National Coordinator for Health Information Technology (ONC).* [Online] Available at: https://www.healthit.gov/topic/federal-incentive-programs/meaningful-use [Accessed 13 June 2018].

Patel, V. & Kaelber, D., 2014. Using aggregated, de-identified eletronic health record data for multivariate pharmacosurveillance: A case study of azathioprine. *Journal of Biomedical Informatics,* 52(UNK), pp. 36-42.

PatientsLikeMe, 2018. *PatientsLikeMe.* [Online]
Available at: https://www.patientslikeme.com/
[Accessed 13 June 2018].

Pfizer, 2011. *Pfizer Conducts First "Virtual" Clinical Trial Allowing Patients to Participate Regardless Of Geography.* [Online]
Available at: http://press.pfizer.com/press-release/pfizer-conducts-first-virtual-clinical-trial-allowing-patients-participate-regardless-
[Accessed 1 May 2018].

Phelan, A. et al., 2016. Exclusion of Women of Childbearing Potential in Clinical Trials of Type 2 Diabetes Medications: A Review of Protocol-based Barriers to Enrollment. *Diabetes Care,* 39(6), p. 1004–1009.

PI, 2018. *Citeline.* [Online]
Available at: https://pharmaintelligence.informa.com/products-and-services/data-and-analysis/citeline
[Accessed 13 June 2018].

Raghupathi, W. & Raghupathi, V., 2014. Big data analytics in healthcare: promise and potential. *Health Information Science and Systems,* 2(3), p. UNK.

Regalado, A., 2015. *Technology Review.* [Online]
Available at: https://www.technologyreview.com/s/534591/us-to-develop-dna-study-of-one-million-people/
[Accessed 07 June 2018].

Richards, L., n.d. Qualitative Research Design. In: *Thinking Research.* s.l.:s.n., pp. 73-103.

Richesson, R. & Nadkarni, P., 2011. Data standards for clinical research data collection forms: current status and challenges. *Jamia,* 107(UNK), p. UNK.

Sackett, D. L., 1981. How to read clinical journals: I. why to read them and how to start reading them critically.. *Canadian Medical Association Journal,* 124(5), pp. 555-558.

Saczynski, J., McManus, D. & Goldberg, R., 2013. Commonly Utilized Data Collection Approaches in Clinical Research. *The American Journal of Medicine,* 126(11), pp. 946-950.

Shah, J. et al., 2010. Electronic Data Capture for Registries and Clinical Trials in Orthopaedic Surgery: Open Source versus Commercial Systems. *Clinical Orthopaedics and Related Research,* 468(10), pp. 2664-2671.

Siemens, T., 2016. How Will Technology Drive Global Clinical Trial Change by 2025?. *Applied Clinical Trials,* 24(12), p. 42.

Sills, G. & Brodie, M., 2009. Antiepileptic Drugs - Clinical Drug Development in Epilepsy. *Encyclopedia of Basic Epilepsy Research,* UNK(UNK), pp. 1477-1485.

Sinaci, A. & Erturkmen, G., 2013. A federated semantic metadata registry framework for enabling Interoperability across clinical research and care domains. *Journal of Biomedical Informatics,* 48(UNK), pp. 784-794.

Spier, R., 2002. The History of the Peer Review Process. *Trends in Biotechnology,* 20(8), pp. 357-8.

Sundar, R. C. A. et al., 2018. Clinical Outcome of Patients with Advanced Biliary Tract Cancer in a Dedicated Phase I Unit. *Clinical Oncology Journal,* 30(3), pp. 185-191.

Syngeron Ltd, n.d.. *India Mart.* [Online] Available at: https://www.indiamart.com/proddetail/clinical-data-management-6719967833.html [Accessed 07 May 2018].

Torjesen, I., 2015. Drug development: the journey of a medicine from lab to shelf. *The Pharmaceutical Journal,* Unk(Unk), p. Unk.

Transparency Life Sciences, 2018. *Transparency Life Sciences.* [Online] Available at: https://www.transparencyls.com/transparency-protocol-crowdsourcing [Accessed 13 June 2018].

Tufts, 2018. *Tufts Centre for the Study of Drug Development.* [Online] Available at: http://csdd.tufts.edu/news/complete_story/pr_ir_jan_feb_2018 [Accessed 9 January 2018].

UHSS, 2013. *US Dept of Health and Social Sciences.* [Online] Available at: https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html [Accessed 29 June 2018].

Underwood, G., 2016. The clinical trial of the future. *Pharma Times,* Unk(Unk), p. Unk.

US National Library of Medicine, 2017. *Clinical Trials.* [Online] Available at: https://clinicaltrials.gov/ct2/search/map [Accessed 24 April 2018].

Wang, L. & Motti, E., 2015. *Pharmaceutical Outsourcing.* [Online] Available at: http://www.pharmoutsourcing.com/Featured-Articles/174597-The-Increasing-Shift-of-

Clinical-Trials-to-CROs/

[Accessed 22 November 2017].

Waters, J., 2017. *Capilano University: Phenomenological Research Guidelines.* [Online] Available at: https://www.capilanou.ca/psychology/student-resources/research-guidelines/Phenomenological-Research-Guidelines/

[Accessed 05 February 2018].

Watson, C. et al., 2015. data capture via mobile electronic clinical outcome assessments (ecoa): Meeting New Demands for Patient Reported Outcomes (PRO). *Journal of mHealth,* UNK(UNK), p. UNK.

WHO, 2006. *Electronic Health Records: Manual for Developing Countries,* Geneva: World Health Organisation.

WHO, n.d.

Williamson, P. et al., 2012. Developing core outcome sets for clinical trials: issues to consider. *Biomed Central,* 13(UNK), p. 132.

WMA, 1964; amd 1975, 1983 & 1989. *Declaration of Helsinki,* Helsinki; Tokyo; Venice; Hong Kong: World Medical Association (WMA).

Wong, C., Siah, K. & Lo, A., 2018. Estimation of clinical trial success rates and related parameters. *Biostatistics,* UNK(Jan), p. UNK.

Woodcock, J., 2017. *Workshop at the National Academies of Sciences, Engineering, and Medicine.* Kansas, Endpoints News.

Woolf, V., 1929. *A Room of One's Own.* 2 ed. St Albans: Panther.

Zivin, J., 2000. Understanding Clinical Trials. *Scientific American Inc,* UNK April, 282(4), pp. 69-73.

APPENDICES

APPENDIX I – ETHICS SUBMISSION PACKAGE

Sheila Kelly, Stud. #: 16306176
TCD MSc HI, Dissertation (2017-18)

Revised Ethics Application (29 Jan 2018)

Summary of Changes

1. **EC recommendation:** Change the fact that the data will be destroyed at a particular time to say that the data will be destroyed once the examination period is completed.

Change made on page 2 of the Key Informants Information Sheet

I will erase all electronic voice recordings at the end of my research and when the examination period has concluded.

2. **EC recommendation:** Include a bullet point on the informed consent form to say that the respondents are expressing their own views and not the views of their companies/organisations. If this is not the case you will need to get permission from their employer to interview them. If so, please include an additional information sheet and consent form which will be sent to employers requesting their permission to interview their employee.

Change made on page 3 of the Key Informants Information Sheet/Informed Consent Form

I am expressing my own views and not the views of my employer or other organization.

3. **EC recommendation:** Please clarify where you got/will get the contact details for the key informants.

Outlined in the body of the research proposal (page 5):

Below I have provided a list of proposed key informants that I intend to contact through Linkedin, and ask them if they are willing to participate in the research. If they accept my invitation and agree to participate, I will provide them with the key informants' information sheet that I have prepared as part of the TCD Ethics application.

School of Computer Science & Statistics
Research Ethics Application

Part A

Project Title: **Challenges and Enablers in the Collection of Health Data for use in Phase II-III Clinical Trials**

Name of student: **Sheila Kelly**

Name of Supervisor: **Lucy Hederman**

TCD E-mail: **kellys77@tcd.ie** Contact Tel No.: **087 6819492**

Course Name and Code (if applicable): **MSc in Health Informatics**

Estimated start date of survey/research: **February 2018**

I confirm that I will (where relevant):

- Familiarize myself with the Data Protection Act and the College Good Research Practice guidelines http://www.tcd.ie/info_compliance/dp/legislation.php;
- Tell participants that any recordings, e.g. audio/video/photographs, will not be identifiable unless prior written permission has been given. I will obtain permission for specific reuse (in papers, talks, etc.)
- Provide participants with an information sheet (or web-page for web-based experiments) that describes the main procedures (a copy of the information sheet must be included with this application)
- Obtain informed consent for participation (a copy of the informed consent form must be included with this application)
- Should the research be observational, ask participants for their consent to be observed
- Tell participants that their participation is voluntary
- Tell participants that they may withdraw at any time and for any reason without penalty
- Give participants the option of omitting questions they do not wish to answer if a questionnaire is used
- On request, debrief participants at the end of their participation (i.e. give them a brief explanation of the study)
- Verify that participants are 18 years or older and competent to supply consent.
- If the study involves participants viewing video displays then I will verify that they understand that if they or anyone in their family has a history of epilepsy then the participant is proceeding at their own risk
- Declare any potential conflict of interest to participants.
- Inform participants that in the extremely unlikely event that illicit activity is reported to me during the study I will be obliged to report it to appropriate authorities.
- Act in accordance with the information provided (i.e. if I tell participants I will not do something, then I will not do it).

Signed: S Kelly Date: 12 Jan 18

Student

Part B

Please answer the following questions.		*Yes/No*
Has this research application or any application of a similar nature connected to this research project been refused ethical approval by another review committee of the College (or at the institutions of any collaborators)?		No
Will your project involve photographing participants or electronic audio or video recordings?		No
Will your project deliberately involve misleading participants in any way?		No
Does this study contain commercially sensitive material?		No
Is there a risk of participants experiencing either physical or psychological distress or discomfort? If yes, give details on a separate sheet and state what you will tell them to do if they should experience any such problems (e.g. who they can contact for help).		No
Does your study involve any of the following?	Children (under 18 years of age)	No
	People with intellectual or communication difficulties	No
	Patients	No

School of Computer Science and Statistics
Research Ethical Application Form

Details of the Research Project Proposal must be submitted as a separate document to include the following information:

1. Title of project
2. Purpose of project including academic rationale
3. Brief description of methods and measurements to be used
4. Participants - recruitment methods, number, age, gender, exclusion/inclusion criteria, including statistical justification for numbers of participants
5. Debriefing arrangements
6. A clear concise statement of the ethical considerations raised by the project and how you intend to deal with them
7. Cite any relevant legislation relevant to the project with the method of compliance e.g. Data Protection Act etc.

Part C

I confirm that the materials I have submitted provided a complete and accurate account of the research I propose to conduct in this context, including my assessment of the ethical ramifications.

Signed: M Kelly Date: 12 Jan '18
Student

There is an obligation on the lead researcher to bring to the attention of the SCSS Research Ethics Committee any issues with ethical implications not clearly covered above.

Part D

If external or other TCD Ethics Committee approval has been received, please complete below.

External/TCD ethical approval has been received and no further ethical approval is required from the School's Research Ethical Committee. I have attached a copy of the external ethical approval for the School's Research Unit.

Signed: M Kelly Date: 12 Jan '18
Student

Part E

If the research is proposed by an undergraduate or postgraduate student, please have the below section completed.

I confirm, as an academic supervisor of this proposed research that the documents at hand are complete (i.e. each item on the submission checklist is accounted for) and are in a form that is suitable for review by the SCSS Research Ethics Committee

Signed: Lucy Hederman Date: 12 Jan 2018
Supervisor

Completed application forms together with supporting documentation should be submitted electronically to the online ethics system - https://webhost.tchpc.tcd.ie/research_ethics/ When your application has been reviewed and approved by the Ethics committee, hardcopies with original signatures should be submitted to the School of Computer Science & Statistics, Room 104, Lloyd Building, Trinity College, Dublin 2.

Ethics Application Guidelines – 2016

CHECKLIST

Please ensure that you have submitted the following documents with your application:

1.	• SCSS Ethical **Application Form**	Yes
2.	• **Participant's Information Sheet** must include the following: a) Declarations from Part A of the application form; b) Details provided to participants about how they were selected to participate; c) Declaration of all conflicts of interest.	Yes
3.	• **Participant's Consent Form** must include the following: a) Declarations from Part A of the application form; b) Researchers contact details provided for counter-signature (your participant will keep one copy of the signed consent form and return a copy to you).	Yes
4.	• **Research Project Proposal** must include the following: a) You must inform the Ethics Committee **who** your intended participants are i.e. are they your work colleagues, class mates etc. b) How will you recruit the participants i.e. **how** do you intend asking people to take part in your research? For example, will you stand on Pearse Street asking passers-by? c) If your participants are under the age of 18, you must seek both parental/guardian AND child consent.	Yes
5.	• Intended **questionnaire**/survey/interview protocol/screen shots/representative materials (as appropriate)	Yes
6.	• **URL** to intended on-line survey (as appropriate)	N/A

Notes on Conflict of Interest

1. If your intended participants are work colleagues, you must declare a potential conflict of interest: you are taking advantage of your existing relationships in order to make progress in your research. It is best to acknowledge this in your invitation to participants.
2. If your research is also intended to direct commercial or other exploitation, this must be declared. For example, *"Please be advised that this research is being conducted by an employee of the company that supplies the product or service which form an object of study within the research."*

Notes for questionnaires and interviews

1. If your questionnaire is paper based, you must have the following **opt-out** clause on the top of each page of the questionnaire: *"Each question is optional. Feel free to omit a response to any question; however the researcher would be grateful if all questions are responded to."*
2. If you questionnaire is **on-line**, the first page of your questionnaire must repeat the content of the information sheet. This must be followed by the consent form. If the participant does not agree to the consent, they must automatically be exited from the questionnaire.
3. Each question must be **optional**.
4. The participant must have the option to 'not submit, exit with**out submitting**' at the final submission point on your questionnaire.
5. If you have open-ended questions on your questionnaire you must warn the participant against naming **third parties**: *"Please do not name third parties in any open text field of the questionnaire. Any such replies will be anonymised."*
6. You must inform your participants regarding **illicit activity**: *"In the extremely unlikely event that illicit activity is reported I will be obliged to report it to appropriate authorities."*

Sheila Kelly, Stud. #: 16306176
TCD MSc HI, Dissertation (2017-18)

Key Informants Information Sheet

<TBC>February 2018

Name <TBC>

Title <TBC>

Company/Address <TBC>

Re: MSc in Health Informatics - Key Informant Interview for Dissertation

Dear <TBC>

As mentioned in my Linkedin communication, I am currently studying for an MSc in Health Informatics at Trinity College Dublin (TCD), and, as part of my research, I have undertaken to complete a dissertation on:

Challenges and Enablers in the Collection of Health Data for use in Phase II-III Clinical Trials

As a key informant in this field, I would like to invite you to participate in an interview to answer a series of questions related to this topic, as outlined below:

1. Janet Woodcock, Director of FDA's Centre for Drug Evaluation and Research recently remarked that the clinical trial data collection process is broken. Do you agree with this statement, and what do you see as the barriers or issues that exist today in how data is collected in phase II-III clinical trials?
2. In your opinion, what factors are changing how health data is collected in phase II-III clinical trials? What are the advantages of such changes? What are the disadvantages of such changes?
3. In the longer term, what innovations may impact how data is collected for phase II-III clinical trials? What advantages will these innovations bring? What issues may these innovations fail to address?

I plan to conduct the interview over the phone or in person (subject to availability), and expect it to take approximately 30 minutes to complete. I will record your answers in writing and as electronic audio recordings.

