APPLICATION OF AUTOMATED FEEDBACK FOR THE IMPROVEMENT OF DATA QUALITY IN WEB-BASED CLINICAL COLLABORATIONS

Stephan Glöckner

(ID: 64 34 99)

Dissertation

submitted in total fulfilment of the requirements for the degree of

Doctor of Philosophy

Supervisors: Prof. Richard Sinnott (University of Melbourne)

Prof. Wiebke Arlt (University of Birmingham)

Co-Supervisor: Dr. Reeva Ledermann

December 2016

Department of Computing and Information Systems

Melbourne School of Engineering

The University of Melbourne

Produced on archival quality paper

ABSTRACT

Background

Biomedical research typically relies on data collected from patients in clinical settings. This is currently a fraught process due to the diversity and heterogeneity of data management systems, the numerous data standards and the sensitivities around the access to and sharing of such data. To tackle this, international biomedical registries are often established and targeted to specific diseases and communities. The quality of the data in such registries is essential to ensure that the clinical research findings can be translated into clinical care. However, at present clinical data management systems developed for biomedical research rarely perform quality assurance procedures during ongoing data collection. Similarly, clinical trials typically perform data quality assessment at the end of the trial. This is too late. We argue that data quality assurance procedures for cost reduction and data process improvements have to be implemented as an integral and ongoing part of disease registries and the data that they are used to collect.

Such an approach requires all aspects of data collection efforts are considered including the intrinsic and extrinsic motivational factors of data entry personnel and the organisations in which they work. Technical solutions that encourage better behaviour and hence improve data quality are thus encouraged.

Hypothesis

The web-based interactions between data entry users and data management systems can be used to improve data quality. Leveraging technological advancement of web-based registries, new feedback mechanisms can be used to improve the overall data quality of the data that is captured by registries. This should lead to streamlined and improved data capture methods that support the users and ultimately benefit the clinical research more generally. This thesis proposes that web-based data quality feedback can motivate registry data entry personnel, increase their contributions and ultimately improve the quality of registry data and its (re-)use to support clinical trials.

Methods

To explore causes of low data quality and user motivation, a survey and an assessment of quality indicators in a multicentre clinical setting was performed. Based on this, we developed and evaluated a stage wise framework for web-based feedback and measured data quality trends including the factors that can impact user motivation in their data entry. This was explored in the International Niemann-Pick Disease Registry (INPDR) and two major international clinical trials associated with the European Network for the Study of Adrenal Tumours (ENSAT). We also considered the role of patients in data collection through mobile applications supporting data collection with the context of the Environmental Determinants of Islet Auto-immunity (ENDIA) clinical study.

Results

Researchers are motivated when they see the contribution resulting from their data entry and the improvement in treatment of patients. The results of the survey and the framework evaluation highlight the effectiveness of web-based automated data quality feedback. It was discovered that data quality feedback to researchers and the research community improve the data quality. Case studies showed an increase of data quality over the period of observation of this research – noting that these studies are still ongoing. The stage wise framework to evaluate data entry user behaviour after feedback was applied to one trial, which showed that feedback encouraged users to enter both more and higher quality data.

Conclusions

Recent literature confirms the need for data quality feedback as an ongoing and near-real time activity associated with data capture. Centralised data monitoring requires a general framework that can be adjusted for a variety of trials and studies. The proposed stage wide research method must be improved to measure the outcome of data quality feedback against a control group and/or where known benchmarks exist.

Data quality dimensions need to be adjusted to all research interests. In the age of big data and mobile health possibilities, further research needs to be performed with regards to the upcoming challenges of data trustworthiness and record eligibility to tackle current and future research objectives. The findings highlight how biomedical research registries have to be designed with focus on data quality and feedback mechanisms.

DECLARATION

This is to certify that:

I. The thesis comprises only my original work towards the PhD,
II. Due acknowledgement has been made in the text to all other material used, and
III. The thesis is less than 100,000 words in length, exclusive of tables, maps, bibliographies and appendices.

Signed,

_____

Stephan Glöckner

December, 2016

PREFACE

This thesis has resulted in attendance and presentations at several conferences through posters, abstracts as well as several refereed conference/journal papers. These have included:

JOURNAL and CONFERENCE PAPERS

Sinnott, R., Beuschlein, F., Effendy, J., Eisenhofer, G., **Glöckner, S.**, & Stell, A. (2016). Beyond a Disease Registry: An Integrated Virtual Environment for Adrenal Cancer Research. P1-18, *Journal of Grid Computing, DOI: 10.1007/s10723-016-9375-x*.

Glöckner, S., Arlt, W., Bancos, I., Stell, A. J., & Sinnott, R. (2015). Improving Data Quality in Disease Registries and Clinical Trials: A Case Study from the ENSAT-CANCER Project. In A. Maeder & J. Warren (Eds.), *8th Australasian Workshop on Health Informatics and Knowledge Management (HIKM 2015)* (Vol. 164, pp. 25–32). Sydney, Australia: ACS. **Best Paper Award**

Glöckner, S., & Sinnott, R. O. (2016). Increasing Clinical Data Entry Quality Through Automated Web-based Feedback. In *Proceedings of the 4th Annual Doctoral Colloquium Computer and Information Systems Melbourne*.

Glöckner, S., & Sinnott, R. O. (2015). Web-based Feedback as a Tool for Data Quality Improvement in Biomedical Research. In *Proceedings of the 3rd Annual Doctoral Colloquium Computer and Information Systems Melbourne* (p. 28). Melbourne. **(incl. Poster)**

Stell, A., Sinnott, R., **Glöckner, S.**, & Effendy, J. (2015). Advanced Clinical Study Monitoring through Distributed Registry Data Management. In *Proceedings of eResearch New Zealand 2015*. Queenstown.

Glöckner, S., Stell, A. J., Effendy, J., Bancos, I., Arlt, W., & Sinnott, R. (2014). Improving Data Quality in Disease Registries and Clinical Trials: A Case Study from the ENSAT-CANCER Project. In *Proceedings of the 13th ENS@T Scientific Meeting*. Nice. **(incl. Poster)**

Talks and Presentations

- 4th & 3rd CIS Doctoral Colloquium (Melbourne),
- 8th Australian Workshop on Health Informatics (HIKM, Sydney),
- 13th ENSAT Meeting (Nice), ENSAT-CANCER General Assembly (Venice),
- 8 **invited talks** focused on Data Quality in Clinical Research (Athens, Padua, Florence, Turin, Paris, Wurzburg, Munich, Birmingham)

Papers under consideration

Glöckner, S., & Sinnott, R. (2016). *The Impact of motivational and environmental Factors on Data Quality in Disease Registries: A survey in the ENSAT-CANCER Project, in progress for Journal of Biomedical Informatics*.

This paper focuses on the analysis of the ENSAT survey with focus on data trustworthiness.

Glöckner, S., Eisenhofer, G., Beuschlein, F., Effendy, J., Stell, A., & Sinnott, R. (2016). *The Changing Landscape of Biomedical Research in the Age of High Technology Multidisciplinary Teams and Big Data: Focus on European Adrenal Research*, submitted to Journal of Contemporary Clinical Trials, May 2016 (under review).

This paper focuses on the impact that the ENSAT-CANCER has made on global adrenal tumour research. The paper is currently under review.

Papers in progress

R.O. Sinnott, W. Hu, J. Wu, **Glöckner, S.**, ENDIA Study Group, *Usage Patterns and Data Quality: A Case Study of a National Type-1 Diabetes Study*, in progress for 7th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, Workshop on Clusters, Clouds and Grids for Life Sciences, 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, Madrid, Spain, May 2017.

This paper focuses on the usage patterns and data quality of the ENDIA clinical study.

R.O. Sinnott, W. Hu, **Glöckner, S.**, *Usage Patterns and Data Quality for the International Niemann-Pick Disease Registry*, in progress for 7th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, Workshop on Clusters, Clouds and Grids for Life Sciences, 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, Madrid, Spain, May 2017.

This paper focuses on the usage patterns and data quality of the INPDR clinical study.

ACKNOWLEDGEMENTS

After six years of research about the improvement of data quality in clinical registries, I have to thank numerous people.

First up, I would like to thank my supervisor Prof Richard Sinnott, who motivated me through all stages of this journey. Without his positive pressure and attitude, I would not have accomplished my thesis in a timely manner. I am very grateful and really appreciate all the time he spent on supervising me. I hope that we can collaborate in future projects and also finally beat him in badminton one day.

In addition, I want to thank my second supervisor Prof Wiebke Arlt, University of Birmingham, who helped inspire this thesis with her clinical knowledge. I am especially grateful for her support during the application for travel grants that allowed me to gain more expertise and background in site-specific data processes that have underpinned my own research. I really appreciate that even in very busy times, she had time for me and my questions.

I am also thankful for the support and feedback from my co-supervisor in Melbourne Assoc. Prof Reeva Lederman with her knowledge about Information Systems and my PhD committee chair Prof Frank Vetere with his critical response in many aspects of this research, but especially on user interactions and feedback.

I am also very grateful for the software development and support of Anthony Stell, who developed the ENSAT registry and PMT eCRFs alone and William Hu for ENDIA and INPDR. They helped realise the feedback loops and log-based analysis systems that have been used throughout this thesis as the basis for evaluation of data quality and the impact of data quality feedback.

I would like to acknowledge the support of Dr Irina Bancos, Mayo Clinic, during my initial approaches to evaluate data quality for EURINE-ACT and her help in teaching me about the pathophysiology of the adrenal gland. Furthermore, I have to thank Jemie Effendy for his support with his Python scripts (chapter 3) that allowed me to export data from several databases so that I could analyse them in a clinical context. I am also very thankful for the support from Salla Eliisa Toikkanen, Helmholtz Centre for Infection Research, in the last stages of my PhD with her statistical expertise (chapter 7). I am also grateful to the ENSAT network and the data entry personnel who provided the time for me to visit and better understand their daily data entry processes.

My heartfelt thanks go to Prof Graeme Eisenhofer, University of Dresden, who provided me with the opportunity to explore the area of adrenal research and who originally connected me with a highly motivated international community focused on improving clinical research. It was through Prof Eisenhofer that I received the chance to meet Prof Sinnott and start this project based on my research interests and background.

Finally, and most importantly, I want to thank my family and especially my wife Marie. Her continuous support, love, encouragement and quite patience in all stages of this research were like a rock for me. Also very inspiring was the time I could spend with my daughter Sanya and son Ben who showed me the true purpose of this life. I thank my parents Annette and Gerhard, for their support in allowing me to explore challenges they haven't really expected me to accomplish. Lastly, I thank God for the possibility to take this journey and who blessed me with these awesome people I met on my way. Thank you so much.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

GLOSSARY

ACC	Adrenocortical carcinoma, a malignant adrenal tumour. It secretes corticosteroid hormones (Arlt, 2011).
ACTH	Adrenocorticotropic Hormone is a hormone measured during a Dexamethason Test (DST).
ADDN	Australian Diabetes Data Network.
ADIUVO	a clinical trial associated with the ENSAT registry.
APA	Aldosterone producing adenoma, is an adrenal tumour that causes the production of higher levels of aldosterone, which subsequently causes lower renin levels that lead to primary aldosteronism (or Conn's Syndrome) (Schirpenbach & Reincke, 2007).
Big Data	in the biomedical context, big data is used to describe the increasingly voluminous and heterogeneous data that can now be generated by sequencing machines, sensors, mobile devices, clinical devices or indeed people. Big data often is described by volume, velocity (speed of production), variety, and in the clinical domain especially, data often has need for veracity (trustworthiness).
Data Mining	aggregation and analysis of large (big) datasets to make data more readily understandable. Data mining is also used in a range of data-driven research approaches (instead of hypothesis-driven research) (Batini & Scannapieco, 2006).
DC	data completeness.
DIKW Pyramid	Data, Information, Knowledge, Wisdom Pyramid. A framework used to understand the context and purpose of data collection and its processes used to improve research knowledge (Division & Pharmaceuticals, 2001)**.**
DQS	data quality score.
DST	Dexamethason Test is a test that checks if an adenoma is hormone-secreting (Nieman, 2010).
eCRF	electronic case report forms, a digital data capture system used for trial related case/subject information that is stored in a clinical database (ECRIN, 2013).
ECRIN	European Clinical Research Infrastructure Network.
EHR	electronic health record, in this work also referred to as EMR (electronic medical record), is a system that captures and stores a digital copy of patient related case/subject information – typically in a hospital database.

EDC	electronic data capture, in contrast to manual data entry (MDE) a direct feed of patient related information from a medical database to a research database (Bacchieri & Della Cioppa, 2012)**.**
ENDIA	Environmental Determinants of Islet Auto-immunity.
ENSAT	European Network for the Study of Adrenal Tumours is an international leading research network that focuses on adrenal research and on improvement of diagnosis and treatment of adrenal tumours.
EURINE-ACT	Evaluation of Urine Steroid Metabolomics for the Differential Diagnosis of Adrenal Cortical Tumours is a prospective diagnostic trial, focused on differentiating benign and malignant adrenocortical tumours.
EU FP7	European Union Framework Program 7, a framework programme for research and technological development, which was a funding scheme of the European Union between the years 2007-2013. The successor is Horizon 2020.
FAMIAN	Combined FDG-PET and 123I-Iodometomidate Imaging for Adrenal Neoplasia is an interventional study performed by ENSAT to verify the thesis that a combination of FDG-PET and 123I-Iodometomidate more accurately detects adrenocortical adenomas, rather than extracting the tumour during surgery and obtaining a diagnosis by a pathologist.
FDA	Federal Drug Association, a federal agency in the USA that defines standards and provides guidance especially for biomedical research.
FIRST-MAPP	a clinical trial associated with the ENSAT registry.
GCP	Good Clinical Practise is general standard for the performance of clinical trials, defined by the International Conference of Harmonisation (ICH).
Gold Standard Test	in clinical research a new diagnostic approach is generally compared to a gold standard test, which is defined as the most efficient test, currently used in routine care. This test has the highest values of sensitivity (probability to detect a disease – true positive rate) and specificity (probability to exclude patients that do not have the disease under test– true negative rate).
h-index	is an author-level metric that attempts to measure both the productivity and citation impact of the publications of a scientist or scholar. The index is based on the set of the scientist's most cited papers and the number of citations that they have received in other publications.
HU	Hounsfield Units – a quantitative value of radio density of Computer Tomography (CT) results. In adrenal research HU supports the determination of whether an adrenal tumour is benign or malignant (Arlt, 2011).