The answers you provide will be used to form the basis of a key informant section of my dissertation and I may quote your responses directly in the body of my document. The final dissertation will be submitted to the examinations board at Trinity College and, if successful, will be posted on TCD's website for future reference.

Before submitting the final document for examination I will provide you with a copy of your answers for final review, and will provide you with the sections of my dissertation where you may be quoted. I will erase all electronic voice recordings at the end of my research and when the examination period has concluded.

In line with Trinity College Dublin's ethics guidelines I would like to emphasise the following points:

- Your participation is voluntary. There are neither anticipated risks nor benefits to you by participating in the research, and you may choose to participate in the research or to decline.
- If you choose to participate,
    - I must obtain written approval of your informed consent. Please sign the attached "informed consent" document.
    - You may withdraw from the research at any time
    - You may choose not to answer all questions posed
    - In the unlikely event illicit activity is reported to me during the study I will be obliged to report it to appropriate authorities

Note that I am conducting this research on my own behalf for the purposes of my studies at TCD, and not on behalf of my employer, ICON plc.

Please can you confirm by return email if you are willing to participate in the interview?

Sincerely

Sheila Kelly
kellys77@tcd.ie

Trinity College Dublin

Informed Consent Form

Researcher: Sheila Kelly

BACKGROUND OF RESEARCH: As part of an MSc program in Health Informatics at Trinity College Dublin (TCD), the researcher has undertaken to complete a dissertation on: **Challenges and Enablers in the Collection of Health Data for use in Phase II-III Clinical Trials**

PROCEDURES OF THIS STUDY: As part of the planned research, key informants will be interviewed over the phone or in person (subject to availability), and the answers provided will be used to form the basis of a key informant section of research dissertation.

PUBLICATION: The final dissertation will be submitted to the examinations board at Trinity College and, if successful, will be posted on TCD's website.

DECLARATION:

- I am 18 years or older and am competent to provide consent.
- I have read, or had read to me, a document providing information about this research and this consent form. I have had the opportunity to ask questions and all my questions have been answered to my satisfaction and understand the description of the research that is being provided to me.
- I agree that my data will be used for scientific purposes and I have no objection that my data may be published in scientific publications
- I am expressing my own views and not the views of my employer or other organization.
- I understand that if I make illicit activities known, these will be reported to appropriate authorities.
- I understand that I may stop electronic recordings at any time, and that I may at any time, even subsequent to my participation, have such recordings destroyed (except in situations such as above).
- I understand that, subject to the constraints above, no recordings will be replayed in any public forum or made available to any audience other than the current researchers/research team.
- I freely and voluntarily agree to be part of this research study, though without prejudice to my legal and ethical rights.
- I understand that I may refuse to answer any question and that I may withdraw at any time without penalty.
- I have received a copy of this agreement.

_____
Key Informant's Name

_____ _____
Key Informant's Signature Date:

_____ _____
Researcher's Contact Details Date:

3

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

MSc HI – Year 2
Research Proposal

Challenges and Enablers in the Collection of Health Data for use in Phase II-III Clinical Trials

Prepared for:

Lucy Hederman

15 January 2018

Prepared by:
Sheila Kelly
Stud #: 16306176
M: +353 87 6819492
E: kellys77@tcd.ie

Table of Contents

1. INTRODUCTION

Janet Woodcock, of the FDA's Centre for Drug Evaluation and Research, told a recent workshop at the National Academies of Sciences, Engineering, and Medicine that "the clinical trials system is "broken" and there needs to be new ways to collect and utilize patient data" (Woodcock, 2017)

The process of collecting data for clinical trials is disparate. In an average clinical trial, multiple systems are used to collect patients' health data. Often these systems are incompatible and are configured to optimize workflow for a particular operational department (Hublou, 2016). For example, in a typical laboratory system, patient data is gathered with the purpose of managing patient samples, while in an imaging system, patient data is gathered to manage image sharing and review.

Typically, pharmaceutical companies outsource the data collection process to contract research organisations (CROs), and in many cases only certain aspects of the clinical trial will be outsourced. So, for example, in a phase II or III trial, one CRO may manage the laboratory data, while another CRO may collect the main patient information or case record from (CRF), and the pharmaceutical company may manage some aspects of the clinical trial itself (Wang & Motti, 2015)

As part of the phase II-III clinical trial process, the case record form (CRF) is the central data collection document used to collate patient information. In the last 10-15 years, the CRF has now morphed into the eCRF, as typically this data is collected using electronic data capture (EDC) systems (Krishnankutty, et al., 2012). The CRF or eCRF charts the patient's progress though the clinical trial by using a predefined set of research criteria and assessments.

2. RESEARCH QUESTION

Title/Research Question: Challenges and Enablers in the Collection of Health Data for use in Phase II-III Clinical Trials

3. MOTIVATION & AIMS

3.1. Motivation

The motivation for this research topic came from direct experience of the challenges faced in the collection of health data for clinical research, while working in the industry.

Clinical research is under increasing pressure to reduce costs and timelines. Estimated costs to bring a new drug to market are in the region of $2.6bn (Mullin, 2014) over a 12-year period (Torjesen, 2015)

A preliminary literature search outlines the many challenges in the clinical data collection process (Richesson & Nadkarni, 2011) (Vadim Tantsyura, et al., 2007), and, additionally, highlights how advances in technology can transform this process by giving researchers direct access to pertinent data and potentially reducing timelines and costs (Bhavnani, et al., 2017).

Initiatives such as EHR4CR (EHR4CR, 2015), which is developing a new working model to extract data directly from EHRs for clinical research has the potential to transform the clinical research industry. To date, this model is not used widely, and patient data is still collected in separate systems solely designed for clinical research (Murphy, et al., 2007)

3.2. Aims

Using an evidence-based approach, the main aim of this research is to identify the potential barriers to, and enablers for, change to improve how data is collected for use in phase II-III clinical trials, with a focus on:

3.2.1.1. Outlining the current practice for data collection in phase II-III clinical trials with a focus on the collection of eCRF data

3.2.1.2. Identifying gaps or pain points in this process

3.2.1.3. Identifying new developments in the industry that are impacting this process, such as GDPR (Luveaux, 2016); overall technical advancements and how they can improve this process (Underwood, 2016); and a switch in thinking from historic experience towards evidence-based research (IQVIA, 2017)

3.2.1.4. Reviewing new ways of harnessing health care data for clinical trials, such as direct access to EHRs (Cowie, et al., 2017); the growing importance of health outcomes to guide clinical research (Williamson, et al., 2012); and developing clinical research as a treatment option (NIH, 2016).

4. APPROACH TO THE RESEARCH METHODOLOGY

To answer the research question, an outline of how data is managed today for clinical trials, will be presented, with a focus on new trends and emerging technologies that are shaping how health data is collected for clinical trials, using the following approach:

1. Review the literature related to health data collection for clinical trials in journals and industry publications, with a focus on current and new initiatives to collect health data. Compare these achievements to the current working practices.
2. Conduct interviews with key informants in the industry, such as senior managers of key clinical research data management providers, such as Medidata and Oracle; CIOs in the clinical research industry; and other key opinion leaders from pharma and/or academia.
3. Below I have provided a list of proposed key informants that I intend to contact through Linkedin, and ask them if they are willing to participate in the research. If they accept my invitation and agree to participate, I will provide them with the key informants' information sheet that I have prepared as part of the TCD Ethics application.

   List of Proposed Key Informants

   I propose to interview the following key informants:

   i. Bill Byrom, independent contractor and specialist in eClinical Products
   ii. Lars-Olaf "Lollo" Ericksson, independent contractor & specialist in Clinical Trial Feasibility
   iii. Dr Brendan Buckley, Prof of Medicine & Pharmacology at UCC
   iv. Dermot Kenny, Head of Clinical Data Management, Novartis
   v. Dipak Kalra, President of the European Institute for Health Records
   vi. Representative from Medidata, name *to be confirmed*
   vii. Representative from Veeva systems, name *to be confirmed*
   viii. Marie McCarthy, Director Wearables Innovation, ICON

This list is not conclusive and I may approach other industry leaders and ask them to participate in the research.

Note: Key informants i-iv (above) all worked for my current employer, ICON plc, at certain points in their careers. Key informant viii is currently employed by ICON plc, but I do not having any reporting line to this individual.

4. A proposed table of contents for the dissertation is attached in appendix I.

5. WORK COMPLETED TO DATE

To date, the following research activities are underway:Key informants have been identified.

2. Questions for key informants have been drafted and have been circulated to a small group of colleagues for input. The draft questions are, as follows:

   - Janet Woodcock, Director of FDA's Centre for Drug Evaluation and Research recently remarked that the clinical trial data collection process is broken. Do you agree with this statement, and what do you see as the barriers or issues that exist today in how data is collected in phase II-III clinical trials?

   - In your opinion, what factors are changing how health data is collected in phase II-III clinical trials? What are the advantages of such changes? What are the disadvantages of such changes?

   - In the longer term, what innovations may impact how data is collected for phase II-III clinical trials? What advantages will these innovations bring? What issues may these innovations fail to address?

3. Ongoing research of journal articles, web sites and industry publications

4. Attending supervisor meeting to finesse the research question and the focus of the research

6. APPROXIMATE TIMETABLE TO COMPLETE DISSERTATION

Working backwards from the due date of 26th June 2018, key milestones to complete the dissertation have been identified, as follows:

3.1. **November 2017 Milestones**
- 3.1.1. Finalise research question
- 3.1.2. Submit research proposal to supervisor and present it to class mates

3.2. **January 2018 Milestones**
- 3.2.1. Finalise literature review
- 3.2.2. Present progress report to academic staff
- 3.2.3. Finalise selection of key informants
- 3.2.4. Finalise key informants' questionnaire
- 3.2.5. Submit key informatics questionnaires to TCD ethics board
- 3.2.6. Meet with supervisor

3.3. **February 2018 Milestones**
- 3.3.1. Conduct key informants' interviews
- 3.3.2. Meet with supervisor

3.4. **March 2018 Milestones**
- 3.4.1. Final key informant responses & implement into dissertation
- 3.4.2. Prepare first formal draft of the dissertation
- 3.4.3. Meet with supervisor

3.5. **April 2018 Milestones**
- 3.5.1. Finalise dissertation
- 3.5.2. Meet with supervisor

3.6. **May 2018 Milestones**
- 3.6.1. Submit softcopy of dissertation to supervisor
- 3.6.2. Meet with supervisor
- 3.6.3. Send dissertation to print

3.7. **June 2018 Milestones**
- 3.7.1. Submit final dissertation (hardbound copies)

Figure 1 below outlines the planned timetable to complete the dissertation.

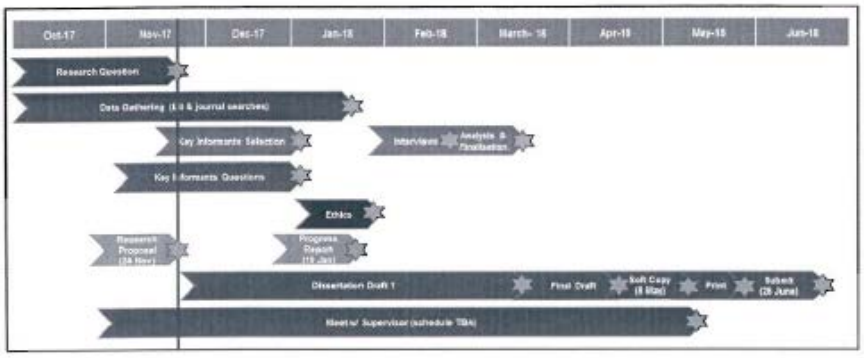


Figure 1: Proposed Timeline to Complete Dissertation

WORKS CITED & BIBLIOGRAPHY

Anderson, M., 2017. *Huffington Post.* [Online] Available at: https://www.huffingtonpost.com/margaret-anderson/top-ten-medical-research_b_13975834.html [Accessed 2017 November 2017].

Beresniak, A. et al., 2016. Cost-benefit assessment of using electronic health records data for clinical research versus current practices: Contribution of the Electronic Health Records for Clinical Research (EHR4CR) European Project.. *Contemp Clin Trials,* Jan(46), pp. 85-91.

Better Value Organisation, 2014. *Key Informant Interviews.* [Online] Available at: https://www.betterevaluation.org/evaluation-options/key_informant_interviews [Accessed 2017 November 22].

Bhavnani, S. et al., 2017. 2017 Roadmap for Innovation—ACC Health Policy Statement on Healthcare Transformation in the Era of Digital Health, Big Data, and Precision Health. *Journal of the American College of Cardiology,* 10(18), p. Unk.

Brennan, Z., 2017. *Endpoints News.* [Online] Available at: https://endpts.com/fdas-janet-woodcock-the-clinical-trials-system-is-broken/ [Accessed 22 November 2017].

CDC, 2016. *Centres for Disease Control & Prevention.* [Online] Available at: https://www.cdc.gov/copd/index.html [Accessed 15 May 2017].

Cowie, M. et al., 2017. Electronic health records to facilitate clinical research. *Springer,* 106(Unk), pp. 1-9.

Doods, J., Botteri, F., Dugas, M. & Fritz, F., 2014. A European inventory of common electronic health record data elements for clinical trial feasibility.. *PubMed,* Jan(10), p. Unk.

EHR4CR, 2015. *EHR4CR.* [Online] Available at: http://www.ehr4cr.eu/ [Accessed 23 November 2017].

Hublou, R., 2016. *Pharmaceutical Processing.* [Online] Available at: https://www.pharmpro.com/article/2016/09/biggest-challenges-delivering-clinical-trials-time-within-budget
[Accessed 22 November 2017].

Institute of Medicine (US) Forum on Drug Discovery, Development, and Translation, 2010. Challenges and Opportunities: Workshop Summary. In: *Transforming Clinical Research in the United States.* Washington DC: The National Academies, p. Unk.

IQVIA, 2017. *IQVIA Analyst and Investor Conference.* New York, IQVIA, p. 82.

Krishnankutty, B., Bellary, S., Kumar, N. & Moodahadu, L., 2012. Data management in clinical research: An overview. *India Journal of Pharmacology,* Mar-Apr(44 (2)), p. 168–172.

Lu, M. et al., 2014. Improving efficiency and reducing costs: Design of an Improving efficiency and reducing costs: Design of an adaptive, seamless, and enriched pragmatic efficacy trial of an online asthma management program.. *PubMed,* May(38), pp. 19-27.

Luveaux, S., 2016. *Big data and the new EU data protection legislation.* London, European Medicines Agency.

Mullin, R., 2014. *Scientific American.* [Online] Available at: https://www.scientificamerican.com/article/cost-to-develop-new-pharmaceutical-drug-now-exceeds-2-5b/
[Accessed 22 November 2017].

Murphy, E., Ferris, F. & O'Donnell, R., 2007. An Electronic Medical Records System for Clinical Research and the EMR–EDC Interface. *Investigative Opthalmology & Visual Services,* 48(10), pp. 4383-89.

NIH, 2016. *National Institutes of Health.* [Online] Available at: https://www.nih.gov/health-information/nih-clinical-research-trials-you/need-awareness-clinical-research
[Accessed 22 November 2017].

Richesson, R. & Nadkarni, P., 2011. Data standards for clinical research data collection forms: current status and challenges. *Jamia,* 107(UNK), p. UNK.

Torjesen, I., 2015. Drug development: the journey of a medicine from lab to shelf. *The Pharmaceutical Journal,* Unk(Unk), p. Unk.

Tufts, 2017. *Tufts Centre for the Study of Drug Development.* [Online] Available at: http://csdd.tufts.edu/ [Accessed 22 November 2017].

UCLA CENTER FOR HEALTH POLICY RESEARCH, UNK. *Key Informant Interviews.* [Online] Available at: http://healthpolicy.ucla.edu/programs/health-data/trainings/documents/tw_cba23.pdf [Accessed 22 November 2017].