ICC/RICC	Intraclass Correlation/Residual Intraclass Correlation is used in clinical monitoring to provide a quantified degree "to which observations within a cluster differ from those between clusters" (Guthrie et al., 2012). The ICC supports the data quality dimension of data comparability. The RICC "quantifies the degree of clustering" (Wynants, Timmerman, Bourne, Van Huffel, & Van Calster, 2013) and also provides the possibility to visualise data inconsistencies for data monitoring purposes.
IF	The impact factor (IF) of an academic journal is a measure reflecting the yearly average number of citations to articles published in that journal. It is frequently used as a proxy for the relative importance of a journal within its field.
INPDR	International Niemann-Pick Disease Registry.
Interventional Informatics	also referred as I^2, a novel approach that uses informatics as intervention to improve decision making and clinical outcomes of at risk populations (Payne, Lussier, Foraker, & Embi, 2016).
Kaplan Meier Survival Analysis	A statistical method used to plot the survival rate of a specific disease in an observed time period. The time periods are typically plotted hierarchically and aggregated. In this model short time periods (more critical) are situated on the left side in the graph.
Ki67	is a result of immunohistochemistry, which supports the differentiation of a adrenal tumour and whether it is benign or malignant (Fassnacht & Allolio, 2009). This is also used in combination with the Weiss Score.
MDE	manual data entry, in contrast to electronic data capture (EDC), a direct entry of patient/case related information into a clinical (research) database.
mHealth	an approach to support health related research through mobile devices and associated applications. mHealth offers many potential opportunities to capture and indeed deliver health-related data (World Health Organization, 2011).
NAPACA	non-aldosterone producing adrenocortical adenoma, also named 'incidentaloma' a benign form of an adrenal tumour that is commonly incidentally detected during routine imaging.
NCI	National Cancer Institute, a part of the US National Institutes of Health that coordinates the process of cancer related health information to conduct research focused on the prevention, diagnosis and treatment of cancer.
NPD	Niemann-Pick Disease is a rare metabolic disease caused by accumulation of fat cells in the liver, spleen, bone marrow and brain. NPD is categorised into three types (A, B, C) (Schuchman, 2007)

pCRF	paper case report forms, an analogue of trial related case/subject information that is often stored in folders in locked cupboards in hospital settings.
Personalised medicine/precision medicine	in this thesis, personalised medicine is described as the ultimate goal of translational research. Personalised medicine focuses on the detection and treatment of a disease specific to an individual using advanced knowledge now available since the sequencing of the human genome.
Pheo	pheochromocytoma, a benign or malignant adrenal tumor in the adrenal medulla, which secretes catecholamines.
PPGL	Paraganglioma & Pheochromocytoma, a paraganglima is a catecholamine-secreting tumour, which is not located in the adrenal medulla (commonly occurring in the head and neck lymphnodes)
PMT	Prospective Monoamine Tumour trial is a prospective diagnostic trial focused on improving the diagnosis of pheochromocytoma (Pheo).
SDV	Source Data Verification is an approach used to detect data inconsistencies via re-entering selected data items into eCRFs. In general, SDV is used during external audit procedures, e.g. to detect user training needs.
Sensitivity	see Gold standard test.
SEER	Surveillance, Epidemiology and End Results program is a framework of the NCI, which provides source data about incidence, prevalence and mortality of certain cancer types.
SJR	SCImago Journal Rank (SJR indicator) is a measure of scientific influence of scholarly journals that accounts for both the number of citations received by a journal and the importance or prestige of the journals where such citations come from.
SOPs	Standard Operation Procedures are procedures defined in a clinical context that are used to harmonise processes to improve the comparability of diagnostic procedures or treatments.
Specificity	see Gold standard test.
Translational research/Translational informatics	Translational research has the aim to 'translate' findings of basic research into clinical routine care. Translational informatics supports this approach and provides concepts that move away from general reductionism (where problems are broken down to smaller problems) to a move towards system thinking (Payne & Embi, 2015). Translational informatics promotes data-driven decision making and research, leveraging the increasing amount of data available from –omic analytics and information about individual phenotypes (Tenenbaum, 2015).

VRE	Virtual Research Environment is an advanced web based research environment offering extended functionality to a research community. This typically encompasses research databases, analytical capabilities with collaboration support.
Weiss score	is a system used by pathologists to analyse the morphology of adrenal tumour tissues. It consists of 9 parameters. The more parameters that are positive, the more likely the tumour is a malignant tumour (Fassnacht & Allolio, 2009). The Weiss score alone is not a good diagnostic tool, hence it is used in combination with Ki67.
WHO	World Health Organisation

1 INTRODUCTION

This chapter provides an overview of the context in which the thesis was developed. Beginning with an overview of the current landscape of modern clinical trials and biomedical studies where new information technologies are increasingly being applied to obtain increased impact and improvements. We consider the way in which big data and especially big data solutions can be applied to tackle the challenges of dealing with diverse data holdings. Following this broad overview, we focus especially on the specific problems that motivate this research as a whole. Specifically, we consider the impact of data quality on clinical research; emphasising the need and the challenges of data quality in inter-organisational settings. The chapter introduces the hypothesis for this work. Finally, the chapter concludes with a summary of the aims, scope and the significance of this research as a whole.

1.1 Context

Breakthroughs in biomedical research should ideally lead to improvements in clinical care across the population in a timely manner. The application of new information technologies in the biomedical domain can greatly simplify the work all protagonists involved in clinical research, however such technologies give rise to new problems and challenges that must be solved. One of these problems, which is the focus for the work conducted in this thesis, is the quality of data that is collected and analysed *during* the lifetime of clinical and biomedical research studies. Data quality has a huge impact on research outcome and clinical research objectives, especially in the field of rare diseases and in multicentre clinical research settings. It is the case that the majority of clinical research studies currently rely on data that is manually entered by humans (and not captured and used directly from existing clinical healthcare management systems). Patient related information is either entered directly into trial databases (or disease registries) where the original source comes from pre-existing paper records or electronic hospital records. This process of transferring data into a research study database has a huge impact on the quality of the trial data, and ultimately the quality of the research that can be undertaken in the clinical trial or research studies more generally. To minimise human error, quality assurance procedures such as double data entry are often performed to reduce inconsistencies in the data sets that are collected. However, whilst double data entry is an effective approach to reduce errors, it is very inefficient with regards to both the time and the costs for the trial as a whole. Furthermore, issues with data quality are often only identified after the trial has completed, by which it is too late to take any steps to rectify data quality issues. Instead technical solutions should be implemented into the data management capture and analysis systems to verify and assess the data quality throughout the course of a given clinical trial or biomedical research collaboration more generally.

1.1.1 Research Fields

To understand the context of data quality in a given clinical collaboration whether it is a trial, study or a disease registry, the lifecycle of data must be considered. The general process of collecting patient related information is often divided into: data collection (question the patient), data capture (electronically or manually), optional data transfer and the storage in a clinical database management system for research. The Data, Information, Knowledge and Wisdom (DIKW) Pyramid is often used to describe the purpose of data capture, which can then be translated into information and knowledge for the improvement of patient care. To improve data quality in biomedical research and especially in translational clinical research, we argue that research fields in medical data quality assurance and clinical research quality assurance must be improved. We show through a range of case studies how

and why data quality issues arise and how they can be tackled in a systematic manner that is applicable to many research settings.

It is also important to consider, that such approaches are also taking place during a period of rapid evolution in both technology and clinical and biomedical research.

1.1.2 Translational Informatics

Since the sequencing of the whole human genome in 2002, biomedical research has undergone massive changes with regards to education, research and funding (Payne & Embi, 2015). Simultaneously, the rapid development of technological advancements provides the possibility of the collection of large (big) data sets. Data analytics and data mining now represent major research areas that are applicable to many research disciplines. Big data is a buzzword that has captured the imagination of many researchers and infrastructure providers alike. With the application of any new technologies into research areas, both new opportunities and new challenges arise.

Data aggregation often focuses on the collection of a range of diverse data from potentially distributed data holdings to investigate specific phenomenon. This is often used to support a reductionist approach. In itself, data aggregation is often essential for many research areas; however translational clinical research and translational informatics focus on supporting system thinking, stemming from the analysis of such aggregation of data resources. Technological advancement also provides the opportunity to investigate research as a whole (Payne, Embi, & Sen, 2009). This data aggregation approach often focuses on data-driven research rather than on hypothesis driven research. Shah and Tenenbaum identify that data-driven research can enable proactive, predicative, preventive, participatory and patient-centred health. They argue that this approach can be used to discover patterns of diseases in all patient-related data (Shah & Tenenbaum, 2012).

> *"Data intensive science allows scientists to analyse bigger and more complete systems efficiently, and complements more traditional scientific processes of hypothesis generation and experimental test to refine our understanding of the natural world."* (Kelling et al., 2009)

Biomedical research typically starts with a research hypothesis. This hypothesis is drawn from previous research or from informed guesswork of researchers about something that needs to be observed and/or better understood. The researcher defines and investigates a self-defined problem and explores solutions for this and potentially their influence in other areas. However, in many cases this is a simplified (reductionist) view of the way in which research is really undertaken. The counterpart of this approach is *holism*. The advantage of computer-based systems in biomedical research is that they offer the opportunity for a shift from hypothesis-driven research to more complex data-driven research.

> *"Data-driven medicine will enable the discovery of new treatment options based on multi-model molecular measurements on patients and learning from the trends hidden among the diagnoses, prescriptions, and discharge summaries of millions of patient encounters logged by clinical practitioners." (Shah & Tenenbaum, 2012)*

Despite this rhetoric of (big) data–driven research, there is still a great need for hypothesis-driven research. Indeed, different approaches are now possible to researchers in allowing freedom for defining research questions and a multitude of ways in which such research can be undertaken.

> *"The accepted scientific method consists of formulating a hypothesis and then testing it by experimentation which, at least in theory, attempts to disprove the hypothesis."* (van Helden, 2013)

Data-driven research offers an approach that needs to be considered and tested, and ultimately validated using more traditional approaches. It is the case that for many biomedical research activities it is now only possible to undertake the research through data-driven approaches.

> *"By contrast, the new 'omics' technologies allow us to generate massive quantities of data rapidly and thereby enable us to take a far less biased approach to tackle a given problem."* (van Helden, 2013)

Key aspects of new scientific discoveries made possible by access to increasing amounts of biomedical information allow tackling areas such as personalised medicine and clinical decision making. For example, pharmaceutical industries now increasingly develop personalised drugs based on the targeted genetic information of the patients (Costa, 2014; Ramsey et al., 2011).

Such a data-driven approach can potentially transform biomedical science and simplify research more generally. Big data-based machine-learning and use of Cloud and high performance computing infrastructures are already benefiting many other fields of research (physics, biochemistry). There is increased interest in using such data-driven approaches in the clinical and biomedical domains. However, more data by itself will not improve research alone and they will not immediately allow that research to be translated into a clinical healthcare setting. In the clinical domain especially, data needs to be captured and structured in accordance with researcher needs as well as meet a range of guidelines for data access and data analysis that are not typically present in other research disciplines such as physics and experiments associated with the Large Hadron Collider (as one example). The simple aggregation and sharing of data is challenged by the nature of the data that is being shared, e.g. due to privacy and confidentiality considerations.

Irrespective of these challenges, there is a clear need to accelerate the speed of translation of findings in clinical research into clinical care. Improvements in data quality is fundamental to drive the process of medical discoveries and allow to evolve and sustain biomedical research for all scientists involved (Martin-Sanchez & Verspoor, 2014).

Importantly high data quality can enhance clinical research and its ability to draw significant conclusions out of (digitised) data holdings. The translational domain has hitherto largely focused on system thinking, but data-driven research is increasingly possible and there is now a major shift to data-driven research more generally.

This thesis focuses on several research areas, most notably: data quality assurance in information systems, translational/clinical research informatics, the e-Infrastructures and platforms that underpin research collaborations and the impact of Human-Computer Interaction (HCI) on data quality and research collaborations more generally. The rest of this chapter provides the motivation and problem statement, together with the primary aims and scope of the thesis and its significance in the field as a whole.

1.2 Motivation

According to the World Health Organisation (WHO), 13% of all deaths worldwide are caused by cancer and more than 100 cancer types are known that need individual screening and treatment[1]. However, the cause of many cancer types is not yet established and the treatment of the many cancer types is currently not established. Cancer research often focuses on monitoring and evaluation accompanied by screening programs and associated treatments. For many years, cancer care and cancer research more generally focused on investigating the effects of radiation therapy, chemotherapy, targeted therapy or surgical removal of primary tumours and understanding the biological processes in cancer metastasis. Given their prevalence, research into many forms of cancer (such as prostate cancer,

[1] http://www.who.int/cancer/en/

malignant melanoma, breast cancer, lung cancer, colorectal cancer) is able to benefit directly from disease reports from many different institutions and organisations both nationally and internationally. These can include reports derived from analysis of data from general practitioners, hospitals and/or targeted (comprehensive) cancer centres. However, this prevalence can also be a disadvantage. Information in such reports is derived from a snapshot of data at a given point in time, yet it is rarely the case that the underlying data behind such reports is systematically made available and similarly it is not used to automatically update such reports when new data is collected or indeed when issues arise and the reports need to be corrected with regards to incidence, mortality, prevalence, survival calculations or the effectiveness of treatments. Whilst common cancer diseases such as prostate cancer, breast cancer or lung cancer can have an established health reporting pattern and processes for data collection and analysis, e.g. data captured and included in regional or national cancer registers, this is typically not the case for rarer cancer types. Cancer forms such as adrenal cancer or thyroid cancer are not reported because of their low incidence rates in the community. In such scenarios, targeted research registries need to be established to aggregate data. Such data typically needs to be collected across national boundaries. In such contexts, insights into data processes and data quality evaluation are essential to underpin any statistically significant conclusions for these rare diseases and importantly, to take steps that can automate the quality of the data that is captured and stored in such international disease registries. At present the collection and quality of data in such resources is largely based on best effort. Up to now there has been little systematic (automated) support for improvements in the quality of data that is entered.

1.2.1 Data Processes of Clinical Registries

Data captured and stored within clinical data registries can serve many different purposes: health reports, health protection, research and for the evaluation of preventive and therapeutic processes (Hentschel, Pritzkuleit, Schmid-Höpfner, & Katalinic, 2011). The data process is often divided into three general areas: data entry, data processing and data assessment (see Figure 1-1 for the example of cancer registries).

Data process	Analysis criteria	Frame
Data entry	Dataset	**Infrastructure Cancer Registries (Germany)**
	Sources	
	Software	
	Reporting channels	
Data processing	Reporting procedures	
	Coding	
	Record linkage	
	Comparison of mortality rates	
Data assessment	Data quality	

Figure 1-1 Assessment criteria of data processes

The development of epidemiological and clinical registries can improve the quality of reporting. Epidemiology is focused on the health of populations and therefore, epidemiological registries have the primary purpose to publish reports applicable to the population. For cancer this can be calculations

and reports of cancer prevalence, incidence and mortality. The distribution of mortality is often a special case, since official death certificates as source documents are typically needed. Clinical cancer registries monitor the cancer progress of individuals or smaller cohorts to evaluate diagnostic, treatment, rehabilitation and follow-up care of cancer patients based upon general established guidelines and standards (Hofstädter & Klinkhammer-Schalke, 2011). To improve data quality, region and/or nationwide registration of cancer patients is important, because different regions and countries can have many different factors that should be considered in relation to their own cancer patients (chemical industry, socio-economic indicators, smoking rate, etc.) (Katalinic & Hentschel, 2008).

To improve the outcome of clinical trials, every step in the data process needs to be understood, monitored and potentially improved with quality control methods. In every step the process and the source documents have to be assessed. Such patient-specific data can be either generated by devices (biochemical assessments, blood pressure, weight, etc.), from reports of other medical institutions (pathology report, genetic assessments, etc.) or indeed contributed by the patient themselves. In terms of data quality, it is assumed that devices provide less error than the recall of humans (recall bias). Therefore, data from devices or other documents should be preferred or assumed to have higher quality and trustworthiness. It is possible that this information is entered into hospital records directly and re-used (recycled) into research registries through manual or automated mechanisms. If the patient data is written to paper, then it often requires manual data entry into the registry by individuals within the hospital setting. If the patient data is included in electronic health records, then in principle it is possible to directly export it into clinical registries (electronic data capture).

Irrespective of the mechanisms used, the implementation of quality assurance procedures to assess and evaluate the data processes is essential, especially in a clinical setting. Data and patient-related information must be accurate at the stage where it is created and added to paper or electronic records. The steps taken to transfer it to registries should also be monitored and improved wherever possible. The question that drives this research is, what influences the overall data quality of data entered into such registries and what is the impact on future reuse in clinical trials and studies and their associated outcomes?

1.2.2 Example of Cancer Registry Data Analysis and undocumented Data Quality

To understand the challenges that arise we consider an explorative study involving a German dataset of malignant melanoma that was used to calculate 5-year survival rates (Kaplan-Meier survival). This data was collected from epidemiological cancer registries across Germany in 2011. The study used the variables for the calculation set out in Table 1-1.

Table 1-1 Variables of Survival Calculation

Variable	Description
Date of Birth	Calculation of age (adjustment with relative survival)
Date of Death	Calculation of survival
Date of First Diagnosis	Data for reliable diagnosis (pathology report)
Date of Follow-up	Every data of clinical examinations that confirm that the patient is still alive
Date of Events	Example: patients moves from one registry cover region to an other
Date of Final Report	Last status of the patient (alive or dead)

With a tumour size greater than 1mm, the tumour and the adjacent sentinel lymph node were removed (SLNE). In case a tumour cell was discovered in the extracted sentinel lymph node a total lymphadenectomy was performed (LAD). Patients with both interventions have a poor survival rate. Based on this, every patient was assigned to a special tumour stage (Balch et al., 2009). Patients with a very pathogenic and lethal form of skin cancer were classified in stage IV, while patients with a not so malignant form were classified in stage III as shown below in Table 1-2.

Table 1-2 Patients included in Survival Analysis (Glöckner & Schoffer, 2012)

Stage	Patients (without Stage X)	Percentage
0	8.111	22,7
I	18.733	52,5
II	5.434	15,2
III	2.553	7,2
IV	839	2,4

The first step of the analysis involved a comparison of SLNE patients with SLNE and LAD patients stratified by gender. The five-year survival rate of male patients with SLNE and LAD was 0.48 whilst male patients with SLNE had a survival rate of 0.56. In other words, a male person diagnosed with malignant melanoma and a treatment combination of SLNE and LAD had a 48% chance of survival within the first 5 years after initial diagnosis. Female patients with SLNE and LAD had a calculated survival rate of 0.54 and those with only SLNE had a survival rate of 0.71. This indicates that males with LAD and SLNE have the worst prognosis compared to males with just SLNE or female patients with LAD and SLNE (see Figure 1-2 & Figure 1-3).

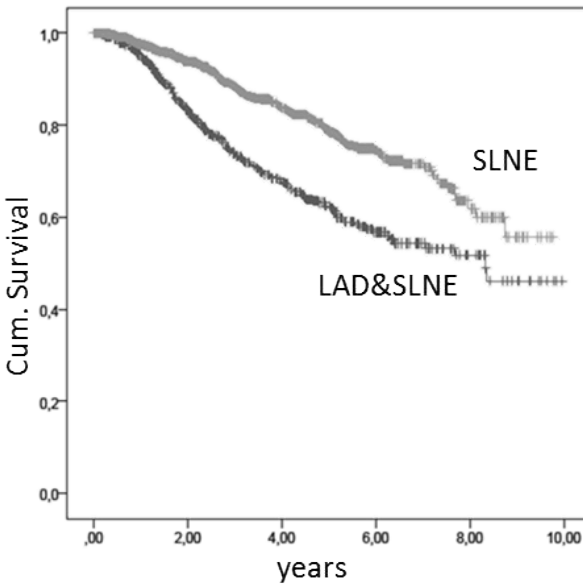


Figure 1-2 Kaplan-Meier male Patients with LAD and LAD & SLNE

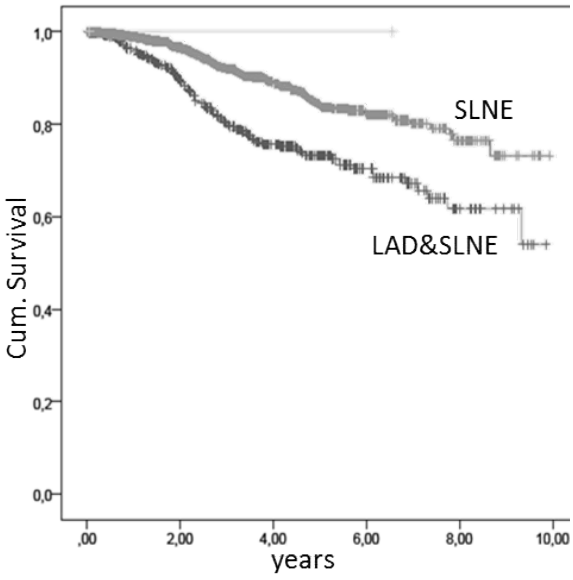


Figure 1-3 Kaplan-Meier male Patients with LAD and LAD & SLNE

The graph shows a smooth curve up to 8 years, since there are fewer patients to follow up after 8 years. What is not shown in the graphs is the completeness and quality of the data that was provided for the analysis. As described, cancer registries often receive their data from many different stakeholders and organisations such as comprehensive cancer centres, hospitals, organ cancer centres or pathology laboratories. These stakeholders typically don't report directly to single established national cancer registries but to population-based registries at the regional or country level. The steps for collecting and checking the data that goes into such registries can vary greatly and standardisations must be improved. Similarly, the analysis that is undertaken by institutions and individuals using the data in cancer reports is often specific to the needs and demands and quality of the data. However, not every source of information reports on the correct and/or complete information required to address such efforts. We argue that a quality assurance instrument has to be continuously implemented into cancer and other disease registries focused on the continuous and systematic evaluation of data, and ultimately on the ongoing improvement of data quality.

In 2009, Bray and Parkin published two articles focused on the evaluation of data quality in cancer registries. Their work included the primary research areas of data quality, medical informatics and epidemiology (Bray & Parkin, 2009; Parkin & Bray, 2009). The researchers identified that the application of tools for the data management of cancer registries provided possibilities to measure data quality and raise discussions on how to improve data collection processes to underpin cancer reports. The authors described four primary quality dimensions for the evaluation of data quality in cancer registries: *completeness, validity, timeliness* and *comparability*. Every dimension has its own approach to provide quantitative values, graphs and visualisations used to show cancer researchers whether the data has high or low quality based on these dimensions.

For the measurement of completeness for cancer registries, they regarded completeness as "the extent to which all of the incident cancers occurring in the population are included in the registry" (Parkin & Bray, 2009). Comparing the completeness of a given cancer registry, they assumed that incident and mortality values for every type of cancer should be similar worldwide. As an example, for mortality to incidence ratio (MIR) of deadly cancer types like lung cancer, they assumed that this should be near 1, while cancer types with a good prognosis (e.g. breast cancer) would have a much lower ratio. The M:I ratio is considered as an observed/expected age adjusted rate (Parkin & Bray,

2009). The authors compared national registries with the worldwide data and discovered that the investigated cancer registry in Japan had a reported 82.8% completeness, in contrast to a cancer registry in Korea that only achieved 50.7% data completeness levels.

In a pre-project for this thesis, a completeness analysis was performed of available data for the five most common types of cancer (stomach, prostate, lung, breast and colorectal cancer). Scandinavian cancer registries (Norway, Sweden, Denmark) were assumed to have the most complete dataset for all reported cancer types (Bray & Parkin, 2009; Parkin & Bray, 2009) since every person was identified with an 11 digit identification number assigned to the person at birth (Larsen et al., 2009). This helps minimise the possibility of duplicates and gives the advantage of better follow-up possibilities. The evaluation of data completeness was performed using age adjustments (European standardized age). For the evaluation of data completeness the calculation of a MIR suggested from Bray & Parkin (2009) was used. The MIR is a quantitative method to estimate data completeness. In this case, all reported deaths of patients from cancer were divided by the calculated incidence of the cancer type.

The Surveillance, Epidemiology and End Results Program (SEER) of the National Cancer Institute (NCI) provides standardized, age, sex and region adjusted incidence and mortality rates of cancer. Cancer of the pancreas, oesophagus or liver are often fatal diseases and therefore have a very high MIR rate. SEER focuses on defining such rates and using MIR as a quality indicator for registry completeness. Comparison of cancer registries should be based on having a similar MIR in relation to case completeness. In a hypothetic case of a MIR for prostate cancer with the value 0.95, a complete epidemiological cancer registry needs to have the MIR of 0.95 (defined by SEER). The calculated completeness rates are related to this value. The calculation of record completeness of cancer registries is shown in Equation 1-1.

Equation 1-1 Calculation of Record Completeness in Cancer Registries

The European Cancer registry data is available online at http://eco.iarc.fr/EUREG/Default.aspx. In a comparison of Norway (assumed complete coverage) versus all German cancer registries in 2010, the data shows that Norway has more complete data for stomach, prostate and colorectal cancer, whilst Germany has a higher level of completeness with regards to lung and breast cancer (see Figure 1-4).

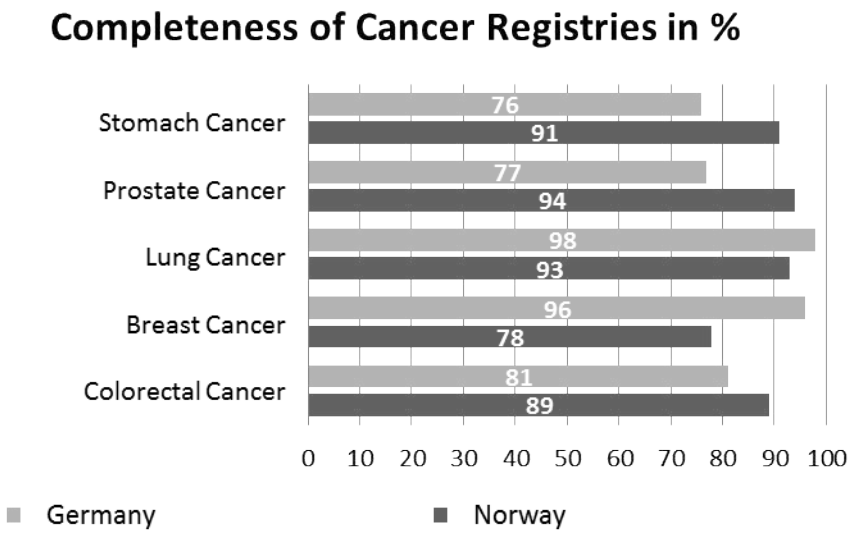


Figure 1-4 Registry Completeness Germany vs. Norway calculated with Mortality-Incidence Ratio

This calculation shows that no general conclusion can be made regarding Scandinavian countries (in this example Norway) over other countries. Similarly the data collection and data processes often differ between registries (Bray & Parkin, 2009; Larsen et al., 2009; Parkin & Bray, 2009). Importantly, the MIR is only one method used to calculate the data quality dimension of registry completeness. There are many other quantitative and qualitative methods that can also be used to calculate data quality dimensions. Other dimensions like validity, timeliness and comparability have their own methods to quantify the overall data quality (from these dimensions) of disease registries. The case studies outlined here have focused on cancer, but the same approaches and methodologies for comparison of data is applicable to many other diseases areas such as diabetes, mental health and other non-chronic diseases.