Underwood, G., 2016. The clinical trial of the future. *Pharma Times*, Unk(Unk), p. Unk.

Vadim Tantsyura, V., Yuan, O. & Sirichenko, S., 2007. *Challenges and Opportunities.* New York, MIT.

Wang, L. & Motti, E., 2015. *Pharmaceutical Outsourcing.* [Online] Available at: http://www.pharmoutsourcing.com/Featured-Articles/174597-The-Increasing-Shift-of-Clinical-Trials-to-CROs/ [Accessed 22 November 2017].

Williamson, P. et al., 2012. Developing core outcome sets for clinical trials: issues to consider. *Biomed Central*, 13(UNK), p. 132.

Woodcock, J., 2017. *Workshop at the National Academies of Sciences, Engineering, and Medicine.* Kansas, Endpoints News.

Wood, W., Bennett, A. & Basch, E., 2015. Emerging uses of patient generated health data in clinical. *Molecular Oncology*, 9(UNK), pp. 1018 -1024.

APPENDIX I – PROPOSED TABLE OF CONTENTS (DISSERTATION)

Dissertation: Proposed Table of Contents
Chapter 1 Introduction
1.1 Introduction
1.2 Background
1.3 Research Question and Study Aims
1.4 Overview of the Research
1.5 Overview of the Dissertation
Chapter 2 Current Situation
2.1 Introduction
2.2 How the data collection works today in clinical trials
2.3 Pain points in the current process
2.4 New technologies than may enhance the data collection process for clinical trials
2.5 Conclusion to State of the Art
Chapter 3 Research Design / Methodology
3.1 Introduction
3.2 Outline of Literature Search
3.3 Selection of Key Informants
3.4 Key Informants' Questionnaire
3.5 Ethical Considerations
3.6 Conclusion
Chapter 4 Implementation and Results
4.1. Introduction
4.2 Analysis of Key Informants's Responses
4.3 Analysis of the Literature
4.4 Conclusion
Chapter 5 Conclusion and Future Work
5.1 Introduction
5.2 Strengths and Limitations of the Study
5.3 Dissemination of Findings
5.4 Implications for How Health Data is Managed in Clinical Trials
5.5 Recommendations for Future Research
5.6 Reflections of the Study
5.7 Conclusion
References
Bibliography
Appendices – to be decided – for example:
Appendix 1: List of Key Informants
Appendix 2: Key Informants questions and answers

12

APPENDIX II – KEY INFORMANTS' SIGNED CONSENT FORMS

Tigran Arzumanov's Signed Consent Form

Sheila Kelly, Stud # 16306176
TCD MSc HI, Dissertation (2017-18)

Trinity College Dublin

Informed Consent Form

Researcher: Sheila Kelly

BACKGROUND OF RESEARCH: As part of an MSc program in Health Informatics at Trinity College Dublin (TCD), the researcher has undertaken to complete a dissertation on: **Challenges and Enablers in the Collection of Health Data for use in Phase II-III Clinical Trials**

PROCEDURES OF THIS STUDY: As part of the planned research, key informants will be interviewed over the phone or in person (subject to availability), and the answers provided will be used to form the basis of a key informant section of research dissertation.

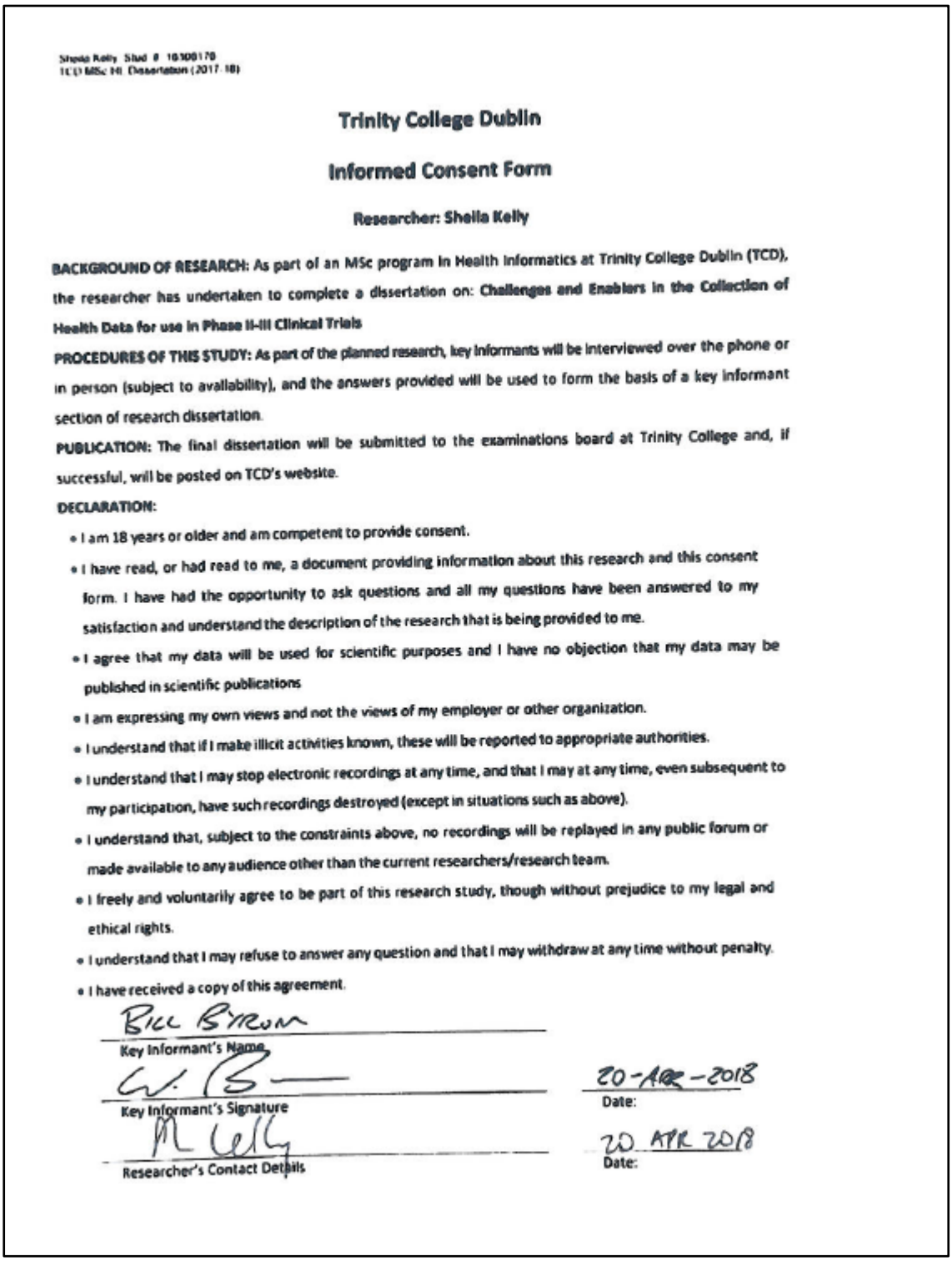
PUBLICATION: The final dissertation will be submitted to the examinations board at Trinity College and, if successful, will be posted on TCD's website.

DECLARATION:

- I am 18 years or older and am competent to provide consent.
- I have read, or had read to me, a document providing information about this research and this consent form. I have had the opportunity to ask questions and all my questions have been answered to my satisfaction and understand the description of the research that is being provided to me.
- I agree that my data will be used for scientific purposes and I have no objection that my data may be published in scientific publications
- I am expressing my own views and not the views of my employer or other organization.
- I understand that if I make illicit activities known, these will be reported to appropriate authorities.
- I understand that I may stop electronic recordings at any time, and that I may at any time, even subsequent to my participation, have such recordings destroyed (except in situations such as above).
- I understand that, subject to the constraints above, no recordings will be replayed in any public forum or made available to any audience other than the current researchers/research team.
- I freely and voluntarily agree to be part of this research study, though without prejudice to my legal and ethical rights.
- I understand that I may refuse to answer any question and that I may withdraw at any time without penalty.
- I have received a copy of this agreement.

Tigran Arzumanov	
Key Informant's Name	
Key Informant's Signature	23 Feb 2018
	Date:
Researcher's Contact Details	23/02/'18
	Date:

Bill Byrom's Signed Consent Form



Sheila Kelly Stud # 16308170
TCD MSc HI Dissertation (2017-18)

Trinity College Dublin

Informed Consent Form

Researcher: Sheila Kelly

BACKGROUND OF RESEARCH: As part of an MSc program in Health Informatics at Trinity College Dublin (TCD), the researcher has undertaken to complete a dissertation on: **Challenges and Enablers in the Collection of Health Data for use in Phase II-III Clinical Trials**

PROCEDURES OF THIS STUDY: As part of the planned research, key informants will be interviewed over the phone or in person (subject to availability), and the answers provided will be used to form the basis of a key informant section of research dissertation.

PUBLICATION: The final dissertation will be submitted to the examinations board at Trinity College and, if successful, will be posted on TCD's website.

DECLARATION:

- I am 18 years or older and am competent to provide consent.
- I have read, or had read to me, a document providing information about this research and this consent form. I have had the opportunity to ask questions and all my questions have been answered to my satisfaction and understand the description of the research that is being provided to me.
- I agree that my data will be used for scientific purposes and I have no objection that my data may be published in scientific publications
- I am expressing my own views and not the views of my employer or other organization.
- I understand that if I make illicit activities known, these will be reported to appropriate authorities.
- I understand that I may stop electronic recordings at any time, and that I may at any time, even subsequent to my participation, have such recordings destroyed (except in situations such as above).
- I understand that, subject to the constraints above, no recordings will be replayed in any public forum or made available to any audience other than the current researchers/research team.
- I freely and voluntarily agree to be part of this research study, though without prejudice to my legal and ethical rights.
- I understand that I may refuse to answer any question and that I may withdraw at any time without penalty.
- I have received a copy of this agreement.

Bill Byrom
Key Informant's Name

Key Informant's Signature — Date: 20-APR-2018

Researcher's Contact Details — Date: 20 APR 2018

Jeff Lee's Signed Consent Form

Sheila Kelly, Stud #: 16305178
TCD MSc HI, Dissertation (2017-18)

Trinity College Dublin

Informed Consent Form

Researcher: Sheila Kelly

BACKGROUND OF RESEARCH: As part of an MSc program in Health Informatics at Trinity College Dublin (TCD), the researcher has undertaken to complete a dissertation on: **Challenges and Enablers in the Collection of Health Data for use in Phase II-III Clinical Trials**

PROCEDURES OF THIS STUDY: As part of the planned research, key informants will be interviewed over the phone or in person (subject to availability), and the answers provided will be used to form the basis of a key informant section of research dissertation.

PUBLICATION: The final dissertation will be submitted to the examinations board at Trinity College and, if successful, will be posted on TCD's website.

DECLARATION:

- I am 18 years or older and am competent to provide consent.
- I have read, or had read to me, a document providing information about this research and this consent form. I have had the opportunity to ask questions and all my questions have been answered to my satisfaction and understand the description of the research that is being provided to me.
- I agree that my data will be used for scientific purposes and I have no objection that my data may be published in scientific publications
- I am expressing my own views and not the views of my employer or other organization.
- I understand that if I make illicit activities known, these will be reported to appropriate authorities.
- I understand that I may stop electronic recordings at any time, and that I may at any time, even subsequent to my participation, have such recordings destroyed (except in situations such as above).
- I understand that, subject to the constraints above, no recordings will be replayed in any public forum or made available to any audience other than the current researchers/research team.
- I freely and voluntarily agree to be part of this research study, though without prejudice to my legal and ethical rights.
- I understand that I may refuse to answer any question and that I may withdraw at any time without penalty.
- I have received a copy of this agreement.

Jeff Lee
Key Informant's Name

Key Informant's Signature — Date: 07Feb2018

Researcher's Contact Details — Date: 7 Feb 18

Marie McCarthy's Signed Consent Form

Sheila Kelly, Stud #: 16306176
TCD MSc HI, Dissertation (2017-18)

Trinity College Dublin

Informed Consent Form

Researcher: Sheila Kelly

BACKGROUND OF RESEARCH: As part of an MSc program in Health Informatics at Trinity College Dublin (TCD), the researcher has undertaken to complete a dissertation on: **Challenges and Enablers in the Collection of Health Data for use in Phase II-III Clinical Trials**

PROCEDURES OF THIS STUDY: As part of the planned research, key informants will be interviewed over the phone or in person (subject to availability), and the answers provided will be used to form the basis of a key informant section of research dissertation.

PUBLICATION: The final dissertation will be submitted to the examinations board at Trinity College and, if successful, will be posted on TCD's website.

DECLARATION:

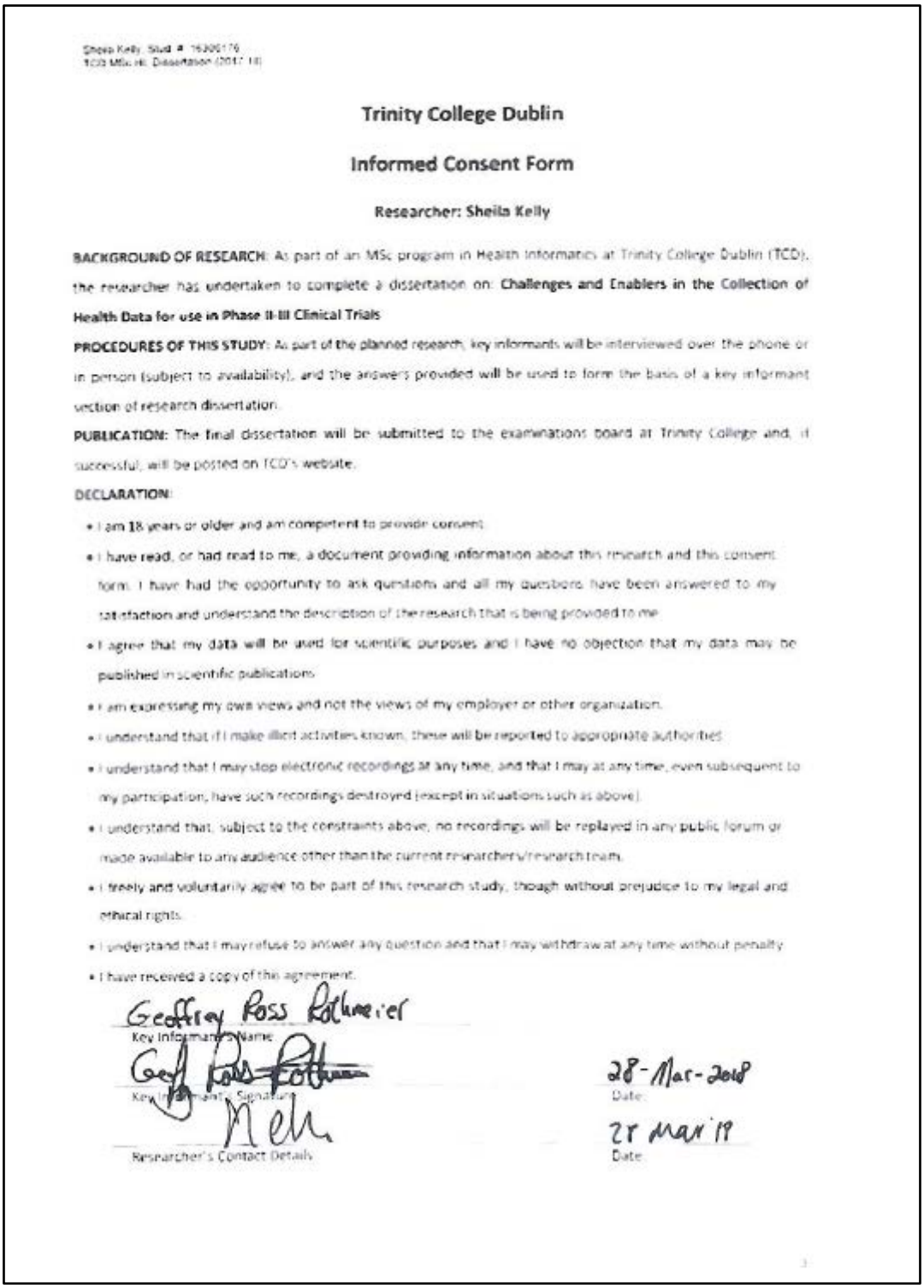
- I am 18 years or older and am competent to provide consent.
- I have read, or had read to me, a document providing information about this research and this consent form. I have had the opportunity to ask questions and all my questions have been answered to my satisfaction and understand the description of the research that is being provided to me.
- I agree that my data will be used for scientific purposes and I have no objection that my data may be published in scientific publications
- I am expressing my own views and not the views of my employer or other organization.
- I understand that if I make illicit activities known, these will be reported to appropriate authorities.
- I understand that I may stop electronic recordings at any time, and that I may at any time, even subsequent to my participation, have such recordings destroyed (except in situations such as above).
- I understand that, subject to the constraints above, no recordings will be replayed in any public forum or made available to any audience other than the current researchers/research team.
- I freely and voluntarily agree to be part of this research study, though without prejudice to my legal and ethical rights.
- I understand that I may refuse to answer any question and that I may withdraw at any time without penalty.
- I have received a copy of this agreement.

MARIE Mc CARTHY
Key Informant's Name

Key Informant's Signature — Date: 9 Feb 18

S Kelly
Researcher's Contact Details — Date: 9 Feb 18

Roth Rothmeier's Signed Consent Form



Sheila Kelly Stud # 16206176
TCD MSc HI Dissertation (2017-18)

Trinity College Dublin

Informed Consent Form

Researcher: Sheila Kelly

BACKGROUND OF RESEARCH: As part of an MSc program in Health Informatics at Trinity College Dublin (TCD), the researcher has undertaken to complete a dissertation on: **Challenges and Enablers in the Collection of Health Data for use in Phase II-III Clinical Trials**

PROCEDURES OF THIS STUDY: As part of the planned research, key informants will be interviewed over the phone or in person (subject to availability), and the answers provided will be used to form the basis of a key informant section of research dissertation.

PUBLICATION: The final dissertation will be submitted to the examinations board at Trinity College and, if successful, will be posted on TCD's website.

DECLARATION:

- I am 18 years or older and am competent to provide consent.
- I have read, or had read to me, a document providing information about this research and this consent form. I have had the opportunity to ask questions and all my questions have been answered to my satisfaction and understand the description of the research that is being provided to me.
- I agree that my data will be used for scientific purposes and I have no objection that my data may be published in scientific publications.
- I am expressing my own views and not the views of my employer or other organization.
- I understand that if I make illegal activities known, these will be reported to appropriate authorities.
- I understand that I may stop electronic recordings at any time, and that I may at any time, even subsequent to my participation, have such recordings destroyed (except in situations such as above).
- I understand that, subject to the constraints above, no recordings will be replayed in any public forum or made available to any audience other than the current researchers/research team.
- I freely and voluntarily agree to be part of this research study, though without prejudice to my legal and ethical rights.
- I understand that I may refuse to answer any question and that I may withdraw at any time without penalty.
- I have received a copy of this agreement.

Geoffrey Ross Rothmeier
Key Informant's Name

Key Informant's Signature

28-Mar-2018
Date

Researcher's Contact Details

28 Mar 19
Date

Louis Smith's Signed Consent Form

Sheila Kelly, Stud. #: 16306176
TCD MSc HI, Dissertation (2017-18)

Trinity College Dublin

Informed Consent Form

Researcher: Sheila Kelly

BACKGROUND OF RESEARCH: As part of an MSc program in Health Informatics at Trinity College Dublin (TCD), the researcher has undertaken to complete a dissertation on: **Challenges and Enablers in the Collection of Health Data for use in Phase II-III Clinical Trials**

PROCEDURES OF THIS STUDY: As part of the planned research, key informants will be interviewed over the phone or in person (subject to availability), and the answers provided will be used to form the basis of a key informant section of research dissertation.

PUBLICATION: The final dissertation will be submitted to the examinations board at Trinity College and, if successful, will be posted on TCD's website.

DECLARATION:

- I am 18 years or older and am competent to provide consent.
- I have read, or had read to me, a document providing information about this research and this consent form. I have had the opportunity to ask questions and all my questions have been answered to my satisfaction and understand the description of the research that is being provided to me.
- I agree that my data will be used for scientific purposes and I have no objection that my data may be published in scientific publications
- I am expressing my own views and not the views of my employer or other organization.
- I understand that if I make illicit activities known, these will be reported to appropriate authorities.
- I understand that I may stop electronic recordings at any time, and that I may at any time, even subsequent to my participation, have such recordings destroyed (except in situations such as above).
- I understand that, subject to the constraints above, no recordings will be replayed in any public forum or made available to any audience other than the current researchers/research team.
- I freely and voluntarily agree to be part of this research study, though without prejudice to my legal and ethical rights.
- I understand that I may refuse to answer any question and that I may withdraw at any time without penalty.
- I have received a copy of this agreement.

LOUIS SMITH
Key Informant's Name

Louis SMith
Key Informant's Signature

5 Feb 18
Date:

M Mc
Researcher's Contact Details

5 Feb 18
Date:

Richard Young's Signed Consent Form

Sheila Kelly, Stud. #: 16308176
TCD MSc HI. Dissertation (2017-18)

Trinity College Dublin

Informed Consent Form

Researcher: Sheila Kelly

BACKGROUND OF RESEARCH: As part of an MSc program in Health Informatics at Trinity College Dublin (TCD), the researcher has undertaken to complete a dissertation on: **Challenges and Enablers in the Collection of Health Data for use in Phase II-III Clinical Trials**

PROCEDURES OF THIS STUDY: As part of the planned research, key informants will be interviewed over the phone or in person (subject to availability), and the answers provided will be used to form the basis of a key informant section of research dissertation.

PUBLICATION: The final dissertation will be submitted to the examinations board at Trinity College and, if successful, will be posted on TCD's website.

DECLARATION:

- I am 18 years or older and am competent to provide consent.
- I have read, or had read to me, a document providing information about this research and this consent form. I have had the opportunity to ask questions and all my questions have been answered to my satisfaction and understand the description of the research that is being provided to me.
- I agree that my data will be used for scientific purposes and I have no objection that my data may be published in scientific publications
- I am expressing my own views and not the views of my employer or other organization.
- I understand that if I make illicit activities known, these will be reported to appropriate authorities.
- I understand that I may stop electronic recordings at any time, and that I may at any time, even subsequent to my participation, have such recordings destroyed (except in situations such as above).
- I understand that, subject to the constraints above, no recordings will be replayed in any public forum or made available to any audience other than the current researchers/research team.
- I freely and voluntarily agree to be part of this research study, though without prejudice to my legal and ethical rights.
- I understand that I may refuse to answer any question and that I may withdraw at any time without penalty.
- I have received a copy of this agreement.

RICHARD YOUNG
Key Informant's Name

Key Informant's Signature — Date: 05-FEB-2018

Researcher's Contact Details — Date: 5 Feb '18

APPENDIX III – KEY INFORMANTS' INTERVIEW TRANSCRIPTS

Interview with Tigran Arzumanov, VP, Clinerion

23 February 2018

SK: Reads Q1

TA: I wouldn't quite say that the process is broken. I would say that it is fractured. Broken means that it is completely not working which I think wouldn't be the right statement. I think it is working – it is just working very badly. To me, the biggest reason why this is happening is – well one of the reasons why this is happening – is the whole concept of source data verification. I think that the concept of source data verification was very useful when it came up, but it has been obsolete for some time now. And because our industry is very conservative – the way that things happen – concepts outlive their usefulness but then they stay – they continue to put breaks on the path of innovation. So, the whole concept that you have to record the information on some paper or some other media, and you have to record information on another media, and you have to manually compare the two, I think is has outlived its usefulness. Data has to be entered only once. Because by enshrining SDV processes, we are essentially making it compulsory to create data entry errors and look for them and this is the very part why the process is fractured – just an entrenchment of the SDV concept. Another thing is that data comes from multiple sources and the use of electronic health care records for data in clinical trials – everyone understands that this is needed but there are a lot of challenges to it. Some of the challenges are objective – they are there just because people cannot figure out how to do things, and some of the challenges are in perception – and perception is changing. It is gradually changing. But I am concerned now that I am jumping ahead to the other two questions.

SK: The questions sort of flow into one another. I don't mind that

TA: So, you kind of see the attitude of people towards data sharing is changing with time. Our company is the use of electronic patient health records for patient recruitment and for real world evidence, so we do not yet actively engage in the use of electronic healthcare records in clinical trials as a source of information for data for the clinical trial database, but this is something that we actually discuss with several partners, so, I think, eventually a project like this will come. And the multiple roadblocks are first technology – technology is a very objective roadblock here – the ability to create standardised data. First off, technology which is being used by hospitals or healthcare institutions is different – it creates problems by itself. Then coding conventions are different – and this is actually more serious than a

technology problem – this is a convention or a standard. And in some cases, it is relatively straightforward to map one standard to the other, and in some cases, it is very hard. And then the final roadblock is the attitude towards data sharing. This is probably the hardest one because it cannot be objectively solved. The only way to solve this on a large scale is time. Some things society is ready for and some things they are not ready for. And they the future is here – it is just not evenly distributed. There are some healthcare institutions that are very progressive in terms of data sharing is concerned, and some healthcare institutions that are not. And what is very interesting is – this is actually mirrored by the situation in pharma and there are also very conservative pharma people and that means that they are trying to design processes for data collection which could work 10 years ago and by the time by put them in practice, they would already be 10 years too late. So, they make assumptions on what doctors are like and their assumptions are extremely conservative, and in the end they end up with unworkable solutions. So, it is kind of like an absurd situation, where, in theory, everyone wants the same thing. So, both the doctors and the pharma people want to have an efficient process of putting data in a clinical database, reduce the number of errors and streamline the process and so on. And in the process, everyone is extremely worried about patient privacy, because all of the data issue concern patient privacy – but incidentally, not all of them. Some of them are also about transparency because hospitals are apprehensive of letting other people access their data because that would potentially put them on the spot with some of the statements they are making.

So, on the one hand they are concerned about transparency, and on the other hand they are concerned about privacy. And in our experience, concerns about privacy are bigger than transparency, so and in principle a pharma company does not want to breach privacy – they do not benefit from it. And hospitals do not want to breach privacy. Both sides are making things difficult by putting patient privacy at the top of the considerations. And there is nothing wrong with putting privacy at the very top, but they start to think about excessive, very difficult procedures or outright stop things because of privacy concerns. The worst thing we had – a hospital decided not to work with us because their lawyers said that working with us may give the perception of selling patient data. This to me is completely outrageous – a lawyer gives advice about perception, but this is the reality. Patient privacy is such a red flag it is actually stopping things, and things are stopped in the name of patient privacy in a situation where patient privacy is not really an issue, and what is being lost in translation here is that patient privacy is actually a balance. On the one hand there is complete patient privacy and on the other…complete patient privacy is linked with suffering and death. Ultimate patient privacy is when you pick up a patient on the street and you cannot access their patient record because it is locked to anybody other than their treating physician and

they die because of an allergy. A lot of opportunities are missed because of patient privacy, because for a patient to receive a treatment – information that a patient needs a treatment, needs to be written down somewhere and it is not happening that patients are suffering. It is very complex.

SK: I have looked at Sweden and other countries where they have looked at opt-out situations. There data is already used for research and it's up to the patient to opt-out and say no, I don't want it to be used, and I think those sort of methods may overcome that issue regarding privacy – because the patient starts to take control of their data.

TA: This is exactly right, because its fine when a patient makes a decision on how his or her data is being is used. It is not fine when somebody hurts a patient in the name of patient privacy without consulting a patient about it.

SK: That's one of the big disadvantages of trying to get this data out of hospitals and to use it for research purposes or for feasibility purposes, so it is possible today to take this data out but of course we have these standards issues and, institutions giving permission, but there are advantages. If we could get this data out, what else could we use it for? What else – what are the realms of possibility for which we could use this data (looking into the future), beyond feasibility? Are there other use cases, do you think?

TA: There are plenty – imagination is the limit. We have been talking to a few parties, and a few things we already do – feasibility and identification of patients, so we actually enable healthcare institutions to identify patients that could be suitable for a trial. Now another use case is real-world evidence. Usually this is aggregated, and the data is being used to answer questions and questions can be very diverse. What is the patient's route, what kind of treatments are being used for this specific disease, what is the prevalence, what is the load on healthcare institutions to justify another drug that may decrease the workload of healthcare institutions, and so on. And, then, in some cases we also work in making source data available for analysis. This, of course, has to be approved on a project by project basis but pharma companies could benefit a lot from hospital data in making themselves more efficient, so one project we are discussing with pharma companies is payment-by-result. So, a drug may cost a different amount of money whether it is used for indication 1 or indication 2. Now if things are left to hospital reporting, they are going to report in many places that the drug has been used for the cheaper indication, so the pharma company needs some objective data source to be able to justify or to provide data for the payment. In some cases, it may be outcome based, so payments are going to made dependant on how many patients recovered. Health insurance could benefit from having access to hospital data as they could precisely quantify the probability or the cost that they need to associate with the patient, as

they could have access to outcome data which is local to their specific geography. So, if the health insurance company has access to on-demand real-time data they could make better decisions and be more efficient. Then there are a lot of things that can be done already on the treatment part of things. So, decisions support systems, systems which give doctors hints on how the patient can be treated. Already, and it is happening in US, pharma companies provide discounts. So, when there is information that a certain patient needs a treatment, a voucher can immediately be issued by the system – a discount for this drug. Some of these things are – let's say – ethically ambiguous, but nevertheless, if a pharma company can issue a discount, and has the right to do this legally, and this decreases the financial burden for a patient, why not, if they can do that in an efficient way? So, you can persuade payers or decision-makers in the healthcare domain to make informed decisions by providing data that is being used to provide emotionally moving pictures. Let's say if you want to initiate a public campaign and you show a map that says 133 people have died today because of this issue, and 3 people have died in the city where you are – this goes against the cold logic of using data, but people are not rational beings and such data may actually move people into action. And dynamics of diseases, and so you can investigate things like epidemics of flu spreading pretty much in real time from one city to another, so lots and lots of things that can be done if health data is published.

SK: That's very interesting. Just going back to a couple of things you raised there about privacy and that being a barrier to having access to data. How do you propose to get access to data in hospitals? I am aware of what EHR4CR is trying to do in terms of going into the site and putting in some sort of warehouse or infrastructure that allows the data to be aggregated and then they only take out the queried data. Is your model something similar or how do you extract the data from the site?

TA: It is similar but there are some differences. The EHR4CR project relied on i2B2, i2B2 is open source software developed in the US, and they essentially have to install the i2B2 software and build their warehouse on top of it. We built our technology from the ground zero, completely by ourselves and we run an ETL process (extract transfer load) to get hospital data into our server which is based in the hospital. And the data is heavily indexed, and this allows us to make our searches very quickly. But the big issue with the i2B2 tool is originally that it was not supposed to include mechanisms for actual data transfer. i2B2 in itself has actually been built for manual data upload and this means that any process for automatic data uploads from the hospital is subject to a patchwork of solutions and a lot of cases there are no solutions and they are just manually uploaded from time to time. As a result, you end up with a network of hospitals where the timing between data refreshes is

very different and that makes the data ok to use for feasibility purposes but very hard to use for patient identification in any kind of consistent way.