1.2.3 Data Quality Feedback

One major challenge of data quality is dealing with complexity. Ideally the mean of all data quality dimensions should be expressed in a single score, which describes the overall data quality (see Table 1-3). However not all dimensions can be applied equally in clinical research. General data quality dimensions can include (registry) completeness, data comparability, data validity and data timeliness for registries and data completeness, data correctness (accuracy), data concordance, data plausibility and data currency (timeliness) for clinical research more generally (Weiskopf & Weng, 2013). There are other quality dimensions that can be applied for other research purposes depending upon need and focus. For example, the trustworthiness of source documents is often essential for data mining as explained in Nurse, Agrafiotis, Goldsmith, Creese, & Lamberts (2014).

Table 1-3 Example Overall Data Quality Score

Validity	Comparability	Completeness	Timeliness	Trustworthiness	…	Data Quality
80%	95%	90%	95%	70%		= 86%

The overall data quality (Quality Score – QS) is the arithmetic mean of all used quality dimensions (QD) defined by Equation 1-2. The overall data quality of a registry is a quantitative value that expresses the quality of processes, reports and data contained within that registry. This thesis explores whether this score can be used to motivate individuals to increase their own overall data quality and hence the data quality of the registry as a whole? Specifically, we consider whether combining the data quality score with feedback to researchers can potentially influence the quality of the data as a whole. Care must be taken to ensure that the feedback does not demotivate the individuals tasked with data entry. Instead approaches should seek to inspire them to improve their own data quality related processes and practices.

Equation 1-2 Calculation of the overall Data Quality

$$\overline{\quad}$$

In conclusion, data processes are an important aspect of data quality since they have a large influence on the quality of data. Poor quality data can lead to erroneous conclusions. Even high quality data that does not document how the quality of the data has been achieved and measured can lead to challenges in understanding and reusing the data. The established methods of evaluating data quality in disease registries have to be applied across the wider clinical context. Solutions that work with single disease or isolated registries should be avoided (if possible). Too many data inconsistencies can lead to a lower research impact. Therefore, data quality dimensions have to be defined to transform the data to the needs of the researcher and the intended use of the data by the community involved.

1.2.4 Usefulness of Improving Data Quality

Data quality dimensions such as trustworthiness can lead to questions such as, are the data reported by these institutions reliable and believable (trustworthy)? Are reports provided by national institutions, the EU or even the WHO actually correct and based on data of high quality?

There are many types of disease registries. One type of registry that has directly impact on individual patients and can lead to discovery of new conclusions are clinical trial-related registries. Compared to registries more generally, databases for clinical research cannot often use a comparison database that was collected by other institutions. MIR or other quality indicators that are applied on cancer registries for example can't be transferred to research registries. However, clinical research registries need to visibly demonstrate and measure data quality to ensure the validity and repeatability of insights gained from any studies that have used the data. What quality dimensions and statistical assessments can an evaluation of data quality in clinical research registries include?

Only a few research groups worldwide have investigated the possibilities to increase data quality in clinical research. In contrast to the previous described research about the evaluation of data quality in cancer registries, research regarding the evaluation of data quality in research registries more generally tends to focus on clinical research informatics, human-computer interaction and data quality research. As noted, data quality evaluation normally happens at the end of a clinical trial and its values are not often reported, since it is assumed that 'high-quality' and positive research based on rigorous and high quality data is published. At the end of a given trial, the principal investigator or data-manager typically 'cleans' the data, which typically involves adding average values for missing data or deleting data items that were used for the final analysis. Such late-stage data amendments can significantly weaken the overall quality and results of the clinical trial. Instead data quality evaluation should be conducted throughout the lifetime of the clinical study.

Previous work leading to this research focused on data monitoring, on-site monitoring and the application of quality assurance procedures in clinical trials, however these are often rejected by the clinicians and study sponsors, because they make the trial more expensive and can lead to discovery of problems that may not actually arise or be directly related to the study in question.

There is no direct measure related to the impact of low data quality on the outcome of clinical trials, however data from surveys of business companies showed that the estimated average costs that are associated with 'dirty data' are around $8 Million USD[2]. It was identified that with the implementation of data quality assurance methods, overall costs could be reduced by 10-40% and the outcome (revenue and sales) increased by around 20%. Similar rates were estimated by the Federal Drug Agency (FDA) with regards to clinical research. It was identified that through implementation of risk-based monitoring and data quality assessment by the FDA, trial costs could be reduced by 20-25% (Franco, Hronec, & Slizgi, 2013).

Ongoing procedures for evaluating data quality during clinical trials have to be implemented into the data management system. Compared to other more general data quality research activities, clinical trials have other quality dimensions. Research registries do not need record completeness, but they may often need eligibility checks to determine whether records meet the minimum criteria for a given study or trial objectives. Often this is to determine if the patients meet inclusion and/or exclusion criteria for the study.

On the other hand, evaluating data quality or applying quality methods in research does not directly increase data quality. Factors that directly influence data quality often include environmental factors like the software, hardware and especially the human element. Indeed, the human factor often has

[2] https://halobi.com/2015/02/infographic-data-quality-in-bi-the-costs-and-benefits/

the highest impact on data quality. Considerations such as whether the registry user (data entry personnel) has been trained or is actually motivated to enter research related data are important factors. The complexity of the data to be entered and indeed whether the information required is actually available are all important factors to consider with regards to data quality along with the processes that should be taken to improve data quality.

1.2.5 Summary

Clinical research informatics face numerous challenges with regards to capturing heterogeneous data in a single registry (J. Chen, Qian, Yan, & Shen, 2013). Quality control procedures to achieve high data quality are rarely established or implemented in data management systems, especially those that cross organisational and jurisdictional boundaries. This thesis explores the reasons for low data quality and develops a general framework that supports clinicians, investigators and researchers in capturing higher quality data that can make an impact on their research. We consider the requirements for technological features that support the collection of high quality data during clinical trials with specific focus on web-based feedback mechanisms during data entry.

1.3 Research Hypothesis

The aims of this thesis are to explore and develop a framework for the automation of a web-based tool to increase the quality of data in clinical research registries. Authorities in clinical research informatics have previously concluded that there is no general approach or tool that could fit all clinical research needs (Nahm, 2012). However, the need for high quality data collection in rare diseases especially, requires an approach to systematically provide support to improve data quality. This requires a framework that can be applied in many disease areas and in many scenarios. As such the primary goal of this research is to develop such a framework for data quality that assesses user motivation, which can be applied in web-based data management applications for clinical research across a range of (rare) disease areas.

To ground this work, an important point to consider is the range of disease registries that are selected for clinical research and the experiments that can be considered for web-based feedback. Every registry has different users and intentions for research. Users can be researchers, investigators or in some cases patients. Some registries work in a national setting, whilst others are based on international needs and demands, with associated language issues. Furthermore, research can be driven by data or based on specific research hypotheses. These dimensions are important to consider in how feedback mechanisms can work.

There has been a profusion of translational tools and data processing pipelines that have been created for clinical and biomedical research, yet there has been a limited amount of detailed assessment regarding the application of these to establish best practice and specifically, recommendations and practices that ensure that they meet the rigorous requirements demanded when translating research findings into clinical care settings. Translational clinical research faces problems of patient recruitment, evaluation of tools, lack of academic-industry cooperation, low proof of efficiency and effectiveness, and a need for training of researchers (Bornstein & Licinio, 2011). The field of basic science requires improvement for independent reproducibility of science which in turn depends upon simplified access to clinical data from increasing numbers of patients through targeted disease registries that include rich phenotypic information that contextualise any biological insights that might be made. The era of just dealing with information is nearly over, now it should focus on knowledge derived from data that can be used and trusted to improve clinical decision-making and care for the patient and the population more generally.

With the support of basic research, information pipelines and general frameworks must be developed to expedite knowledge transfer from the bench to the bedside to support improved clinical decision-

making (Cantor, 2012). Implicit in this is a need for improved data quality in clinical and translational trials.

The hypothesis of this work is that

"*web-based feedback should be an integral part of data collection processes in disease registries that demonstrably improves the overall quality of data that is entered into such registries*".

We explore this through a range of case studies related to a variety of disease registries utilising a variety of feedback mechanisms.

1.4 Significance

The first intended outcome of this work is to provide an overview of data quality and whether data quality feedback can be applied to clinical research scenarios or not.

Clinical research quality assurance practices will be explored and a data quality framework developed. This will explore and update guidelines that are now, we argue, outdated, e.g. the GCP-Guideline from 1996. Aspects of user motivation are tackled while exploring the interactions of data entry personnel, and we consider specific extrinsic and intrinsic motivational factors that can impact on data quality. Furthermore, data quality dimensions and their weightings and overall impact on clinical research will be investigated.

Multicentre trials about rare diseases benefit from the implementation of new technologies in clinical research, because clinical data management systems have several features to support data quality assurance and simplify data entry. However, not all funding schemes for clinical research support multicentre trials. This thesis will investigate the impact of funding schemes on research success, measured in publications with high bibliometric scores (impact factor or h-index). The outcome of that investigation will show the importance to enhance clinical data management systems and to support the research intentions of multicentre trials.

On a practical level, the intended outcome of this research is to explore data quality measures in the context of diseases, however we focus predominantly on rare diseases since they are the most demanding with regards to high quality data collection due to the rarity of the conditions and sparseness of such data more generally. In this research, we consider adrenal tumours and Niemann-Pick disease as key case studies and different aspects of rare disease related research (data driven vs. hypothesis driven). As noted, the measurements of data quality during clinical trials are rarely published or reported. Rather the literature provides overviews and review articles about the measurement of data quality and common inconsistencies, but little in the way of practical techniques to improve data quality. Through the application of data quality assurance mechanisms in disease registries and their impact in supporting clinical trials and related studies, we show and quantitatively measure the impact that these web-based feedback tools can make. We also consider more common conditions including type-1 diabetes mellitus and the role of the patient as the data entry personnel.

The third outcome of this thesis is the development of a framework that can be applied across a range of clinical research registries to improve data collection processes. This enables future studies to benefit from the mechanisms and framework proposed, and subsequently allows performance of other observational studies focused on an investigation of the effectiveness of data quality feedback in clinical data collection in inter-organisational settings. Data quality feedback is rarely researched, but all authorities describe the need for improvements in the quality of data to expedite the performance of clinical trials and studies and thereby reduce the costs of clinical research more generally.

1.5 Overview

In undertaking this research it is important to review possible background information that might impact on data quality feedback approaches. We consider related research in the field in chapter 2 and the extent that this work related to other fields. Building on this knowledge, a data quality feedback assessment exercise is considered based on several cohort studies to demonstrate the validity of the approach, i.e. does feedback work (chapter 3). This work focuses primarily on direct face-to-face feedback with multiple centres across Europe tasked with performing data entry into an international disease registry. To better understand the more detailed insights and working mechanisms of data quality feedback, a survey that investigates data processes and user motivation with regards to data quality processes is performed (chapter 4). This data quality feedback approach is considered in a research network based on a rare disease (Niemann-Pick Disease Types A, B and C and the international registry that has been established to support global research into this condition) to compare differences between research-driven and quality-driven research intentions (chapter 5). Chapter 6 focuses on the patients themselves and how they can be data providers through use of targeted mobile applications. We consider whether they are more diligent in data entry and ultimately provide better quality data. Chapter 7 investigates other motivational factors that can influence data entry and especially the quality of data entered into international disease registries. Specifically, we investigate whether centres involved in particular clinical trials associated with the disease registry have better quality data for those data items that are required for the studies associated with the registry. Ultimately we consider the overall impact of data processes and user motivation on data quality and whether web-based feedback can help improve data quality. We conclude the work as a whole and the extent that the research hypothesis has been met.

This thesis is structured into seven further chapters within four parts. These parts are: Introduction, Exploration, Application and Discussion/Conclusions (see Figure 1-5). Following the introduction, three chapters explore data quality feedback with various recommendations identified based on preliminary results (chapters 2-4). Following this, the gained knowledge will be applied to develop and explore a general framework to detect if web-based feedback can be applied to improve the quality of data in clinical research registries. We consider data quality evaluations with regards to weights and quality dimensions to gain more insights into data quality issues and user motivation (chapters 5-7). The final chapter contains the conclusions and suggestions of further research opportunities.