SK: So, your product opens up a real-time solution – is that correct?

TA: So as standard our refresh rate is 24 hours, and this kind of opens up some indications like oncology where usual treatments are assigned within a week to 10 days. In specific cases we are able to open up a one-minute refresh rate. We don't do that by default as it puts an additional load on the hospital systems but if there is a specific trial that requires that we can enable that. And we are starting at our first hospital in France - 1 hour south of Paris, where they are actually going to have true real time access to hospital records based on a HL7 protocol, so there, there is no delay at all. Because we did not marry our system with an obsolete product that is no longer being developed, we have much more freedom in developing connectors to different data formats and eventually moving closer towards real time. And another thing we have which is unique to us, which we have patented, is a new technology for anonymised identification as we call it. One of the big problems in technology such as ours is pseudonym use. Pseudonym is like a fake ID, you take a patient, you assign them a fake ID, and you maintain the list that connects one to the other. So, in the US, a pseudonym is one of eighteen identifiers which have to be removed in order to be considered anonymous, or a statistician or a qualified individual has to certify the probability of identification is law by intended recipient using conventional means. In Europe, pseudonyms with the GDPR that's coming up are considered protected health information, and the punishment for leaking pseudonymised data is the same as leaking patient data which creates issues. So, we came up with a technology called anonymised identification and the essence of the technology is that data we get from the hospital has no identifiers of any type - none at all. And, then once we run our feasibility filter and have the record that we are interested in, we push them back to the hospital and we give the hospital a process for comparing this anonymised record with the source data. And you can imagine that a patient record is very quickly becoming unique – because it has time stamps and even if you have two identical people seeing two identical doctors, very quickly there's going to be some divergence between the data in the lab test. And also, for clinical trials use, and even if theoretically there is a negligible possibility that a comparison is going to use the same result, it will not matter, because both patients will be eligible. So, by doing this we are very able to increase our footprint quicker because hospitals are much more enthusiastic or relaxed in sharing anonymised data that they are about sharing data with any identifier, because a pseudonym is a one to one identifier – it is very dangerous. You can imagine a situation, when somebody looks at a pseudonym and then calls somebody at the hospital

and then says – do me a favour, look up, who is that, and all you need to do is have one person with access to the list and that's it. All the rest of the data points never identify a patient uniquely. So, in the US, where the attitude towards pseudonyms is more relaxed, it also is a major innovation because it makes the job of the person certifying data anonymised much easier. Because out of the 18 identifiers, the only 2 that are relevant to clinical research are time stamps and pseudonyms. So, removing a pseudonym as a factor already removes half of the controlled data identifiers, and these are the more dangerous ones because they depend on the size of the hospital. One of the parameters that this person will be looking at is how big the hospital is. With a pseudonym is does not matter how big the hospital is because they link to the patient on a one-on-one basis.

SK: That's very interesting in terms of how you capture the data. I think from my three questions you have probably answered a lot and you have given me lots of fodder for my dissertation. Thank you.

Interview with Bill Byrom

23 February 2018

SK: Reads Q1

BB: I think she actually said the clinical trial process is broken, but you are absolutely right she was talking about data, but she was talking more about we have very little real-world evidence as part of a drug submission. So it's all highly controlled, randomised clinical trial data, and I think her point there is that's great, and I don't think anybody's saying that we don't need that, but I think what they are saying is that needs to be supplemented with less controlled, more real world evidence type data to really understand the effectiveness of the drug as it would be used in the patient populations. So, I guess the controlled trials are very good at figuring out, you know, what size of signal are we getting, what are is the effects size, some aspects of the safety profile are and because we're able to control for all of the factors and randomise patients etc. we're able to study that in some detail. But that picture isn't necessarily telling the regulators everything they need to know about how a drug is used in practice. And, I think that they're struggling with and it kind of ties in with 21st Century Cures Act – the need for more real-world evidence data. It is a little tricky for me to kind of well how do we get that – because at the moment, we get real world evidence quite often when a drug is already on the market because it is easier then to collect it, and you can look at registries and you can go back and look at certain things. When it's not actually on the market, you are still having to do it in a fairly controlled manner so it's not quite the same as real world evidence, but I suppose she wants something getting a bit closer to that. Maybe what we'll see is that we'll continue to do phase II where we are learning about the effect size in patients and figuring out the optimal dose and that sort of stuff, but as we get into phase III, perhaps what she is looking for is some randomised controlled trials or pivotal trials to confirm things and perhaps some additional trials that look at less rigorously controlled, more relaxed, more real world things in which we are trying to assess the effectiveness of the drug.

SK: And do you think that some of that data might come from EHR or mobile health/mHealth data or sensor data?

BB: Pre-market, I'm not sure that the data can come from EHR that easily – no easier than we currently do in a clinical trial because we'd be putting someone into a study to receive a drug they never had before, so there would be no historic data in there, so the question is would it be – is it an advantage to use EHR or EDC, I personally think we should still

continue to use the tools that have. Now we might be collecting less data and less frequent data, but I don't think the toolset that we is – we can still use that. I am pretty worried about the quality of data we can pull from EHR. We hear some good case studies, but they are usually using a particular hospital or a particular installation of a particular EHR system. When we want to do something on a more global nature – we're going to have multiple EHR systems to try to pull data out, all of which have different ways of recording stuff, a lot of unstructured data. I just feel we're a long way off from that being really useful. So, my thinking is, if it is pre-market approval, we could still be using the portal solutions, the patient reported outcomes solutions and very light-touch EDC systems that we currently use in late phase – we could be using them here - I don't see why not.

SK: Reads Q2

BB: The advantages are we have some really quite robust technology. Some really quite highly robust products with a lot of functionality which enables us to do some quite sophisticated things around the data cleaning process, etc. – with EDC, for example. But, I think as we go into the more real-world evidence studies – these tend to be much bigger in terms of sample sizes. They tend to require, perhaps, less intensive data collection, and so the tools that we currently use are a little bit of a sledgehammer to crack a nut, when it comes to operating to those types of trials. Because, again, with the real-world evidence, because of the size and number of centres that you include, you are often pulling doctors and researchers who are less familiar with the way we do phase II & III trials. They may be absolutely research-naïve in a sense, in that they haven't done this sort of stuff before, so to give them a full-blown instance of (named) EDC or something else to collect their data on, would just be a huge barrier to them. It's just that they are so complex. If you are a professional clinical trialist, you are kind of used to it. Some systems that are a lot simpler and are specifically aimed at late phase or registry type studies, and I think that those are the sorts of systems we would start to use for these larger RWD studies.

SK: And you think of the site burden today. If you have to log onto a number of portals or you have a number of workflows – do you see any enhancements that could be made there?

BB: There are these initiatives where investigator IDs and stuff (Exostar) - I don't know really. For the professional clinical trialist they would subscribe to that and that would be straightforward. I don't know is the answer.

SK: That's fair enough. If there was a solution there it would have been implemented by now, so I can understand where you are coming from. I guess if you look at the patient collected data/ePRO data – that to me still seems like a cumbersome process - the patient

having to break a cycle of care to in order to enter that data. Do you think there's any room for improvement there? What do you think is changing the way that is developing?

BB: So, when you say breaking the cycle of care, what do you mean?

SK: Well I just think that if I'm a patient taking medication for example for a condition where I need to track going to the bathroom rather than me just living my day and going through when I'm filling out an ePRO I actually have to record everything I'm doing, so I have to sort of stop my day, for moments at a time. I am wondering if there are easier ways to do that.

BB: Maybe, but I think patient reported data is seen to be so important, and increasingly important not just so in the phase II & III, but also in the real-world evidence piece – it's got a huge component in there. So, the question is how do you do that the most convenient manner. Obviously, bring your own device will make things a little more convenient, but the point you are making is around is sometimes the frequency in which we are asking for stuff is a little intrusive and it probably does interrupt their what would be considered routine care. What we sometimes see in late phase, we might see asking the patient to complete things on a much less frequent basis would be one approach. I personally think this is becoming increasingly important, so I don't think we'll see any less of this. We'll only see more.

SK: And I think also on my experience looking at ePRO collection, sometimes the patient doesn't get any value from it. So, the patient is giving their data, it is sent to a central repository somewhere and they never see that data again. They don't really understand the value of what they are providing. Do you think there is room for enhancing that experience – the patient getting some sort of metrics back or realising that their data is meaningful and for a greater cause?

BB: I'd love that to be the case and you are hitting on something quite important and that is what sort of feedback can we provide to the patient that is useful to them so that we will continue to engage them and it worth them using this app for six months or whatever it is during the study, but it doesn't start to become part of the intervention itself though it's getting the balance between giving them something that would be useful towards giving them something that is actually helping them to manage their condition, perhaps more than you want to as actually what you are looking for is what is the effect of the drug on their condition. So, an example would be – you give people a smartphone, and it collects their activity data as well as maybe asking them some questions from time to time, and as they can go into that and look at a dashboard and see how many steps they are doing every day, and they might start to set themselves targets to actually improve the number of steps just because the fact that they are using this app and getting the feedback. And that's great for them but how does that effect the results of the study where actually what we're quite

interested in, is knowing well the fact that you have now taken this drug is it helping you to becoming more active. Well having the app and having the feedback every day is helping them to be more active as well, so you see you can't always disentangle the effects. That's the difficulty so it's finding something that does offer them value and is useful to them and would be motivational to continue using but that doesn't become part of the intervention, so it doesn't confound the drug affects you are trying to measure.

SK: And another aspect of that whole piece – what might be of value to the patient – their data sends an alert, or somebody responds to their data in real time or close to real time, so that if they have an exacerbation or there is some information entered in by the patient that would point to them needing immediate care that there's someone at the receiving end that can actually respond.

BB: Yes, that's really good. That's the sort of thing that would be valuable for them. The sort of things you probably don't want to be doing is reminding them to take their tablets and that sort of stuff, as we're trying to do something that's a bit more real world, so if there are reasons why they are not taking the tablets so that needs to continue to happen – you can actually see what the true effectiveness of the drug is in the kind of real world conditions. But I like that from a safety perspective. So, if you are hitting a threshold where you perhaps need to come in for some kind of more care then that's a good idea.

Q3 – SK reads out.

BB: We'll still be talking about it, won't we? Our industry moves so slow. When does the 21st Century Cures Act become implemented, well that's a few years away isn't it, but it's within that time period. I don't know – I kind of feel that there's going to be so much debate here about what you can practically do in phase II & III, and why and whether randomised controlled trials are still vitally important, which I believe they are. I don't believe we should ever reduce them, it's just how we collect this other data in that setting where we don't have a marketed drug, so we're still having to do it properly in a controlled trial setting, in a sense. I don't know is the answer.

SK: But what about, people talk about the site-less trial – do you think that has any credence?

BB: Probably not. I don't think it would be completely site less. I don't think that's the point. I think some of the models that we see like the Apple Research Kit studies, are kind of good models to collect data from a very large group of people in a fairly controlled manner and that's kind of an attractive model. The fact that we'd still be doing this on non-marketed drugs means that patients actually are going to have to come to site to consent and get into

the study. They may have to come to site to receive their medication because it's not off the shelf or prescribed. So, there is going to be restrictions like that. I suppose some of those virtual trial tools will become useful to try to implement this on a bigger scale. And maybe the way it might start to work is around – so you have like the Pfizer remote study, which was like one single centralised site for the whole of America and that was how that was set to work. I kind of feel there might be more regional oversights. If you've got a lot of research-naïve doctors that are part of this trial, collecting simple data on their patients, maybe within a region you have kind of regional PI who is the person who kind of is overseeing the safety aspects and some of the wider aspects of the trial. Somebody who is an experienced researcher, perhaps just a routine care physician and maybe it's more like a hub and spoke model that a completely decentralised study.

SK: I was talking to someone the other day about the Apple Research Kit and their opinion was that it wasn't really working and that there was very little uptake. So a lot of people had downloaded it but very few were using it. Have you heard anything to that effect?

BB: Well I think that is the problem with apps in general in mHealth. People think they are cool, they download them, they use them for a little bit and then they get disengaged with them. And I think the drop-out rates are pretty high, but I don't know if it matters. If you are doing a huge study and you get a few thousand people who have entered enough data – that's probably enough for that piece of research, but maybe it's back to your point earlier in terms of - it's all very well collecting the data but unless it's engaging in a different way for the patient, they probably won't carry on using it. So, if it's not providing value or reward in some way then perhaps people won't continue to use it.

SK: And what do you think about these personalised health records – do you think they have any value in terms of people collating their own health record?

BB: It's an interesting one, isn't it? I kind of wonder who would do that. I don't seem to have much inclination myself. How's the uptake of that going, and who are these people who are really going to do it.

Interview with Jeff Lee

CEO, mProve Health

22 February 2018

Q1: SK reads question

JL: In 2010 I was a newcomer to the clinical research field, so bring some fresh eyes to these topics, and our focus has been at the intersection of mobile and clinical trials – so that's meant patient recruitment, patient engagement, patient compliance and some patient reported outcomes. I want to be really clear about who I am, and who I am not. Definitely, I am an active follower of clinical trials but I would not consider myself a data management specialist...not an EDC or CTMS company, nor are we in the EHR/EMR space.

My take on Janet's comments, clearly there seems like there is an open lie about the effectiveness of the clinical trial model because the sample sizes are constrained by economics and time, and yet the manifestation of symptoms and results, effect efficacy measures doesn't always occur within that timeframe. Pretty much we are not measuring outcomes in big enough populations for long enough – that's why we are seeing so much emphasis on real world evidence where that has been collected in a more formalised way. What I see, is that your projects reach phase I, phase II, phase III and then they're approved, or they are approved with a REMS program, where's there's additional attention post approval. The formality of the data assessment after approval, other than in a case like REMS, is really about AEs. And that's, of course, important, but I don't think that's the only way to measure efficacy or safety, really, and the necessary level of detail to have the optimal understanding of how these therapies work.

SK reads question 2

JL: What originally inspired my interest in clinical research and the opportunities here at the intersection of mobile and clinical trials was what we called at the time "machine-to-machine" communication, and now people think of it as the internet of things. Ultimately, it's connected sensor-based devices. Back when I started my company, if I could have I would have started on that, but that was so far ahead of the market in 2010 that we just thought it was bad timing. Now there's a lot of interest in this, so I would say that if you abstract the topic, fundamentally, technology should be making it easier to collect more information from patients with the same or less effort. And whether that's bring your own device or patients to provide PRO surveys, data mining, whether that's wearables-based devices. So, staying at the abstract level, we should be expecting more out of our studies. Yes, of course protocols

are getting more complicated, there's more burden on sites, burden on patients, so there's a lot of healthy caution about further loading up our protocols, but what gets me excited about the opportunity is that we should be able to learn more, more quickly and I think that mean in terms of advantages, it can mean that programs get killed earlier. Whether it's because of safety or efficacy signals or just simply because you recognise that you're doing market research collection in phase II and recognise that you may have safety, but you don't have enough comparative effectiveness to have a good market access strategy. Wouldn't it be better to learn that at phase II rather than at phase III? Spending all that money and redirecting those resources to other compounds that have promise. We are excited about the prospect of doing better data collection having generated more insights. And then, of course, the overall changing paradigm and decentralised studies, it is not so much a data management topic, it is more of an access topic. There's a really interesting company who you may be familiar with called Transparency Life sciences that have been, the concept of crowdsourcing clinical research, almost like a kick-starter for clinical trials – getting people who will join a study. You have got a pre-built audience of patients and bringing the study to them using telehealth. And the premise of Transparency I find fascinating. Their idea is that any given sponsor has some backlog of compounds that they think they have market value and patient value, but they can't justify bringing them through a development process because it costs too much money. If it costs $500m to bring a drug to market, maybe the market potential for such a product is only $400m, and they just don't bother, whereas, by offering them a leaner more virtualised study model, maybe the cost of development goes down to $250m, and now you're got a whole layer of therapies that hadn't been commercially-viable to develop but are now. So, I think, that's another key advantage or opportunity given how things are starting to change in phase II-III.

SK: You talk about the protocol becoming more complicated and the issue you have is the protocol varies from trial to trial, and each pharma company has their own standard. Do you think technology will help to standardise that data collection?

JL: I think it's a mixed bag in the near term. I think that there some data collection methods that are standardised that can just go away. Everyone points to the six-minute walk test as just being a very outdated, coarse way of understanding in mobility and factors related to mobility. That ought to just go away with an activity meter. That said, for every one of those cases where there's a need for simplification, there's a dozen or more where there's new novel endpoints – there's new measurements. The way we are using wearables sensors is absolutely not standardised today. There are efforts to standard the data structure of it, but the types of measures, the types of data, how we are valuing it, how we think about the

algorithms. All that stuff is new, so it's hard, it's a difficult time to harness this technology as it does make life more complicating while developing new opportunities.

Question 3

SK: It's one thing collecting huge volumes of data, but if you are seeing any message or it's not pointing to one particular problem, what's the point in collecting it. So that touches on longer term innovations. Can you think of any others and challenges that may be forthcoming in terms of how we are looking at data today?

JL: One topic comes to mind. I am not sure if this is the best, most relevant topic, but it comes to mind in our discussion which I would think of as a false start with research kit. And the idea that research kit offered was very exciting. I almost think of it as democratising data collection because the way it's done currently other than the very gorilla academic level survey monkey kind of stuff, it's not validated; it's not particularly regulatory-grade. Your only options start to be very expensive companies and stifles the ability to do a lot of research. And so, the opportunity to create developer tools can make it easy to collect more information from patients is very exciting, and yet it was a false start. Because it is only IOS. It is only for smartphones owners, which we all knew, even if you added – if Google had come up with some non-competitive mirrored …offering, it still would have been smartphone users only. It would have been a subset of the patient population. However, that gives us a glimpse of the future, where that's how we're going be getting real world evidence – is collecting from information, whether it's passively or actively provided. Whether it's subjective or objective from a population of individuals and I think that it's tools like research kit – it's the first generation of tools that can make it easier to do that. Have the data collection in a faster and less expensive way.

SK: All of this data we are collecting – where are we going to put it?

JL: I wish I had better insights as I have had that same question. The generation of EDC providers is highly vulnerable to disruption because their data structure is organised around data coming in, in batches. Like CRF-style batches. And that's not how it works with sensor devices. I have given (named) EDC provider a lot of credit for being at the forefront of this new data collection model and paradigm because I think it is disruptive to their business, but I'm not close enough to understand a lot of the facets. I hear people talk about it's just terabytes of data – data that's so granular that you can't really understand it. How do you understand exceptions to the data or are you just talking about patterns and averages – that's stuff that I'm interested in, but not super-experienced in.

SK: So, in the longer term what changes do you think are coming in the next 5-10 years?

JL: I do hope to see a few things happen – I hope to see a better grasp of how to leverage sensor-based data collection – that's every part of it – developing novel endpoints and accepting the concepts of ALCOA in software validity applied differently to wearable devices – not applying outdated paradigms to new opportunities. I think that will allow things like wearable devices, research kit, passive data collection to be more embraced.

Another thing we haven't talked about which I feel conversation would be remiss to exclude is the role of EHR in our data collection. It is my understanding that the 21st Century Cures Act – if you look at that as well as the Affordable Care Act here in the US, they had stages of mandates. So, in the beginning, if you look at the Affordable Care Act – you've got to get on electronic records – you have got to get off paper. And they subsidised doctors' offices to be fully electronic by a certain date. That's how they knew they would move everybody over to electronic. And they called it meaningful use, and then there was next stages in meaningful use, and it's evolving. There's a latter stage – I don't know when it's supposed to kick-in of meaningful use, where the data is being collected in a manner it can be used for research – not just as a record of patient care but more for research purposes. And then at some point then the categories of EDC and EHR, theoretically they start to blend. In a five to 10 year horizon, I would expect to see changes there.

SK: What about the patient in all this? We have all of this technology. We can take your data every which way, but a lot of these technologies involve the patient doing something – disrupting their day, especially if they have quite complicated diseases or conditions. What do you think is going to be the solution there?

JL: It's a good question and I think it's very under studied. The ePRO consortium commissioned a study to see how much research existed on the concept of patient burden as it relate to data collection in research studies, and they came up virtually empty. That's a sign that no-one has really has taken a careful look at this. And I think what we will see is that there is burden. We'll have a truer picture of burden that we've ever had before. Whatever we thought we were burdening the patient with on paper, we were all kidding ourselves – all the backfilling/front filling – it was all just a big joke. So, at least we are getting to the point now that we tell if they actually did it – we can ask them how burdensome it was. We can get a benchmark that we either dial up or down. The ePRO consortium – their goal of course is to figure out can we do more data collection – when does it become too burdensome. I will give you one other example – I offer the premise that technology can make data collection easier, so we should be able to collect more, and I will be working on ways to make it simpler for the patients – more intuitive, convenient, quick, passive – where

possible. But at some point, there's no waving a magic wand – there's going to be a burden on the patient and I gave the example of the research kit example as something that's informed our thinking because all of these research kit apps were very exciting but they all failed to have meaningful persistence. They grabbed headlines by having tens of thousands of people download them and begin using them which suggests such a huge opportunity, but I was at conference a couple of years ago and 4 out of the 5 launched research kit apps providers (different academic institutions) were all present and they all confessed their usage levels and they were down to dozens or hundreds out of tens of thousands. The mistake that they made was they created and imbalanced proposition for the patient. It was simply something that you had to put time into and it didn't really give you any value back. It gave you maybe some trends – that's helpful – you could see your scores over time, but it didn't really give you anything else. It's appealing to the quantified self – a sliver of patients, but I think that's a tiny, tiny sliver, so we've got to find ways to make this more appealing to regular patients. So, one of the things we've done is, coming at this from a patient engagement perspective – our diary offering is very balanced between the diary and the survey, and only the features that are of value to the patient – helping the patient with their visit schedules, what to bring, what to expect, downloadable reference materials. Everything from a study id card to gamification to videos of how to do administer an injection – basically communicating to them in a way that gives them a lot of value, support, encouragement, clarity and empowerment in their program. And, also, that's the same app/vehicle that asks for their diary information. And, we have never scientifically measured this, and I'd think to that giving the patient something that's more valuable to them we're making it more of a balanced proposition – yes, we're asking you to give us diary information, but we are going to give a lot back as well. That notion of balance and engagement is pretty much all together missed in clinical trials. In real world, insurers want to have you track your meds and symptoms and all of that, and they are connecting with communities of people with whom you can share information and there are opportunities to get more value out of being part of data collection. Patients-like-me is a good example of that. But in clinical trials it's more restricted for obvious reasons so we have to look for other ways to make it more balanced and meaningful for the patient.

Transcript of Interview with Marie McCarthy (9 Feb 2018)

1. **Janet Woodcock, Director of FDA's Centre for Drug Evaluation and Research recently remarked that the clinical trial data collection process is broken. Do you agree with this statement, and what do you see as the barriers or issues that exist today in how data is collected in phase II-III clinical trials?**

   Some of these statements try to provoke a response or generate discussion.

   Clinical trials support a hypothesis – if you are not getting the right data or consistent – it is difficult to prove that the drug works – in Alzheimer's there has been no drug for 15 years.

   I went to a conference recently, and in one panel discussion, they talked about the patient's participation in a trial, and typically they are on the trial for 12-18 months – 9,000 hours on the trial, but the study only collects about 50 hours of data.

   The data collected is very subjective data - is based on patient recall or clinical observation. It isn't of sufficient quality. Parkinson's depends on the patient's symptoms which are episodic. If you had a continuous stream of data... For pain, the FDA, for analgesics, suggests that drugs may have been rejected because the methodology of identifying the change was poor.

   In Alzheimer's' – they can't identify a homogenous population. The trial data goes across recruitment as well as the response (of the patient)

   Using continuous data - the crossover from clinical research to healthcare

   .

2. **In your opinion, what factors are changing how health data is collected in phase II-III clinical trials? What are the advantages of such changes? What are the disadvantages of such changes?**

   - Continuous data
   - mHealth

   Patient centricity – a lot of endpoints linked to clinical outcomes. One endpoint is for a 6 min walk, but patients don't care about this.

For Duchene's, change in gait is another clinical endpoint but for parents of patients the loss of arm movement is of greater concern because children loose greater independence

Pharma wrong – they need to talk to parents

mHealth is starting to identify what is important to the patient – objective data is not enough. Combining objective and subjective data is critical.

Sensor suite combinations for COPD and asthma – there are so many pieces of tech for the patient – pharma is losing insight that the patient must deal with this

Huge benefits in capturing data remotely – and isolating the patient

So, few trials – not 100% that it's going to work. Signals not yet worked out, and there are privacy issues

In one paper, in a Parkinson's app. 7,000 patients enrolled within 6 hours…which shows the value of mobile solutions as the highest number of patients ever recruited using traditional methods was 1,700 (I will send on the paper for your review)

There is a change in recruitment to engagement, but there are issues related to blinding

Gamification can be used to encourage patients to stay on the trial

3. **In the longer term, what innovations may impact how data is collected for phase II-III clinical trials? What advantages will these innovations bring? What issues may these innovations fail to address?**

New endpoints will give pharma and trial designer's new trial designs

Digitalised streams and more analytics platform will create more automated correlations and adaptive designs

Trials will be prolonged. Real world data will become more important – it will be no longer about the regulators, but the payer, and the cost of the drug will drive change and the real move to use wearables to show quality of life against the existing drug.

I am beginning to see change already – pharma is trying to show trending rather than to prove an endpoint and to show compliance with these devices.

Wearables will be more seamless – technology developments such as smart home, voice recognition and other potentially could be used

For example, patients could show Fitbit data they've already collected

In Radar CNS – depression studies – the patient's engagement in social platforms may assess how well they are. Patients may be recruited into a study bringing your existing digital signature…

In 2014, when I first started reviewing clinical trial protocols from pharma companies, there was little interest in wearables. Suddenly wearables, in last 6 months, I am having amazing conversations with sponsors who are looking to combine a significant mHealth element into their trial design, and sensor suites that I may have been afraid to suggest, because they were too outlandish, are now acceptable.

There are still privacy or ethical concerns. I attended a conference last year, and an insurance company was able to track a person, where they were located, that they were a smoker (because they left their building for 10 minutes regularly). This big brother effect is frightening.

I met a man recently who has an RFID implanted in his arm – and there are a group of 20,000 in Sweden who have done this. Personally, this is a step too far, but future generations may be more accepting.

It depends on the purpose to which you use this data…

The value of wearables…some companies are already offering discounts, if you wear a Fitbit.

Change is evitable.

Interview with Ross Rothmeier

Vice President, Technology Solutions & Innovation Labs (Medidata Solutions)

23 March 2018

SK: Reads Q1

RR: I have heard this phrase quite a bit both from both sponsors and from CROs. It is a harsh statement to make, but I think it has its roots in the fact that no matter how well we think we've planned our trial, we find that there are changes we need to make along the way; they **take longer** than we had expected, and they **cost** more than we had hoped. So, in that context, **I can definitely see** where one would say that the clinical trial process or the data collection **process is broken**. Because if it were working perfectly it would be **efficient, lower cost and higher quality** than it currently is. And if you measure that cost and quality by **how we planned it**, then you can see why someone would say why it's broken. In my opinion, the idea of **collecting data to prove a hypothesis** isn't necessarily what we're challenging here. I think it's still important to prove that drugs are safe & effective and to do so in a controlled and scientifically rigorous and responsible fashion. But we can certainly do a lot to improve it. And those things include using historical data to help both forecast and target more efficient ways of doing things; to plan our work in such a way that we have a higher degree of probability of success in terms of recruiting subjects, collecting data as expected and reviewing it for any anomalies that we have to respond to. So, in that regard I **agree with the statement** and the barriers primarily are not really technological although there certainly is an element of technology. I think a lot of it has to do with letting go of the things we have done traditionally, and I can name a few of those, but for the most part I think our process and our cultural adoption of new ways of doing things is probably the hardest part of fixing or improving the clinical trial data collection process:

SK: And do you think that's driven by the regulators or is that just the data managers themselves or perhaps the technology companies...?

RR: I would lay it as any one constituent's feet. The **regulators** are certainly supporting this by **coming out with guidance** and ways to do things and new opinions, but at the same time, the **regulators are a bit of an adversary**, and that the reason the industry is so **conservative**, in my opinion, is the consequence of doing something wrong is so severe. You get a **483** at a **CRO** and you now have both weakened your competitive position and potentially put a **submission** at risk. You get one as a **sponsor** and you definitely put your submission at risk. So, without intending to, the industry has been conservative because

they are afraid of the consequences of doing something wrong, but they don't know it's wrong until you get an inspector saying so. And, of course, it's different by country, as we all know, the EMEA, the PMDA, the CFDA, US FDA and so on and so forth - all have different prioritisations and things that they are looking at. And I have been inspected by regulators in five countries and I prepared as though I was going to the worst-case scenario in each case, and they are very different experiences without judging or promoting one over another. You just have to be prepared for any contingency, which, I think is what holds us back from being really transformational. Though culturally, Data Managers, CRAs, Clinical Trial Scientists, Statisticians all fall into that place at some point that say well I can't justify not doing something that I have been doing for years because the possibility exists that we'll get an inspection that will require that artefact or that proof point. Regulators themselves – if you ask four regulators something, you'll get five different opinions...or something like that. They are not obligated to be consistent even within a single organisation. So, again, it begs conservative process management.

Q2: SK reads

RR: If I were put it into two words – data volume. We have so more access that we used to. I have been talking to people throughout the industry and within my own company about the extraordinary flood of data that we are going to have to contend with – in terms of sensor data, in terms of newly generated data that we have easier access to such as genomic data, for example, image data, data from various medical devices, let alone wearable health devices, and their variability. Cisco Systems publishes a paper or a prediction online that points this out very vividly. At the current rate, the amount of data generated on the internet by wearable devices will exceed 335 petabytes of data by the year 2020 – two years from now roughly. A petabyte is 1,000 gigabytes. So, if you think about 30035 hundred thousand gigabytes of data in a single month being generated. The sum total of all of the clinical trials that Medidata, arguably one of, if not the largest repositories of clinical data in the industry, to date, over 15 years has accumulated in the neighbourhood of two to two and a half petabytes of data in 15 years. So, the factors that is changing it is the amount of data and then, of course, the impact that is using that data. You cannot expect a data manager or CRA to do the processes that they are used to with the data volumes of yesterday with those kinds of data volumes. So, it pushes us to look at the data more programmatically, specifically using machine learning, artificial intelligence – you can't write an edit check that covers 335 petabytes worth of data so you're going to have to have something looking through the data to whether it's changing or in a way that requires or begs assertion. If the data is digital to begin with, the concept of electronic source challenges the concept of – and

Transcelerate has been very clear about this – data review versus data verification – different processes and with **esource** it becomes an even more obvious answer – you can't review or verify digital data. It is in itself verified and reviewed. The advantage of all of this is – imagine the value that we can derive if we actually do adopt and start looking at this data more effectively. We'll have potentially fewer patients that have to be poked or in some other way intervened with to figure out safety and efficacy of a compound. We'll also be able to reflect back on the data we have collected to more effectively make decisions about which molecules or what trial design or even which therapeutic areas make sense for us to pursue. And using things like genomics data and historical data we can even target certain disease types, and come up with solutions or treatments that have a higher probability and even a demonstrated probability of success which we can even synthesise that and model it. And that of course has downstream benefits to insurance companies or even individual payers. If I'm going to pay a lot of money for this treatment I want to know that it's going to work, and I have data to back that up. And while we used to do that back in the 1980s and 1990s with the clinical trial data we were collecting, if you have 2 or 3,000 subjects on a clinical trial and you go to market and you start selling it to many millions of people, you don't have a high probability or I should say you don't have a lot of assurance that all of those millions of people are going to fall into the few thousand you did those clinical trials on. So, I would say that there are so many opportunities and advantages. The disadvantages like I said are tied largely to our own processes and our own willingness to change and our acceptance of (not sure of wording).