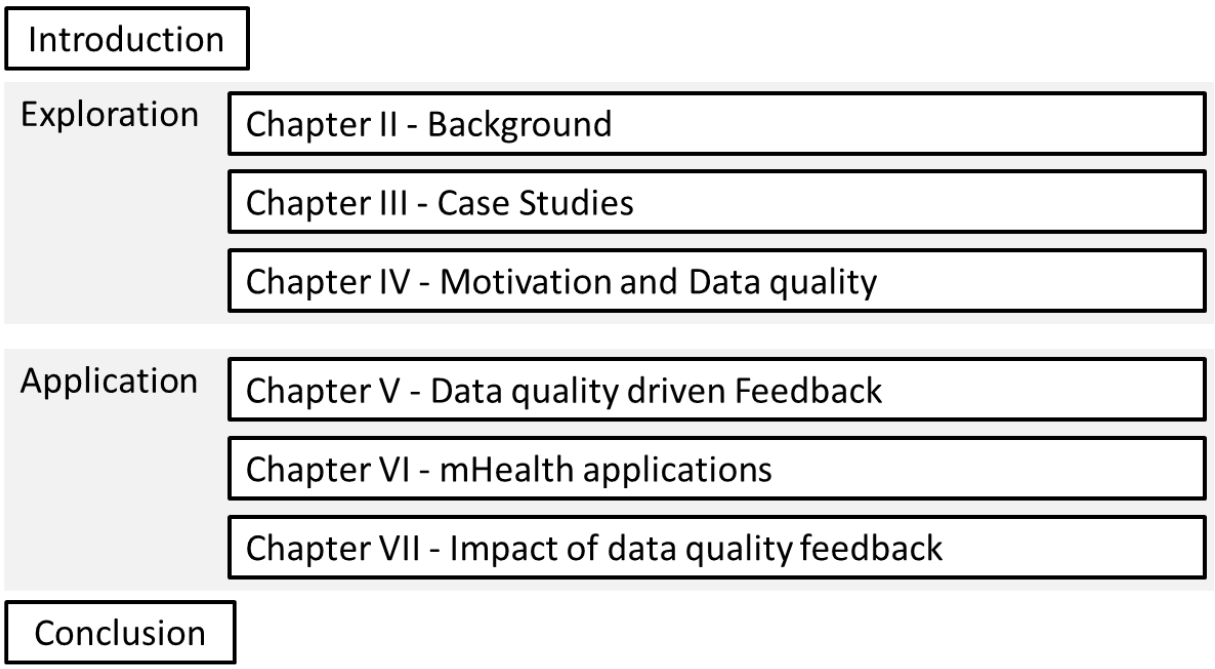


Figure 1-5 Thesis Structure

PART I

- **Introduction** – states the contexts, the motivation, hypothesis, research significance and structure of this thesis

PART II

- **Chapter 2** – gives an overview of Quality Assurance in Clinical Research. In addition, the associated methods to improve quality are discussed and the significance and impact of clinical research informatics identified. Related research to actions that might improve data quality in clinical trials is discussed. Methods like centralised data monitoring, double data entry or electronic data capture are discussed.
- **Chapter 3** – Focuses primarily on a demonstration that investigator feedback and community feedback can work within the setting of two major diagnostic cohort trials associated with an adrenal tumour disease registry.
- **Chapter 4** – Explores the challenges that must be overcome to support automated feedback. For this, a survey within the ENSAT-CANCER research network was performed to investigate user motivation in relation to data quality.

PART III

- **Chapter 5** – Focuses on applying web based feedback to support the data quality of a data-driven research registry. We show how usage through log analysis can identify user behaviour and patterns in access and use of a registry, and subsequently consider aspects that might improve data collection and data entry processes.

- **Chapter 6** – Considers the role of patients and empowering them to enter their own data. We focus specifically on use of mobile technologies within the context of a large diabetes related clinical study.
- **Chapter 7** – Explores data quality evaluations with applied knowledge to a range of data quality dimensions. We consider specifically whether actual clinical trials improve the data quality of research registries. To address this, we consider whether researchers are more motivated to undertake better quality data entry to disease registries if they are direct beneficiaries of the data being entered, i.e. they are involved in a clinical study that requires the data that is being entered into the disease registry.

PART IV

- **Conclusion/Future work** – draws conclusions from the thesis as a whole and identifies areas for future research.

2 STATE OF THE ART

Clinical research relies on information and data. Often such data is captured in clinical research registries. This chapter provides background on clinical research data management procedures, processes and mechanisms and introduces important aspects of clinical research informatics for this thesis. We describe how the quality of data captured in registries and their associated databases can have a huge impact on the outcomes of clinical studies and trials. We provide an overview of common quality assurance procedures adopted in clinical research with specific focus on those that focus on evaluation of data quality and the different data quality dimensions described in chapter 1. This chapter provides the contextual background that underpins the research conducted in this thesis.

2.1 Clinical Research Informatics

2.1.1 Introduction

The primary objective of biomedical research is to improve the knowledge, understanding and decision making of clinicians in clinical care. Ideally biomedical research should lead to guidelines that improve the diagnosis and treatment of patients with particular diseases. The transfer of research outcomes from basic science combined with the extraction of knowledge from clinical research is the primary objective of translational research. Translational biomedical research uses systemic approaches for hypothesis and knowledge generation. Such approaches commonly investigate the relationship and/or interactions between systems with one another, rather than for example, considering a single part of the body.

As an example, the capacity of the human brain is limited. A clinician typically creates a treatment plan based on around a small number of different facts that are discovered in the clinical phenotype (Stead, Searle, Fessler, Smith, & Shortliffe, 2011). Since the Human Genome Project completed in 2003 (International Human Genome Sequencing Consortium, 2004), more and more information for decision making is now available (see Figure 2-1). Data regarding structural genetics, gene expressions or information about proteomics should, depending on the disease, ideally be taken into account by clinicians. Today, it has been identified that around 1000 facts per patient typically needs to be considered (Tenenbaum, 2015). With the support of information systems and the increasing collection of human genotype and phenotype data, even more complex interactions can be studied through massive amounts of data supported by diverse modelling and analysis approaches (I. Buchan & Bischop, 2009). Translational informatics has the ability to deal with and expand the role of “big data”; improve individual decisions; increase capacity by offering learning health care systems with dynamic information transfer and take results and apply/embed them into a clinical care or research setting (Payne & Embi, 2015).

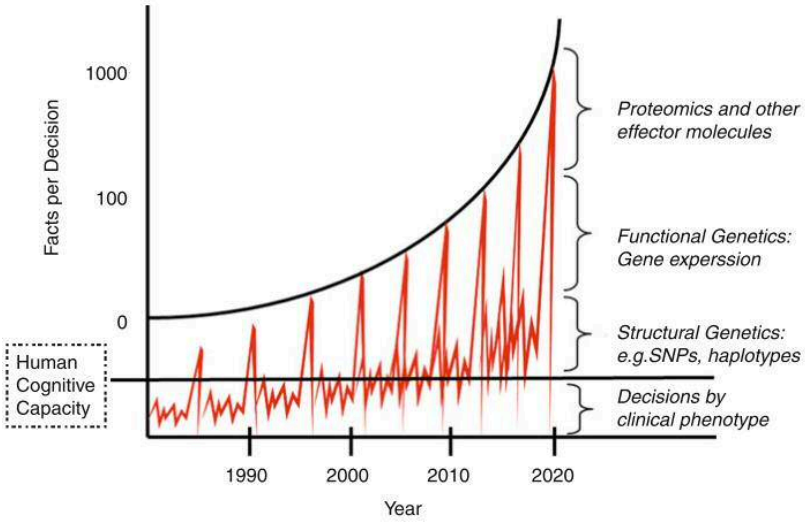


Figure 2-1 Clinical Decision Making (Tenenbaum, 2015)

Health research can be split into several different domains: the understanding of the human genome; the understanding of the underlying biology; understanding of diseases; the continually evolving advancements in the science of medicine and improvement of healthcare (Green, Guyer, & National Human Genome Research Institute, 2011). Genomics are fundamental data sources for the understanding of the biology and underlying basis of diseases that can lead to improved human health and patient management. It is now the case, that generating the genome is cheaper than the generation of knowledge derived from it (Border, 2011). However, there are still problems in transferring knowledge from one domain to another. Similarly, there are still weaknesses in data collection in biomedical trials and discoveries of novel molecules and their potential role in human health. These are often not transferred into the clinical domain due to the complex inter-disciplinary nature of biomedical research (Clinical Research Society, 2014).

These issues suggest that there is a need for mature, validated and interoperable tools, combined with processes and services to address this situation. Translational research, especially in the clinical domain, relies on the quality of collected data. Given this, it is crucial to develop tools that improve data quality and raise awareness on the research and research processes required to collect high quality data that is most useful for diverse clinical research efforts.

This research focuses on how automated feedback mechanisms for data entry and data review can improve this situation. It suggests a novel approach that can be adjusted during particular trials to deliver fundamental information to investigators at the onset of data collection and throughout the course of a given study. The hypothesis of this research is that automating real-time feedback on the quality of data can help improve this situation. To support this, a range of clinical solutions have been developed and used across a range of biomedical research areas.

2.1.2 Problem Statement

Since the beginning of the 21st century, the paradigms of translational medicine, research and informatics has continually evolved and are now fully established and widely adopted (Sung et al., 2003). Translational research moves away from general reductionism and the associated steps to system thinking (Payne & Embi, 2015). The main paradigm of the translational research cycle includes translational blocks (see Figure 2-2). Translational block 1 (T1) is supported by bioinformatics and translational block 2 (T2) is supported by medical informatics. This cycle covers knowledge transfer including evaluation of the effectiveness of the approach across both individual and population-wide health (Khoury et al., 2007).

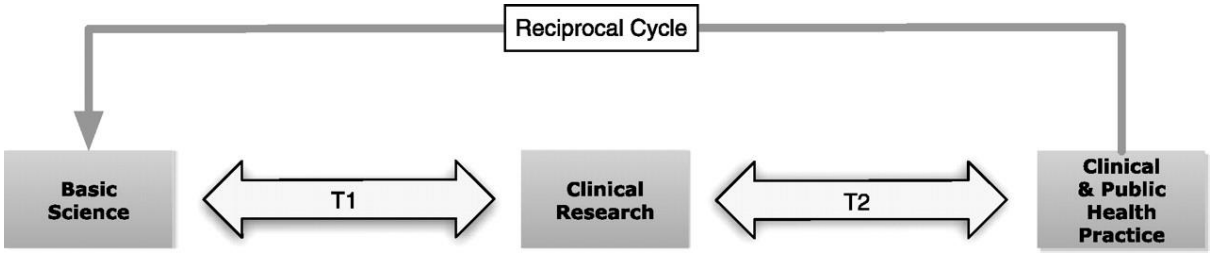


Figure 2-2 Translational Research Cycle (Sung et al., 2003).

The main problem this approach encounters is the slow transfer of findings from basic science into clinical research and beyond (Contopoulos-Ioannidis, Alexiou, Gouvias, & Ioannidis, 2008; Contopoulos-Ioannidis, Ntzani, & Ioannidis, 2003). Additionally this area often has inconsistencies of data, especially in multicentre collaborations where barriers exist for beneficial knowledge sharing of patient data with institutions and indeed transferring discoveries into industrial and/or commercial settings (Bornstein & Licinio, 2011; Kuchinke et al., 2010).

Translational informatics has the aim to increase the speed of newly developed medical entities from the lab to patients with specific diseases (Sarkar, 2010; Woolf, 2008). However it has been identified that over the last decade, the development of new treatments has *decreased*, but the money spent on research and development of new therapies has actually *increased* (Woodcock & Woosley, 2008). Furthermore, it was shown that new findings in genomics often shows no increase in late phase clinical trials (Clinical Research Society, 2014; Green et al., 2011). Currently an average drug development costs of the order of $1 billion USD and the whole process takes around 10 years. Only 15% of new therapies ever get into phase 3 trials and 50% of the novel therapeutics that pass this barrier actually get approved (Ledford, 2011). As such it has been identified that present clinical trials are woefully inefficient for drug development (Allison, 2012) and translational research has to evolve to tackle such problems (N. S. Buchan et al., 2011).

As an example, research in neuroscience showed a drastic shift of R&D funding in 2014 compared to 2009. Far less drug development programs (-52%) were processed by the pharmaceutical industry (Choi et al., 2014). This reduction of neuroscience drug development was caused by high drug costs and the extensive time periods that are taken until a drug actually gets approved (Brady & Insel, 2012). In 2011, the NIH (National Institute of Health) funded the National Center for Advancing Translational Science (NCATS) to counter these slow translational processes. The primary goal of NCATS was to expedite the transfer of newly discovered diagnostics and/or therapeutics to improve human health. It was identified that this should be achieved by promoting collaborative research and partnerships (Colvis & Austin, 2014) so that (inter-)national societies and/or collaborating institutions should be able to accelerate drug development processes. The ability to achieve this depends on both the financial support and importantly the way such support is delivered.

To obtain significant conclusions for new medical hypotheses especially in rare diseases, multicentre clinical studies are needed (Gatta, Capocaccia, Trama, & Martinez, 2010). Multicentre studies offer benefits of an increased sample size and associated improvements in the statistical power leading to better outcomes, however they must tackle the heterogeneity of the performance of the trial procedures to ensure an overall improved generalizability of the results (Guthrie et al., 2012). However multicentre clinical trials also face challenges: the heterogeneous performance of trials can be caused by changes of study protocols or complex protocol designs and diverse patient information that is collected within the individual institutions (Brandt et al., 2006). Other problems include data propagation; errors through wrong randomization or in some cases erroneously calibrated equipment can make a multicentre comparison of values difficult (Venet et al., 2012). A literature review also showed that Randomized Clinical Trials (RCTs) in single centre settings have larger treatment effects,

than intervention trials in multicentre settings (Dechartres, Boutron, Trinquart, Charles, & Ravaud, 2011).

Despite these problems, every genomic discovery and associated new medical development must be proven in a late phase clinical trial to show evidence of the effectiveness and efficacy of the new invention (Green et al., 2011). To prove this, a comparison trial against the gold-standard test or a standard of care treatment *vs* a placebo must be conducted (Vaidyanathan, 2012). In short, there is a great need for faster translation of research and trials into the clinical environment. New methods in clinical research must be cheap, fast, adaptable and more in tune with genomics (Ledford, 2011). To support this, electronic health records (EHR) or phenotypic information must be observed in relation to genotype data, and biomarkers must be used on study populations that can subsequently be shown to lead to new stratified study designs for biomarker trials (Jensen, Jensen, & Brunak, 2012). Translational blocks (T1 & T2 and above) urgently need improved data validation processes to achieve rapid knowledge transfer between the lab and the patient (McShane et al., 2013).