SK: What about issues like data privacy or data protection issues – what's going to happen in that arena do you think?

RR: Well the approach we have taken so far is the best that I can offer right now, but like everybody else I don't want strangers looking at my genotypical data and then deciding whether I'm a good insurance risk or not. So de-identification and responsible use of the data is crucial, and I do think personally, that my willingness to offer that data and have assurance it's only going to be used in ways that I have volunteered it is sensible. It's how we've operated with patient recruitment in the past just on a larger scale. But it's not an easy problem – I contradict myself in a way, because if we start using artificial intelligence and machine learning to look at this data then it's not inconceivable that somewhere along the way that a machine is going to figure out oh that data belongs to X.

SK: And regulations like HIPAA might be thrown out because all data is identifiable to a certain extent – it will be interesting to see what comes down the line.

SK: Reads Q3

RR: Well we are seeing it now with wearables and medical devices. China is investing heavily in this space and given the volume of money and number of people that they represent I think that will be the sneeze that catches cold in the world, but also, I am excited about, even if a little cautiously, some of the work we are doing at my company, around augmented reality and different ways of collecting data. A simple example – we do a lot of blood pressure collection types to ensure that we are collecting a patient's true blood pressure whether they are supine, sitting, standing, walking – well you don't take it when they are walking necessarily. There are lots of different ways to collect blood pressure, and we always challenge the fact that in some **protocols** I've seen an average of three readings or you do this or do that, to try the true essence of a patient's blood pressure. And even with wearables there are variability to calibration and so on that get in the way. What if all of those modalities or ways of collecting blood pressure assume is a two-dimensional world. So, in other words if I'm collecting my blood pressure – if you as a physician or nurse are collecting my blood pressure, there may be other factors in the room that might be affecting me. Maybe I'm nervous being in front of a physician who is collecting my blood pressure to see if I've got high blood pressure which makes me nervous and gives me high blood pressure. Maybe it's cold in the room or maybe it is hot in the room. Maybe I had a disturbing event occur on my way into the doctor's office. Maybe there's a construction vehicle outside or some other event. I was recently in a little micro project that we did using augmented reality to actually collect all of the environmental data – sound, sight, even taste if you wanted to into the context of collecting the clinical trial data, and I think that's an innovation we have not seen, partly because it wasn't easy, but now we can know more - not so much the data – but the context it is collected through sensors that are collecting more information that the two dimensional – I want the systolic and diastolic rates. And I think we can use that – that's an innovation although it's definitely farther out there. My company has published on this – not to pound our chest on it, but I think it makes logically sense. Synthetic control arms are an innovation that I think it's time has come. We can simulate and synthesize patient populations based on historical data again something that we haven't done a lot of in the industry, but, I think is logically a path that we should follow. Artificial intelligence, machine learning – whatever you want to call it, has its place in a world where there's a lot more data and that's precisely the first question. So, what's changing – we just have a lot of data coming and we have to cope with it, and I think we can take advantage of it.

SK: And do you think the method of data collection will change. EDC requires the patient to go to the site and the site enters the data into the EDC system. Do you think there is going to more direct-to-patient focus?

RR: There's no question – we have sponsors telling us now that 80% of the data in their clinical trial does not come through a CRF. If I'd have asked that question even 10 years ago or even 5 years ago, they would have said the other way around probably 80% of the data does come on a CRF (laboratory data often being the outlier). But what that's telling us is that they are depending on patient reported outcomes, they are depending on image data, and they are depending on genomic data or other types of data to contribute to trial design and their submission. And while there might always be a place for collected data per CRF-style collection - EDC gave us a lot but it didn't solve that issue of data transposition and potential points of failure in that two-dimensional place. What it did was get us comfortable using technology in the clinic. It got us comfortable with entering data correctly. It got us comfortable with the attributability and the accountability, so it has served the industry well. When I started though, I think our vision was a little different. We wanted clean patient data within a day of a patient visit or in real-time, and it never accomplished that. We still collected the data on paper and keyed it into an EDC solution. Or, at best, we sat in front of a patient and we fumbled our way through an eCRF. And, I think, now with direct data capture, as you suggested, through ePRO or through devices, we finally crossed that line where the data can be relied on, and because we've got more of it, we've got a better chance of identifying patterns and ensuring the patient quality experience is being held.

SK: Do you think EHR has a role to play here?

RR: I do. I don't know that I fall into a majority opinion here. The reason I do is because that data – the argument that data's not clear and so on and so forth…I go back to my EDC comment – people are better at entering data into forms that they were 10 years ago, and EHR data, while it may not fall into the same regulatory space as EDC data, we have the mechanisms through things like statistical analysis of the data, machine learning to quantify or qualify the cleanliness of that data, and the thing that is does offer is the elimination of what we call swivel chair modality or interoperability – that's a convenience for the site. Another thing that is offers, going back to your point about data privacy, you have a much richer dataset now. You have a historical perspective on a subject – my late father who died of brain cancer a few years ago kept in his briefcase. He had an actual suitcase filled with documents, records and images of his tumour that he would bring from physician to physician to try and get them to see what was going on and offer a solution. Had he had access to an EHR solution – he could bring that data into a clinical trial or clinical space, they would have been far more effective for him.

SK: You also wonder about who can really look at all of that data (paper) and come to a conclusion.

RR: My father was a neurophysiologist, so he had a pretty good grasp of it but did not know where to go with it. So, I would agree. Even the brightest minds – highly trained people – you look at a suitcase of stuff – you need some time and some concentrated time and that's the last thing our patients have. They want solutions. They need solutions.

Louis Smith, Associate Director, Data Analytics (UHG)

Wednesday, 7 February 2018

Transcript of Interview

Q1 SK reads question

LS: I don't it would be completely fair to say that it's absolutely broken. The system gets there in the end. Trial drugs do get approved. I think the bigger problem would be that it is a long way short of being as efficient as it could be, and, obviously I have never been directly involved, at the coal face of collecting the data, but from a distance it would have seemed quite crazy to me that the amount of people that travel to site in person as frequently as they did, and that there was no targeting to that. It would likely that there are huge amounts of sites that either don't need it or maybe don't even want someone turning up at their door, having another meeting that isn't a patient meeting, whereas are probably some other sites that would benefit from a little bit more handholding. So that kind of blanket approach...seemed very inefficient.

And between the FDA and the drug company in terms of wanting to speed things up, versus a slower, more cautious approach, there is a little bit of imbalance between where the risk and reward lies. The FDA probably accepts a little more of the risk, or certainly less of the reward, if they push through a drug, or speed things up or accept a newer or quicker protocol, and this turns out to have horrible side effects…the FDA says it's ok, if it does all work the FDA does not reap the benefit while the drug company does.

And the increased use towards electronic medical records or any kind of a digital data – it's a difficult one to balance out between not wanting to bias your site selection but pen and paper and things like that seem a little bit archaic at this stage.

SK: Basing site selection on whether or not they have an EHR or not – is that what you're saying?

LS: Exactly. I know that gets into other problems about biasing your selection of sites, but that would drive an efficiency in terms of the actual trial.

SK: That's right and some of the challenges you see in clinical research is that often the drug companies want a particular doctor and a particular site. It may not necessarily be a good site, but that doctor may have prowess in a particular therapeutic indication. There's also factors that impact site selection…

LS: Primarily, it's not that it's broken, but I would imagine that there's a lot of fat there and things could be optimised.

SK: ...based on your experience in looking at large datasets, and getting some meaning of that data, what do you think are the issues there the way clinical data is gathered today?

LS: A little bit tricky for me. I never really got my hands on any the true clinical data. It was always the operational data that I was working on….and, the key primary difficulty we had is that there was never a long term, or there didn't seem to be a long term view over gathering data that might be zero value at the moment – but downstream could be potentially useful, and so, for instance, when we choose a site, we make record or note of this investigator has taken part in 20 previous clinic trials, and she's written 10 previous papers etc. We obviously look all of that up on clinical trials.gov or Citeline and look up what they've done in published papers. That all gets noted somewhere and is part of the decision in choosing to go with that investigator or not. It is not stored or captured anywhere, so we cannot go back and do a retrospective analysis of saying I wonder how enrolment rates correlate to previous trials or number of published papers or any of that kind of thing. People don't capture it in a reusable way as it's not going to be part of their ongoing business reporting, and there are numerous other examples of that. And I am not sure if it ever got resolved - a primary problem was having a unique key for investigators that would follow the investigator as they moved from facility A to facility B. In a lot of cases, that wasn't the case. And trying to find where Dr Smyth went when he left New York was a problem essentially.

SK: Or he could be operating out of two different hospitals. And I guess that's also an issue for patients as well. If a patient enrols in trial A you don't know if the same patient is in trial B because the data has been de-identified, and you have no way of tracking the patient journey through the clinical trial process.

LS: No, we don't have any way of tracking that and there are other things that I have never seen data such as distance from patient to the clinic. These are things that are probably known but we aren't gathering in a meaningful way. If you are collecting them individually it means nothing but being able to aggregate them up can drive meaningful insights. And propensity to churn assumingly increases with distance to site, but we don't know this.

SK: And you would call this contextual data or is it more than that?

LS: It is largely contextual – it's a static piece of data. It's at a point in time of why we would choose an investigator. One of the other things that I would have often been interested about as well would have been looking at a network effect. I don't think that's something we

really consider. This would be what would be the effect or influence on a doctor's social network? We might look at investigator and say this lad hasn't got much experience, haven't done very many clinical trials, and discount him. But it could happen to be that they are in a hospital where there's a whole gang of very experienced investigators and they essentially have a good support network, and there's obviously a good flow of patients going in and out of this facility. Considering that one degree of separation, that investigator could look very different from another investigator who similarly has low experience but is off in a clinic – one man and his dog kind of stuff.

SK: So, I see what you say. The criteria to select the investigator is perhaps too narrow.

LS: That data does exist in Citeline – you can see what doctors are at the same facility published papers, co-authoring papers. Things like that to build out a much more fuller picture. So, the data is there, it is whether is it actually being captured or just out of reach when we don't have a unique key to say this Dr Smyth is that Dr Smyth.

Question 2: SK: That gives a nice Segway into thinking about what advances are being made in technology that could help this process. Is it purely a technology solution that would fix it or are we talking about processes? What is the route of this and how could changes be made to help this along?

LS: Once the biggest changes in recent years is how cheap and easy it is to store vast amounts of data, so there's no reason not to store it. And improvements in user experience in terms of apps that could be built into a doctor's workflow that all of this data could be captured/is captured. At some stage all of these things are known they just the workflow does not allow for them to be gathered up and sent somewhere for storage. And if they are never looked at again, it doesn't matter – it costs pennies to save. Even if it was only looked at every once in a while, and it helped save one trial, the storage would probably pay for itself. It's about building in and easing the burden for both the investigator and people working in CROs who are storing this data that they are not so short-term focused on just gathering things that they need just for BI reporting. And these things do become "metricable". When I worked in a prior job, the call centre people for instance, they are meant to log one of twenty different reasons why someone would have called. There's a drop-down menu and we make that an enforced field, so they cannot finish the call until they have selected one, and we see what everyone appeared first in the alphabet order was massively over-indexed. So, let's rejig them and put them into a random order, and whatever one appeared first – I think it was with a K – it was suddenly the one that was over-indexed. So, people were only selecting the first one. They were finding that it was burdensome for them. So, it was basically a conflicting KPI. Their KPI was they had to have so many calls done in

an hour, and how can I spend five seconds at the end of each call looking at which field was most applicable – that adds up over my hour and I don't get any credit for achieving that I just get given out to for not making my target. So, I think building up those kinds of processes.

SK: So, it's really like making the user have a vested interest in or engaging with the technology.

LS: Yes. I can give another example. It's an app that we are using in here with our doctors. It's called check my script, so if they are prescribing a script and they are using their iPad and they type in their details, it will immediately the app in the background will check that the patient is approved for that particular drug, otherwise it will pop back up within five seconds and put a red X beside it or something. The doctor himself or herself does not actually have to do anything different, but they have now by using data, by using iPads, by using EHRs, they can now say Oh sorry, I was going to prescribe this but you are not actually covered or you'd have a massive out of pocket payment – do you still want to go with that or maybe I can prescribe you something different. As opposed to getting your paper script and going down to the pharmacy and getting a huge bill – what a terrible user experience and you have to go back to the doctor and get a different prescription.

SK: And you can see how that would expand to say, do you want the generic drug or the market-leading drug so there could be – there are loads of use cases of applying something like that.

LS: But the idea is that the doctor does not have to do anything. It all happens in the background. You are not adding to the doctor's burden in any way. They fill out the form as they have to anyway

Q3: SK reads

In terms of collecting data, you would hope that things will be a lot more automated. There'd be a lot more targeted response in terms of dealing with patients and sites. Wearables will have a big part in this. The orders of magnitude increases in the level of data at a level of granularity we don't have in phase II and III clinical trials can only lead to increases in insight. That's not to say it would be without difficulties – issues such as patient burden, data privacy, risks around spotting something that you're not supposed to spot (but once it's known it can't be unknown), potential risks around unblinding of the data or the trial, but being able to spot those trends an awful lot sooner it's hard to picture a scenario that won't be an important part of optimising data collection in clinical trials

SK: As the data volumes increase, perhaps the trial lengths reduce, as you are getting to a lot of patients earlier. What do think – will/may that happen?

LS: I would think so. With more data and other explanatory variables and the different between subjective and quantitative responses in terms I slept pretty well last week versus this is exactly how you slept, comparing one day versus the next. We're able to correlate that with how much you actually slept; we are able to correlate that with the weather where you were. We're able to correlate that with events. We're able to correlate that with your heart rate. There are all sorts of measurements. I know people use them in fitness for tracking their resting heart rate, and it's the variability in the resting heart rate that's a measure for how over-trained they are. That's a massive level of insight that just wasn't possible not that long ago. The data is at such a fine level there really could be all sorts of known/unknown patterns there at the moment and things that could be told. People measuring blood pressure in the home, as opposed to weekly or monthly tracking by the investigator. The trends will become more obvious, more quickly because there's less fuzziness around it. It's not, oh yeah, I slept well – here is exactly how well you slept – we can put a more accurate trend, and it should become more obvious much quicker.

SK: And what do you think about that whole concept of – going back to that example you gave about the patient's blood pressure being taken in a home and not having to go to a clinic – maybe that data could be populating an EHR and the EDC or the clinical data collection system simultaneously - or do you think they are going to become one?