An important step to consider is the form of data collection. While -omics analyses are mainly driven by data sets from targeted machines, e.g. spectrometers, clinical trials are predominantly manual data entry methods. Remote Data Entry (RDE) or Electronic Data Capture (EDC) are used in many academic institutions (ECRIN, 2007), however there is no evidence as to whether these methods provide an advantage in reducing errors (Nahm, Pieper, & Cunningham, 2008). Organizations/institutions need better knowledge about the conduct of trials and the user interfaces used for data entry (K. Chen, Chen, Conway, Hellerstein, & Parikh, 2011). Furthermore, since double data entry (DDE) is often too costly, it has been suggested that instead of DDE, randomly selected records must be checked (ECRIN, 2010). Finally, the use of new mobile devices for data entry is considered to improve the data quality and motivation of the staff to enter more and better quality data more often, but has not shown significant beneficial results (K. Chen et al., 2011; Paulsen, Overgaard, & Lauritsen, 2012; Walther et al., 2011).

Human motivational affordances are a very important aspect for the consideration of the improvement of data entry and subsequent improvement of data quality. Besides feedback tools, this research presumes that the reason for low data quality can be due to low staff motivation, and especially those staff actually undertaking data entry. To increase user motivation it was shown that gamified processes in non-game contexts were useful but not well studied within the biomedical domain (Deterding, Sicart, Nacke, O'Hara, & Dixon, 2011). Adding features of community competitions, badges and achievements through web-based data entry, may increase the drive of the user for improved contributions and better quality data entry (Hamari, Koivisto, & Sarsa, 2014).

2.2 Data Quality Assurance Dimensions and Statistics

The term 'quality' is often used in different kind of contexts. It could be used as label for an item, a final mark of a review or a description of lifestyle. Whilst "Quality-Food", "Quality-of-Life" and "Quality-Time" have different usages, but one common purpose: to describe an aspect that something is better than something else. Quality doesn't explicit refer to something expensive, rather it describes whether there is a balance between the usefulness of an object and the related costs/resources that are spent. Quality can also be construed as an attribute that describes an item or process as fit for its purpose.

An accepted general definition of quality is from Joseph Juran, who defined quality as the "Degree to which an inherent characteristic fulfils requirements" (Juran & De Feo, 2012). In regards to the medical domain, the Institute of Medicine (IOM) defines quality "as the degree of which health serves increase the likelihood of desired health outcomes and are consistent with current professional knowledge" (American College of Medical Quality, 2010). Clinical research relies on data captured during clinical

trials, so collected data must represent and be 'fit' for the needs of the scientist (Abate, Diegert, & Allen, 1998).

Another important aspect is that quality is assumed to describe different attributes of a product or trial: design, conformance and useability (Ottevanger et al., 2003). This indicates that all users who are in contact with the product need to participate in the development on the product.

"Quality is everyone's responsibility." W. Edwards Deming

In 1950, Deming adapted and improved a method for quality control from Walter A. Shewhart for consistent quality improvement. Through an iterative process, he provided the interaction of four domains: design (plan), production (do), sales (study) and research (act) (Moen & Norman, 2010). This process is called the PDSA or Deming-Cycle (see Figure 2-3) and due to the general approach proposed for quality improvements, it can be applied to health care as well as many other domains (Deming, 2000). In general, this concept checks the current status of a product or process and compares it with the standard that should be or was achieved, i.e. it is a comparison of a target and the current actual performance.

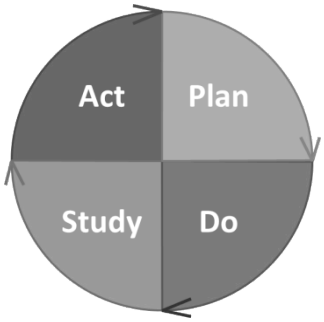


Figure 2-3 Deming Cycle (PDSA)

Furthermore, Deming defines product dimensions: design, development process, time of delivery and price. These dimensions have a huge impact on what the customers will spend to buy or use the product being developed. The Deming Ratio of quality is the ratio between effort and cost:

Equation 2-1 Quality Equation

The larger the usefulness of an item or process and the lower the resources, the higher is the quality. On the other hand, most people expect that something has high quality only when the most expensive available resources are used (implemented). Quality is a term that everyone describes differently, but for quality assurance purposes, high value is easily explained by the ratio of quality and resources (see Equation 2-2). It is important to gain the best value while investing the least possible resources.

Equation 2-2 Value Equation

In conclusion, usefulness equals value. This thesis assumes, that usefulness in clinical research informatics is data that can be used directly for research-oriented clinical trial objectives as well as for other research purposes (in protocol amendments, follow up trials, new clinical studies). Resources in this context are the efforts of all human interactions (software engineers, data managers, principal and local investigators, data entry users) that directly influence data quality. Factors that have

negative impact on data quality will be explored in chapter 3. In general, these factors are explored with the definition of quality indicators.

2.2.1 Data Quality Indicators

Avedis Donabedian provided an insight of quality indicators in health care that influence the quality of a product and its development processes (Donabedian, 1988). He separated quality indicators into three categories: structure quality, process quality and outcome quality. To detect possible indicators for the quality of an object, all categories must be investigated. With the identified problems and resources, improvements can be planned and executed, e.g. in a PDSA Circle (Donabedian, 1988).

The requirement for a Data Quality framework in clinical research needs to address the following aspects:

- Active improvement of quality from every user (not only management);
- Assess quality levels and improve data processes (Donabedian), and
- Ongoing improvement of data during the data collection process (Deming).

Nonnemacher, Nassek, & Stausberg (2014) adapted the Donabedian model for data quality in cohort registries and clinical research registries. The authors applied categories of structure quality, process quality and outcome quality into the monitoring and data management setting. They labelled structure into data 'plausibility' (which was renamed in the 2nd data quality version to 'integrity'), processes into 'organisation' and outcome into 'correctness'.

The researcher defined for each of these levels, several quality indicators with a definition, source of literature, clinical context, comments and the ratio, i.e. how the indicator could be quantified. Table 2-1 provides an overview of all quality indicators of the 2nd version assigned to their recommended weight and threshold. Weights are scaled from 1-9, where 9 reflects the highest importance to be considered for the evaluation of data quality in cohort studies. Thresholds provide an overview of the range of calculated values the quality indicator should achieve. For example, overall data accuracy has to be around over 95%.

Table 2-1 Quality indicators in cohort registries (adapted from Nonnemacher et al., 2014)

	Weighting	Threshold
Plausibility/Integrity (Structure quality)		
Accordance to a reference value	1	> 5%
Concordance	1	> 5%
Consistency	3	> 5%
Records without follow up information (infinite living patients)	2	
Implausible entries	2	
Possible implausible entries (alerts)	1	
Values distribution	1	> 10%
Data fraud (detection of wrongly entered values)	1	> 10%
Value distribution of parameter (medical devices)	1	
Value distribution of parameter (examiner)	2	
Value distribution between results	1	
Value distribution between centres	2	
Examinations at the weekend and public holidays	1	> 2%
Missing records	1	
Missing record items	2	
Missing compulsory items	3	> 5%
Missing optional items	1	> 5%

Unknown items	1	
Items with non-related item entry (e.g. ‘unknown’, ‘na’)	2	
Outliers in biochemical and diagnostic procedures	2	> 0%
Values that are out of range	1	
Values of standards	2	
Unpermitted values of qualitative items	1	
Unpermitted values of qualitative items for coding and missing records	1	
Unpermitted values for coding and missing records	1	
Unpermitted values of quantitative items for coding of outlier of ranges	1	
Items with non-specified units	1	
Observations with unknown primary tumour	2	
Evidence of known correlations	1	
Amount of metadata of examinations	1	
Organisation (Process quality)		
Timeliness	6	
Double entered items	4	> 5%
Recruitment rate	5	< 90%
Death certificate rate	6	
Refusals of examinations	4	
Refusals of records	4	
Refusals of single items	4	
Early cleaned data items	5	> 1%
Synonyms	4	
Homonyms	4	> 5%
Items per record	4	>2%
Reports of pathologists	5	
Rejected reports	5	
Data sources per record	4	
Count of Follow-up data	6	
Correctness (Outcome quality)		
Data accuracy	8	< 95%
Results of SDV regarding items	9	> 5%
Results of SDV regarding records	8	> 5%
Completeness	7	< 95%
Accordance to SOPs and study design	7	> 5%
Representativeness of items	7	< 100%

Here, the data quality dimensions of accuracy and completeness have the highest ratio of threshold and weighting, which indicates the importance of their evaluation for data quality in clinical research/cohort registries. Other important quality indicators of data quality are recruitment rates and results of Source Data Verification (SDV). In version 1.0 of this approach, the professional qualification of data entry users was also part of the process quality with a threshold of < 100%. All these indicators provide the opportunity to calculate an overall data quality score (as described in Equation 1-2 on page 9). However, not all indicators can be applied directly to clinical research registries. It is important to consider that indicators like death certificate rates are only compulsory for cancer registers and not as important for other clinical research registries. To improve data quality as part of a rigorous ongoing process, these indicators must be addressed in every data quality assurance activity throughout the course of a clinical trial. The research group of Nonnemacher et al. (2014) clustered the final data quality score into 5 groups with 100%-81% receiving the grade ‘*very*

good', 80%-61% - '*good*', 60%-41% - '*moderate*', 40%-21% - '*poor*' and below 20% receiving the grade '*very poor*'.

To obtain *very good* data quality grades, ongoing data quality procedures need to be implemented in the quality assurance activities of the study.

2.2.2 Data Quality Assurance

Since the Nurnberg Code 1947, clinical research and research ethics have been jointly developed. The primary reason for this was that after World War II, it was identified that subjects have the fundamental right to be protected against involuntary participation in medical experiments. It is now widely accepted that it is essential that the enrolment of a new patient in the clinical trial is entirely voluntary. This is accomplished when the subject has had enough time to think about participating after reading the informed consent and consequences of being involved in a given study. The Declaration of Helsinki refined the clinical research ethics in accordance to "informed consent based on prospective research design, independent protocol review, confidentiality, and data integrity (Rock, Molloy, & Humphrey, 2010).

Quality assurance for trials is regulated in the Good Clinical Practice (GCP) guidelines. These guidelines include a variety of sections to ensure patient safety and data integrity. GCP was published and declared compulsory for all clinical trials in 1996. To understand the impact of GCP interactions with trial quality assurance (QA), quality control (QC) and monitoring, the following sections provide an overview of the key definitions and suggested practises.

Quality Assurance

Quality Assurance (QA) comprises all actions that improve data. It is defined by:

> *„All those planned and systematic actions that are established to ensure that the trial is performed and the data are generated, documented (recorded), and reported in compliance with Good Clinical Practice (GCP) and the applicable regulatory requirement(s)."* (International Conference on Harmonization, 1996)

In case of an independent audit (external), auditors must have the possibility to gain insights into all performed QA procedures that involve data management processes. All documents (CRFs, study protocols, amendments) should be accessible and QA actions need to be traceable (Campbell & Sweatman, 2002).

Quality Control (QC)

The QC definition in the GCP-Guidelines states:

> *„The operational techniques and activities undertaken within the quality assurance system to verify that the requirements for quality of the trial-related activities have been fulfilled."* (International Conference on Harmonization, 1996)

Therefore, QC is an instrument of QA. QC is more an active action of trial personnel involved in data-related processes. These staff members carry out actions to improve data quality and data integrity. As such, audits of every action used to assess data quality is not part of QC, but it is part of monitoring procedures (Campbell & Sweatman, 2002).

Monitoring

Monitoring as instrument for the evaluation and assessment of data quality is defined as:

„The act of overseeing the progress of a clinical trial, and of ensuring that it is conducted, recorded, and reported in accordance with the protocol, Standard Operating Procedures (SOPs), Good Clinical Practice (GCP), and the applicable regulatory requirement(s)." (International Conference on Harmonization, 1996)

In short, monitoring as a QA tool focuses on controlling and assessing the actual information (data) quality.

A major concern for quality in clinical trials is data monitoring. English, Lebovitz & Griffin (2010) noted in a cost breakdown that 30% of all trial costs are spent on monitoring purposes to improve the quality of collected clinical data (R. Califf, 2009; English, Lebovitz, & Griffin, 2010). Site and data monitoring has drawn strong criticism because of their high inefficiency (Baigent, Harrell, Buyse, Emberson, & Altman, 2008). The FDA recommended using a risk-based approach to monitoring that used centralized techniques to detect data errors, that can subsequently reduce overall trial costs by 20-25% (Franco et al., 2013) and detecting 90% of data issues (J. M. Bakobaki et al., 2012; U.S. Department of Health and Human Services et al., 2013). A key question is whether the high cost is incurred to detect around 95% of all issues or whether a lower cost and a data error detection ratio are feasible for conducted clinical trials? Ideally we would like 100% of data issues to be identified, but if this comes at a cost that is unrealistic, then what rate of errors might be acceptable?

Monitoring in clinical trials is a good example about issues that appear during the performance of medical research. If data monitoring can be performed more efficiently and utilising generally accepted standards, subsequent problems can be reduced and transferral of knowledge to other domains can be expedited. It is essential to apply general accepted standards in healthcare and clinical research with regards to data quality frameworks. Bespoke solutions are unlikely to work in the large.

2.2.3 Data Quality Dimensions

'Quality' is a multidimensional term. In industrialised nations, customers generally want more quality of a product or service for less money. Thus quality is often associated with cost. A customer wanting to purchase a mobile phone has to answer the question, whether he/she is interested in buying a product with high costs and (one assumes) high quality or a product with lower costs and (one assumes) less quality. What is the intention and motivation of buying this product? What features are okay to dismiss for the purpose of saving money?

Quality in healthcare settings has exactly the same intention. Implementing quality control methods that incur extensive resources often leads to high quality (in most cases). However, investing fewer resources can often provide only slightly lesser quality than the higher cost version. Is it possible to find a balance? To answer this question, it is mandatory to evaluate the actual quality and compare it with the targeted quality (target-actual comparison). Following this, the cost for implementing quality control methods must be calculated, to assess if quality assurance of the product or process is worth the actual overall improvement obtained.

Every quality dimension should capture different aspects. The combination of quality dimensions will give a judgement about data quality. Not all dimensions are useful in every project or completely generalizable in a single framework. However, there are concepts that support the development of quality aspects and overall quality assurance with regards to data quality. Data quality dimensions should be separated from data models (schemata) during the development of an information system (Batini & Scannapieco, 2006).

In a review by Bray & Parkin (2009) several key criteria for the evaluation of data quality in cancer registries were explored. These criteria included *Completeness, Comparability, Timeliness and Validity* (Bray & Parkin, 2009; Parkin & Bray, 2009). However, as explained in the introduction, cancer registries

differ to research databases in several aspects regarding data collection including the purposes of the registry itself. While cancer registries often have several data sources and focus on public health reporting, health protection, grading of prevention and treatment and epidemiological research (Hentschel et al., 2011), research registries typically focus on achieving an answer to individual research hypotheses. In their review, registry completeness is mainly used for the calculation of incidence and survival rates. However, registry completeness in trial databases have a different priority to cancer registries. The described dimensions are equally adaptable to research registries. Of high importance for clinical trials are data validity including data completeness, data plausibility and data accuracy (Arts, De Keizer, & Scheffer, 2002).

Weiskopf & Weng (2013) proposed similar dimensions in the clinical research domain. They performed a literature review focused on the reusability of EHRs and they identified five criteria for the assessment of data quality: *Completeness, Correctness, Concordance, Plausibility and Currency*. In their study they determined major terms of data quality dimensions with correctness equal to accuracy and currency equal to timeliness. Additionally, the authors assessed methods of how the data quality dimensions could be calculated. They identified seven main methods: gold standard tests (registry data compared to sources other than the original source data), data element agreement (random items selected and compared with the EHR), elements present (completeness of data elements), data source agreement (data compared to the source data), distribution comparison (statistical analysis compared with key clinical concepts), validity checks (application of item rules) and log reviews (analysis of data entry practises). Table 2-2 shows the total number of review studies clustered in quality dimensions and evaluation methods (Weiskopf & Weng, 2013). They identified that data completeness and accuracy are commonly used dimensions for the evaluation of data quality in EHRs. It is interesting to note the generally low count of log-review methods used for the assessment of any data quality dimension.

Table 2-2 Studies related to data quality dimensions and evaluation methods (adapted from Weiskopf & Weng 2013)

Dimension/ Method	Completeness	Correctness	Concordance	Plausibility	Currency	Total
Gold Standard	24	35				37
Data Element agreement	8	17	7	2		26
Element presence	23					23
Data source agreement	4	1	6			11
Distribution comparison	4		3	4		10
Validity checks		5		2		7
Log review		1			4	5
Total	**63**	**59**	**16**	**8**	**4**	

Other authors referred to *correctness*, *plausibility, organisation* (Nonnemacher et al., 2014) and *reliability* (Hong et al., 2013; Le Jeannic, Quelen, Alberti, & Durand-Zaleski, 2014; Rostami, Nahm, & Pieper, 2009). The European Clinical Research Infrastructure (ECRIN) standard indicates that every

centre needs to perform quality processes to evaluate *accuracy* and *consistency* for data collection (ECRIN, 2013). In 1996, Wang and Strong published a data quality framework for general information systems with the following dimensions: *completeness, accuracy, timeliness, ease of understanding, reputation, reliability, availability, price, (representational) consistency, response time, relevancy* and *amount*. Every dimension was labelled with a scale, e.g. reputation with 1 (bad) to 10 (very good) (Naumann, Leser, & Freytag, 1999; Wang & Strong, 1996).

In the following sections, important dimensions of biomedical data quality will be explained and discussed. These dimensions include *completeness, accuracy, comparability, reputation, timeliness* and *eligibility*.

Data Completeness

Generally, data completeness (C) is measured in combination with data accuracy (validity) checks. Scores are typically calculated as a quotient of the relation between the sum of incomplete items (I) compared to all items (Batini & Scannapieco, 2006).

Equation 2-3 Completeness Calculation (adapted from Batini & Scannapieco, 2006)

$$C = 1 - \frac{I}{\text{all items}}$$

Completeness is a quality dimension measured in many trials and data management systems performing data quality assurance. It is recognised that high data completeness shows strong evidence that trial results are useful (and truthful). However even if an item is complete (e.g. filled out in a web form), then this doesn't necessarily mean that the item is correct. Other dimensions need to be applied to check whether the data is actually correct and useful.

Data Accuracy

Batini & Scannapieco (2006) define inaccuracy as a state such that "the information system represents a real world state different from the one that should have been represented" (Batini & Scannapieco, 2006). Arts, De Keizer & Scheffer (2002) provide a simplified definition in the context of clinical research, where they define accuracy as "the extent to which registered data are in conformity to the truth" (Arts et al., 2002). According to the data quality definition of Wand & Wang (1996) data can be incomplete, ambiguous, meaningless or incomplete (see Figure 2-4) (Wand & Wang, 1996).

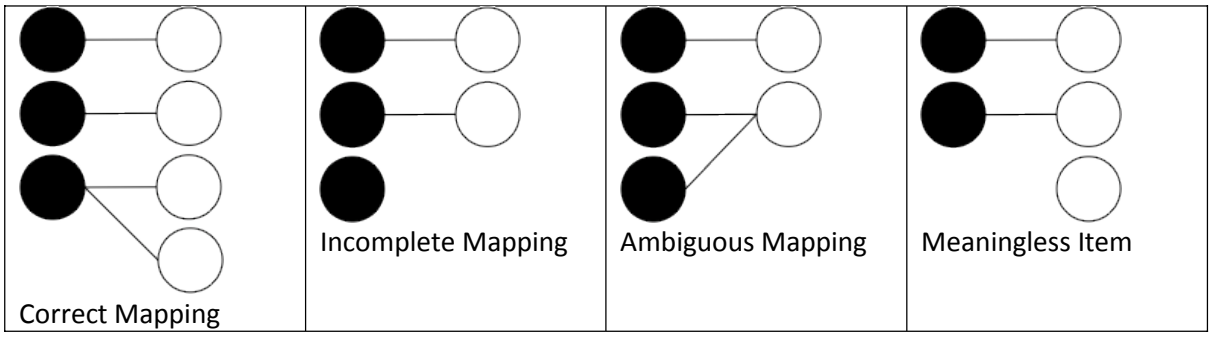


Figure 2-4 Data Quality Representation Deficiencies

Even if the mapping of an item is correct, it doesn't necessarily imply that the value actually represents the truth. Accuracy must include dimensions of timeliness, completeness, reliability, consistency and validity (Batini & Scannapieco, 2006). Accuracy can be distinguished between internal (related to the data) and external (related to the further use of the data) views (Wand & Wang, 1996). Applying range checks on values is one method to improve the internal aspects. Checking if the item represents the most recent value is an external view characteristic.

Arts, De Keizer & Scheffer (2002) performed a case study where they performed a Source Data Verification (SDV) to evaluate the completeness and accuracy of data in a research registry compared to information in hospital records. For every data process (data entry, data extraction, data transfer) they calculated an accuracy score and a completeness score. Accuracy was based on whether the values of the registry were shown in the paper records or the selected category (e.g. yes/no) was correctly selected (Arts et al., 2002).

SDV can be very expensive and time consuming (Tantsyura et al., 2010; Venet et al., 2012). Data accuracy checks without availability of electronic source data is reduced to range checks and cross validation methods. Venet et al. (2012) suggested an approach of evaluating data quality in trials where paper based sources (pCRFs) are in use. The group addressed four main errors: abnormal, infrequent, fabricated and falsified data values. They identified these errors with statistical analyses these errors could be readily detected (Venet et al., 2012).

Data Comparability

For international multicentre trials, every study site should have procedures and processes of how to implement standardized operation procedures (SOPs) for routine care and associated data management. To harmonize SOPs between centres, on-site data monitoring can be performed to investigate local processes and subsequently discover issues that may lead to low data quality. A statistical evaluation provided by clinical data management processes should ideally support visits to reduce costs and time efforts. This approach must also be transferable to other situations and data comparability checks.

Comparability as a data dimension is known in the literature by the terms interoperability, agreement, consistency, reliability and variation. All are collected under the topic of concordance (Weiskopf & Weng, 2013). Guthrie et al. (2012) describe a quality control approach of collected values using the calculation of the intra-class correlation coefficient (ICC). With ICC it is possible to detect performance differences between centres in multicentre clinical trials. In a case study, the authors calculated the ICC and detected 2 significant performance differences. They reported the ICC in a table and looked at the identified values in box plots to easily detect the centre with the highest variations in their measurements compared to other centres (Guthrie et al., 2012). Calculation of the ICC can be used to detect data clustering, e.g. if the same physician performs a measurement more than once. This dependence can be skipped with the use of the residual ICC (RICC). The RICC can provide a direct percentage or degree of data clustering (Wynants et al., 2013).

ICC and RICC can detect data clustering and are therefore very useful for investigators during clinical trials to detect data inconsistencies and subsequently plan actions to reduce them (Guthrie et al., 2012; Wynants et al., 2013).

Reputation

Data reputation is also found in the literature as data provenance or data trustworthiness (Batini & Scannapieco, 2006; Dai, Lin, Bertino, & Kantarcioglu, 2008; Wang & Strong, 1996). Data standards demand that datasets are exchangeable particularly in the biomedical domain where information must be rapidly transferred from basic science to clinical trials and beyond (Richesson & Krischer, 2007). Data provenance is define as “description of the origins of a piece of data and the process by which it arrived in the database” (Buneman, Khanna, & Wang-Chiew, 2001). With a trust score, the scientist can choose whether they wish to use the data directly or send data inquiries to the data provider (Dai et al., 2008).

Nurse, Agrafiotis, Goldsmith, Creese, & Lamberts (2014) explored an approach to measure data trustworthiness of online information using sources such as Facebook, Twitter and Google. The

authors developed a calculation that was dependent on the source competence, the timeliness of the information (*recency*), the user closeness to the event/information (*proximity*) and integrity/quality/provenance of the source. Every item was assigned with a weight to calculate a trustworthiness score. Other authors discussed radar graphs, traffic lights and stars to display the trustworthiness of data for users that need to "make informed decisions" based on the data (Nurse et al., 2014).

In biomedical research this score is effective when methods like remote data entry or electronic data capture are in use. In the case of manual data entry, if the data source is electronic, then the original data can be checked with the entered data directly (at least in principle).

Timeliness

Time-related dimensions like timeliness, currency or volatility provide information on whether data is captured and/or updated in (near) real-time or more generally, if the data represents the most recent values (Batini & Scannapieco, 2006; Lane, Heddle, Arnold, & Walker, 2006). Batini & Scannapieco (2006) provide definitions for terms of time-related dimensions. These include:

- Currency is "the sum [of] how old data are when [they are] received (Age), plus a second term that measures how long data have been in the information system (DeliveryTime – InputTime)" (Batini & Scannapieco, 2006)

Equation 2-4 Currency Calculation (adapted from Batini & Scannapieco, 2006)

- Volatility is "the length of time data remains valid" (Batini & Scannapieco, 2006) and
- Timeliness represents a value of "how current data [is] for [a] task" (Batini & Scannapieco, 2006).

Equation 2-5 Timeliness Calculation (adapted from Batini & Scannapieco, 2006)

Here a value of 0 represents poor timeliness and a value of 1 represent good timeliness (Batini & Scannapieco, 2006).

In the biomedical domain, timeliness scores often represent efficient data capture. The more time between the original data collection and that data entry into a database for example, results in a lower data quality (Richesson & Andrews, 2012). For clinical data management, this score is often combined with a schedule feature that reminds the researcher of certain events and provides indications/notifications to improve timely data entry, e.g. in the biomedical domain, follow up visits of a patient (Brandt et al., 2006). Established tools can provide warnings of pending values or remind researchers that special events may occur in the future (Herzberg, Rahbar, Stegger, Schäfers Michael, & Dugas, 2011). Generally, a cut-off will be set by the data scientist to classify the data as either timely or untimely (Mathers, Fat, Inoue, Rao, & Lopez, 2005).

A narrative literature review of RCTs found that there is an increased efficiency in timeliness of data capture by hand-held computers compared to paper-based approaches (Macefield, Beswick, Blazeby, & Lane, 2013). In the literature, timeliness is reported as either a score, a percentage or a histogram, where it is correlated with completeness or count of errors (Hills et al., 2010; Krzych et al., 2011; Laberge & Shachak, 2013; Mathers et al., 2005; Weiskopf & Weng, 2013). This indicates that timeliness

can always be used in combination with a logging information approach that provides time stamps for entered, changed or deleted values/records in a research registry.

Record Eligibility

Nahm (2012) argues that there is no data quality framework that can be labelled with 'one size fits all'. The reason for this is the heterogeneity of statistical analysis caused by data inconsistencies (Nahm, 2012). Eligibility is a data dimension that is hard to measure. Every clinical trial defines different requirements that need to be matched by all study participants. These criteria are not only the inclusion and exclusion criteria of subjects, but they can also include the minimum required information for the study objective. A proposal for computable protocol-driven research was presented in Sim & Niland (2012).