LS: I think, ultimately, it will start tying in with the new Apple – your phone will contain your EHR

SK: Your own personalised record that you can share

LS: Yes. This is your own EHR; you will be able to track all of that. And obviously the ones that are done in a clinician's office will be more heavily weighted. You could submit in – here's mine, and particularly in place in the US there might be some imperative to create fraud on that. We all stick all of the measurements on our healthy best friend, but when you go to your clinician's office, that obviously not going to hold up very well and that there are around wanting to dirty-up the EHR. You don't want to make too much of a mess of it by filling it with fraudulent information. And, I think, ultimately – storage is so cheap. Why not aggregate that all up there? It will probably be quite a different world when patients start having – and that's what I mean about the risk of unblinding. If a patient is tracking themselves and they have these devices (which are so cheap) and they'll start noticing a change themselves, and say, oh, I guess I'm not on the placebo. Or they see no change and

they say I guess I'm on the placebo. I was expecting the holograph field measure these things. I can now say myself; maybe I'm less inclined to go through with this clinical trial as I think I am on the placebo or the control.

SK: There's also possibly a risk of the patient taking charge of their own treatment. If you can read what's wrong with you on your watch, your watch is telling you something, you may act on that without consulting a physician or clinician.

LS: Definitely, but I was more thinking about the potential. It could potentially be a plus and a minus for clinical trials.

SK: A demotivating factor to know that you have a condition, but you know that you are getting the drug for that condition. It makes you think will people of the future use doctors at all?

LS: Even now, it would be pretty much the first thing anyone would do is Google it. …starting to bleed from my ears…I better Google that. That's everyone's first port of a call before they call a doctor.

Richard Young, VP, Veeva Systems

Wednesday, 7 February 2018

Transcript of Interview

SK: Reads out Q1

RY: A very short answer – I 100% agree with what she says. I think it is totally broken because there is a convergence of a number of different issues or challenges, I suppose you should call them; and we are operating in a very outdated model, and if you think about what those challenges are – they are driven by a very personalised or precision medicine movement, and I think when you overlay that with the advances in technology on one side, and overlay that with the changes that are about to hit us in terms of reimbursement, you can see how everything needs to be revisited.

If you take those three pillars for a second, you know it used to be that pharma companies got paid for selling pills; now you are going to get reimbursed for outcomes and that leads to a whole host of questions for me. For example, the simple scenario, for me, is if you get 100% of the reimbursement you are owed based on 100% of outcome, you need the patient to be 100% compliant with the way you administer that treatment. But we know globally that patients are only about 65% compliant. So, the question is can you survive on only 65% of the revenue you are projecting? So, I think, the reality of that is that you are going to see a lot more of what we call pragmatic trials. This idea of instilling a real-world element. Now, for me, the ultimate pragmatic trial is you gather a group of patients and say here is the study drug. We recommend you this with it, and off you go.

Now, that's going to blow the mind of most people. So, we'll come back to that.

The other thing I said is – personalised/precision medicine – that really is the way you're going to prove what the reimbursement level should be. It's how you're going to prove how the drug actually works, and, I think, with that you are just getting to expose the idea. You don't get to know patients on every possible level, you know. And, now, I think, with movements like the 1,000-genome project(s), we're starting to really explore what that can really mean, not just for the patient in front of you, but for the patient yet to come.

And I think if you look at some of the stats from Tufts. Medidata & Tufts in 2015 talked about just under 1m data points for a phase III study. I question the exact accuracy of that number. It's indicative.

SK: It's a large number.

RY: But, it's not a large number when you look at what we are doing today. So, we're doing our first mHealth study with Autograph, and all we're doing is sleep and step data for 9 months – that's 1 billion records per patient. I've got 12 patients and that's a small study, but now I'm talking about 12 billion records. There is no technology in the world that deals with that right now. There are technologies that could deal with it, but they haven't built out the functionality – I'm one of them. There's progress to be made there. But the volume and variety – I talk about the five Vs of clinical research. Volume, variety, velocity, veracity and value – that should be your calculation. You look at the volume of data you are collecting and if it is appropriate. But understand, regardless of what you think, the volume of data available is going through the roof. ePRO and eCOA will be used in 80% of studies. mHealth is – frankly, we're not even scraping the surface of what mHealth will do for us in the future. You probably saw the recent Apple announcement about the electronic health records, and things like that…so that's volume.

Variety comes in two ways. What I call dry data and wet data. So, there's the mHealth, there's the ePRO - the patient reported outcomes – they're dry measures. The wet measures – are we going to take more PK samples, more genomic samples – we're going to do more analyses that way.

So that's going to stretch us in terms of the volume of data.

The other key phrase for me is – you got to turn data into actual usable knowledge. So how are you bringing all that data together in real-time. So, variety, for me is all of that data – not just in terms of the variety of the data itself. It's the variety of formats, the variety of availability; the variety of level, of scrutiny required.

Of course, from a velocity perspective, everyone wants it here and now. We're part of the iPad generation – everyone wants it here and now and it works instantly, doesn't it. So, that's exactly how we've set the benchmark for everything we do.

Veracity – that's an amazing word. As a data manager, veracity – there's a word I never thought that I'd embrace! Because this goes directly to Janet Woodcock's quote. When I was brought up as a data manager (aging myself!), data unlocked was a career ending event. You didn't open the database without the approval of about 500 people. The only available standard was perfection. Now we're saying *good enough could be good enough*. We're talking about not SDV'ing everything. And people still talk about it with a sense of *this is risky*. My whole point is – it's not, it's common sense. If I give you a billion data points or a billion records of mHealth data, what are you going to do with it? So, I think, veracity is a huge option.

And the way I pitch it is. If you take those four Vs – the fifth is value, and that is the absolute overriding factor for everything. Take those four Vs – add them together and what do you get – the value of the data should be exactly where you need it to be and if it isn't then you know what you've ??? too much or too little, the wrong endpoint or the wrong dataset or you are missing an opportunity, whatever it might be. I think that's where I encourage people to look at. And, I think, when you add all of that together, that's where I think Janet Woodcock's quite right, because too many people are ploughing the traditional path and it takes too long. We're exposing patients to treatments that probably are going to lead to futility. We're not using the data in the right way.

Another phase, I've used, is I want to succeed quickly, but I want to fail early. Too many people still think failure is a negative. We're experimenting here – 97% of what we do is going to be a failure at some point. Rather than turn from that fact, the sooner I fail, the sooner I move on. So, I think, those are the kind of main challenges, as I think to her quote, and obviously there's a lot of subtext (SK note - not sure if this is the correct word) but given the time you may want me to pause there.

SK: No, that's perfect. I think that's a nice Segway into the 2nd piece. If you look at the second question – it's about what's changing health data and how it's collected, and what are the advantages/disadvantages of such changes. And you've talked there about mHealth, which obviously is a palpable change which is happening in the industry. Probably…we're certainly seeing it on a day-to-day basis, but I guess you touched on some of the advantages, because you are getting the data directly from the patient, you are getting "veracity" of data there, but you've also got this problem about the volume of data, and maybe you're not collecting the right data, who knows? We still have yet to explore that. I guess, the industry's slow to change. What's your view on that? In terms of why is the industry so slow to change?

RY: Part of it is fear. But, it's also...it's the way we've structured our technological agreements. The way we have become wedded to a process. I think, also, we still live in an environment where people are building empires. I said this at last year's PCMG. There was a bit of a shameless rip-off of a bank TV advert, which said, if you're not prepared to change your business model every 12 months as a technology company, you shouldn't be here.

The world is moving so fast, you have to think through what you're doing, and customers – sponsors and CROs, should reap the benefit of that. As a tech company, my job is to overcome complexity, is to overcome barriers to make you better, faster, cheaper – whatever the goal is. And, if I don't hit those goals, you should be able to walk away.

Historically, what have we done, and you know this from your past life, we've sought to block vendors into longer term agreements for some financial benefit but, the second you do that, consciously or subconsciously your whole organisation is now we're using this tool for the next five years, and when the new shiny object or opportunity comes up, there's an instant barrier to that adoption. There are pros to that as well as negatives. We can't just change every two minutes for the sake of it, but, I think, that is, definitely, a barrier. People are stuck with old technology, and that gets worse every day. Look at the rate of technology – it's incredible – and, I know there are stats around it which I don't recall off the top of my head. We double the volume of data we generate every year of every two years – I don't know.

SK: Every time I look at it, it changes. It gets quicker and quicker.

RY: That's probably the perfect example. Every time I look at the number it changes. We hamstring ourselves with the technology we pick, but we also, I think, as an organisation, as a global organisation – too many people buy technology very rarely. And, what happens then is that they take that technology and they try to apply it somewhat largely to the world they currently run in, and now we have a discrepancy. So, I think, this rings true of big systems, but, increasingly, going back to mHealth, I can't even tell you how many mobile devices there are now, and how many home testers there are – there are 100s if not 1000s. And, you know, as a technology company, the second I even think about being a portal for those, I probably shoot myself in the foot (with both barrels). For me, there's that dichotomy between innovation and validation, as soon as I stop innovating and validate something, I slow down, but you need someone to validate. But, I think, creating very open infrastructure, a network, if you like for that free exchange of data where you take advantage of those companies. They know what they need to do, so that they can join your clinical trial ecosystem far more easily.

SK: it's a bit like, you know, international phone chargers. I often think of it that way. I remember every mobile phone had a different adaptor, and at some point, someone said this is ridiculous – we need one adaptor for every device. There are still some differences. But, you know yourself, at least there's some model there – that there's open APIs (or whatever they are) that people can plug into

RY: the other thing that people have got to get their head over is that word veracity. It amazes me, even today – I have been talking about this for five or six years, and I don't understand why I'm still talking about it. The 6-minute stress test – it there a least useful test in the world? I can give you an example I'm working on, at the moment. We're doing a validation piece to develop a ? marker using these wearables and talking to the sites about how we want to set a few things up. They have given us examples, and said, "look, when we

do the six-minute stress test, these are very sick cardiovascular patients". What we're realising is a. there are patients who stay in bed for three days beforehand (literally), build themselves up to do the best performance they can, then there are patients who will not spend the time upfront, but after the stress test they are so exhausted, they spend the next three days in bed. Either way, you get an absolute fake result. And then there are patients who treat it as their normal day, who do terribly, and you up saying, "we've done terribly" that's my average day, but you compare them say, but this patient has doubled what they've done since we first spoke, but that's a totally fake comparison, because that patient you are comparing to has spent three days in bed preparing for it. But that's not only a clinical assessment, but the problem is that is how the FDA and the authorities, will say, well, look you've got a big separation, we can't trust this, and, I think, mHealth data gives you that option to say, not necessarily on a purely quantitative basis level, but on a qualitative level I can now see a pattern…

SK: It omits the patient from the process. If you are a sick patient do you really care how you perform on the six-minute stress test? You are more concerned about your day-to-day living standard, I guess, quality of life.

One other thing I wanted to ask you about was the regulators. They always get a bad rap, and they are also told that they are not dynamic enough and they are forcing the clinical trial industry, you know, to comply with unwieldy regulation. What do you think? Do you think it's changing or are they still stuck in the mud?

RY: I think, it's changing very much. I think it's a two or three step process. For me, the first step. I think three or four years ago I would labelled that criticism or directed that criticism to them for sure. And then you see in the last three or four years these white papers get launched, what does RBM mean, what is mHealth…and I see these papers as invitations for you to go to the authorities and say I want to do something different. And, that is a big first step. But I think that the second step that is required is a change in the way I see the regulators engage. Because what I'm still hearing is...you hear the regulators say, come to us with ideas, no-one gets in trouble for coming to us with an idea. We want to hear, we want to talk…but I still hear is, that we go to them with an idea, and what we hear is, if that's what you want to do, that's your decision. That's not an answer. And the fact that there's not an answer, and generally leads to people saying we shouldn't do this, it's too risky.

SK: And based on past experience, I guess, you know, changes haven't been made.

RY: Yes

SK: You touched on it there, just looking at Q3, about innovation and how there's going to be future changes. Certainly, we talked about ? and that being a change. Do you think there's anything else that's going to come down the line? You know, people talk about artificial intelligence, RPA or other things, even blockchain. Do you think that's really going to impact clinical trials in the next 10-15 years or sooner?

RY: I think it is, I think the patients themselves are going to change things most, and part of that is privacy and part of that is demand. You see this in the rare disease where I think it's Scott Dan and Dr Fashenbaum, they've said look my risk profile in a rare disease is very different to that in any other areas, so, I think, you're going to see that mentality creep forward into everyday trials. But, I think, what's really going to come forward is the patient's demand for knowledge and I think patients are frustrated that they donate their time, they donate whatever fluids that you need, and at the end of the day, maybe they get lucky and get on a treatment that helps them, maybe they don't. What we've got to do is to instil education and better practice, and kind of reinvent the pharma industry into a trustworthy position, where patients, they know they are part of a trial, but they know what the outcome, and they are being educated all the time.

SK: So transparent really to them what's happening.

RY: Yes, again, I come back to mHealth because one of the big opportunities is… consider a diabetes patient. A patient pricks their finger, does a blood glucose reading, pricks their figure once a day, twice a day, three times-a-day. That data is collected, that's great. But what if on their phone while they were doing that, they also had their step count, what if they also had their Hba1C result from the lab, what if had a comment from a central nurse saying, you take your step count, your hba1c is under control, what a good job or look something's not quite right, I want to you titrate your insulin or do this, or do this or come and see…that's you pragmatic trial right there – but that is buying the patient's confidence, time and belief in a process in a way that we've never achieved before. And then patient's say I know that if it's not working, I know that you're going to put me on something that is. You'll help. Right now, you've got patients believing I know now that it's not working and that's my only judge, and I don't think you drive compliance when you have people who are questioning the value of what they are doing (SK: and not getting feedback...) And, I think, lack of compliance is a huge issue for us. Remember the old ePRO argument about the car park syndrome. I don't think that's really gone away. Maybe in the ePRO case it had, but in other parts of the clinical trial process, I think, it's still valid. Patients don't lie because they are trying to fool anyone, I think, they just say oh, I forgot to do that, or I should have done that...I don't want to be shouted at, I don't want to let someone down, so I'll say yes, or I won't report that. And

the problem is you can't play that guessing game with clinical data. That's something we've got to get much, much transparent at.

SK: And what the patient burden then? So, say we give them all of these devices and you've got chronic COPD (or something like that), and we've given them seven devices around the house. Some of them (as you say) may not need intervention from the patient, but I can just imagine from my own family that would pose issues. I could see my parents saying…I don't want to use this stuff...and getting frustrated. What do you think are the ways going to be to help the patient along, you know, to…?

RY: I think your example's a good one. I think the problem is and the opportunity is you have such diverse populations. For some patients, maybe it's a generational thing to a degree, but it's also a therapeutic or conditionally thing. Some patients, being at home and not having to see someone every week is brilliant. You gamify...gamification will be absolutely what they want. For other patients, you know what…either a. it's not feasible or b. it's not safe (perhaps we don't want to do that). It's about finding the right balance what do you need that investigator or nurse to do that's actually critical? If you're bringing a patient all the way to a site just to read a chart that, to me, is a waste of everyone's time. If you're bringing him to a caregiver to x-ray or something that requires that skilled intervention, that's high value. If you're seeing poor results for? the patient, the high value requirement is to speak to them and motivate them - see them, check them, whatever. But, I think, the beauty of technology is that we'll soon be able to do that patient by patient. In the short term, I still think we can do it study by study. I did a mHealth study in China. And everyone said that wouldn't work, and you know it did. The only problem was internet connectivity in parts, but it did work. You can overcome these challenges, if you think it through. But, I think, the key thing for me, going back to the patients and doctors is, if you're medical appointments are anything like mine, and even in a clinical trial, what do you get 20 minutes with an investigator if you're really lucky. Or give them 19 minutes of those 20 minutes dedicated to their condition. If some of time is being spent transcribing or fighting with technology or doing something other than focusing on me as the patient, you know, you are damaging our relationship. So, you know, I think it is vital we give them that visit back. That's my number one goal. Give the patient their visit back and their time back? to do what's important not just to check in on something that could have been done remotely or over the phone.