There are two main approaches for confirming clinical records as eligible for a trial. Firstly, compulsory data items can be defined for an ongoing trial and/or they can be defined in the set up of a new study for eligibility criteria to help find records in an existing data registry that can be re-used. Both purposes are used for computer-based clinical trial recruitment (Köpcke, Kraus, et al., 2013; Köpcke, Trinczek, et al., 2013).

During an ongoing clinical trial, confirmation of record eligibility and defining related eligibility scores to records is an indicator of how much compulsory information has been entered for trial subjects. It is the responsibility of the investigator to determine whether mandatory trial items should be weighted. Through such an approach, tools can support users and/or investigators, e.g. by error flags and warnings (important data is missing) or reminders (less important data is missing).

In contrast to Nahm (2012), this leads to the conclusion that in every trial the assessment of data completeness and record eligibility is mandatory to monitor trial performances.

Data Thresholds

Another key aspect for quality assurance and subsequent accreditation of a quality score is the definition of thresholds of items and the weight of the assessed quality dimensions. With regards to thresholds, this limit gives an indicator as to whether the quality of a selected item is adequate or insufficient. It is suggested that this threshold must be defined for all values in a data repository (JCAHO, 1990).

Little research has been done that shows whether thresholds on quality scores/scales are actually effective. In the literature, thresholds are usually subjectively predefined values or limits calculated by linear regression models to predict the likelihood of an event, process or decision procedure. They are typically defined a priori to a trial although they can be dynamically adjusted during the trial (Bacchieri & Della Cioppa, 2012; Batini & Scannapieco, 2006; R. M. Califf, Karnash, & Woodlief, 1997).

The (important) role of thresholds is related to the motivation of individuals in reaching particular goals. Goals can be part of games, or in the context of biomedical research as gamified objects. A goal can encourage a user to reach a certain percentage of data quality (Deterding, 2012).

How this goal (score) is actually calculated can be a generic requirement in the medical context and subsequently used for auditing purposes. Additionally, we argue that this threshold can serve as motivation for the user to achieve better data quality scores.

Data Quality Scores

Generally, scores are used in the medical domain as a grade. This grade is often an arithmetic mean of at least two components that support the assessment of an item (Holle, 1995). In many situations it is a combination of considered quality dimensions, e.g. accuracy and completeness (Wang & Strong, 1996). As described, not every score needs a simple threshold. Some scores need a weight, e.g. whether it is more or less important for the selected research hypothesis. Nonnemacher, Weiland & Strausberg (2007) used in their evaluation of data quality in cohort trials a scale of values between 1 and 9 (Nonnemacher, Weiland, & Stausberg, 2007). They considered that data quality dimensions can be distinguished into four different aspects:

(a) intrinsic aspects, that states that data quality needs to achieve a high amount of accuracy and completeness in its own right;
(b) contextual aspects, that describe data quality itself and considers quality of data in the frame of the research objectives;
(c) the representational aspects, that describe the format (concise and representation) and meaning (interpretability and ease of understanding) of the data, and
(d) accessibility data quality aspects, that define data as a construct that should be sharable and exchangeable between domains, with the highest standards of privacy and security (Wang & Strong, 1996) (see Figure 2-5).

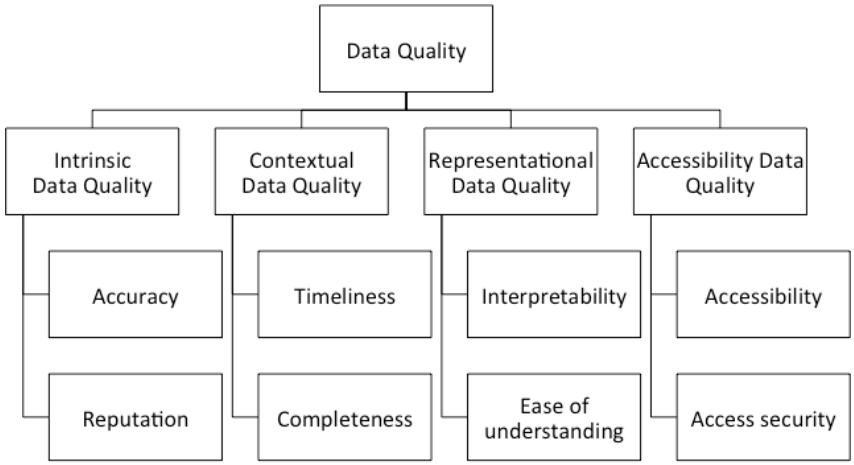


Figure 2-5 Data Quality Dimension Framework (adapted from Wang & Strong, 1996)

Conclusions

Given all of these possibilities of data quality dimensions that need to be considered, a general approach for clinical trials cannot be overly prescriptive. Every trial needs to select their own criteria that are useful to measure data quality that is specific to that trial. This research considers criteria that are possible to capture, measure and report during data entry through web-based feedback. Data quality dimensions need to be reported (feedback) to three different users: the data entry user, the (principal) investigator and the research community. The data entry user needs data quality feedback to achieve better quality, to avoid data cleaning at the end of a trial. Literature shows that feedback of data completeness and data accuracy is usually reported in research papers after a clinical trial has concluded. However, web-based databases and associated information systems now have the possibility to capture, show and we hypothesise, improve this data quality assessment in real-time. Data completeness and data accuracy can be reported during data entry as real-time feedback.

Procedures for data monitoring can also be performed, e.g. centralized monitoring, that can be reported to staff members and investigators on a regular basis. Investigator feedback must include all-important information so that an investigator can start interventions to improve local data processes to collect higher quality data earlier in the trial lifecycle. Important examples are eligibility scores and timeliness scores. Finally, community feedback must include all information needed to challenge the user to be better than others and to set up individualised motivations to improve data quality. Examples considered include leaderboards for overall data quality rankings with suggestions on how to increase data quality scores targeted to the individuals.

To implement feedback tools in data management systems, it is important to investigate what procedures are currently used in clinical trials and what specific requirements new tools should address.

2.3 Web-based Feedback Systems

Participants in clinical trials have to sign and understand the consequences of their involvement through clinical research consent (or parental assent). One part of this consent includes information about what happens with information collected during the trial. In most cases, this information is pseudonymised, i.e. every subject is assigned with a unique computer-generated identification code that only internal staff at the specific centre where the patient was recruited can use to subsequently re-identify the subject. The patient information often needs to be provided to other researchers outside of the immediate healthcare/research unit. The issue here is that local data is primarily stored in paper-based records whilst data that is exchanged between centres is stored in digital repositories. It is generally known, that data transfer from paper-based records to electronic records results in a loss of accuracy and completeness (Arts et al., 2002; Glöckner, Arlt, Bancos, Stell, & Sinnott, 2015; Le Jeannic et al., 2014). To increase the usefulness of captured data for clinical research, data quality feedback is needed to improve the transfer of such information into a translational research context and thereby provide information that can be used to draw statistically significant conclusions in clinical research.

In 2002, Simon De Lusignan and co-workers investigated the impact of feedback on data quality in medical records. In their study they defined ‘useful data quality markers’ and investigated whether feedback increases these marker’s statistical significance. Their literature review concluded, that feedback had been used to improve “clinical outcomes rather than on changes in data quality” (S. de Lusignan, 2002). They discovered that just 3 of their 10 defined markers improved. These items were the READ code, prescriptions linked to diagnosis and acute prescriptions linked to diagnosis. Other prescription items and follow up information was not improved through feedback. They concluded their research with identification of motivational factors that could have had an impact on data entry users.

Data quality feedback has already been implemented by the research group of Brouwer, Bindels & Weert (2006) in clinical data management systems and seemed “to have some positive effect” (Brouwer, Bindels, & Weert, 2006). However, with the improvement of technology and the availability of new (cheaper) devices, novel tools for data quality feedback are far more readily accessible. We consider the background of recent literature that address new approaches for data quality feedback with focus on different data entry domains including manual data entry and electronic data capture mechanisms.

2.3.1 Feedback Realisation

Main origin of errors in data collection is in the original source to registry or trial database transfers (Day, Fayers, & Harvey, 1998). The hospital patient record is considered as the most accurate source of clinical information and the manual double data entry of patient information into a trial registry is

considered as the gold standard of data accuracy (R. M. Califf et al., 1997; Paulsen et al., 2012). Data quality relies on the quality of the original (source) data and the subsequent transcription of this into the registry databases or clinical trial case report forms (CRF). Generally, the collection of patient information and trial-related data are performed in parallel (dual source information system – see Figure 2-6). Due to policy and security regulations, it is not always possible to collect research data directly from patient information systems (single source information systems – see Figure 2-7). There are a multitude of data collection and data sharing models and systems, each with a different purpose and design goal established to meet the demands of the associated stakeholders (Sinnott, Stell, & Jiang, 2011).

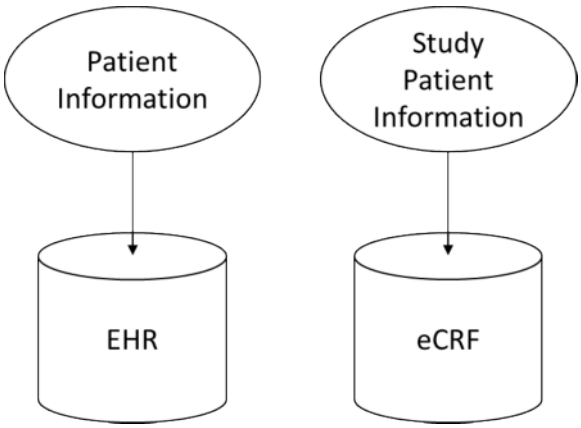


Figure 2-6 Dual Source Information System

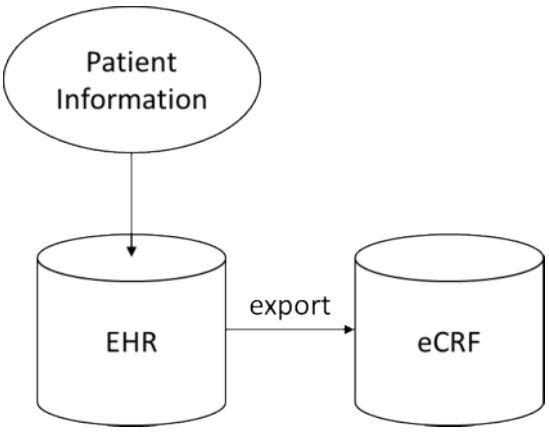


Figure 2-7 Single Source Information System

To tackle issues of data exports, EDC supports the process of information collection and reduces the error rates (Nahm et al., 2008). Research has shown that EDC transfers data with an equal accuracy compared to human data entry, but the time is significantly reduced hence the trial costs are reduced (Walther et al., 2011). It has also been shown that capturing data in electronic CRF (eCRFs) is only effective if it is used in the framework of multicentre trials (Le Jeannic et al., 2014).

Mobile applications (apps) and real-time notifications

In 2014, it was identified that the sales of mobile devices have been increasing each year and are now more widely used to access web-content than desktop PC/Macs (Fulgoni, 2015). During the 2016 Google I/O Keynote CEO Sundar Pichai informed the audience that 50% of all Google queries were now made by mobile phones and every 5th query is a voice query (Pichai, 2016). Even third world countries have an increased mobile device usage (Ibukun, 2014). The mobile phone is now a key part of (almost) everyone's life. Almost everyone now uses such devices for checking social networks, for news or email more than once a day (Fulgoni, 2015). One feature of mobile applications is the ongoing alerts, reminders and notifications that can pop-up on the home screen, reminding everyone of specific events.

The term that describes the use of mobile devices in public health setting is often called mobile health or mHealth. Systematic reviews concluded, that mHealth is beneficial, but less research has been conducted to show significant improvements of such applications on health generally (Aranda-Jan, Mohutsiwa-Dibe, & Loukanova, 2014; Catalani, Philbrick, Fraser, Mechael, & Israelski, 2013; Hamine, Gerth-Guyette, Faulx, Green, & Ginsburg, 2015).

mHealth can be used to improve local data processes as shown in a mobile and web-based approach like SORMAS (Surveillance and Outbreak Response Management System). Here a data management systems using mobile phone apps was implemented into local data processes to improve the reporting/surveillance of an Ebola outbreak for epidemiological purposes in 2014/2015. Patient

information from suspected Ebola patients was captured on paper-based forms, however SORMAS developed a mobile app that improved and simplified the data capture for local medical personal (Fähnrich et al., 2015). Furthermore, with regards to data quality feedback, another research group implemented data quality feedback into a food questionnaire-based trial. They distinguished between data input feedback and data analysis feedback. This data quality tool could be dynamically adjusted by the user, who could define their own data quality requirements with assigned weights (Weber, Presser, & Norrie, 2015).

Another group investigated the increase in data quality achieved through mobile apps as part of a randomised controlled trial (Joos et al., 2016). The improvement of health information can also be accomplished with the help of motivational short message service (SMS) or by the combination of motivational SMS and data quality feedback. The author discovered no differences between the two groups, but mentioned that the study was underpowered (Joos et al., 2016).

Importantly for data capture is the role of the patients themselves. It is now possible to develop mobile applications for patients that allow them to enter their own self-reported data. This offers new possibilities for data capture that extend beyond the kinds of data that is typically collected in a clinical healthcare setting, e.g. real time reporting of quality of life issues that impact patients with particular ailments, or indeed on their general life issues more generally, e.g. the exercise they take or their dietary intake (Sinnott, Han, Hu, Ma, & Yu, 2015).

An important aspect of data accuracy arises from the data source. The ideal situation is a single source system, where data errors can easily be tracked and reminders automatically sent to researchers and/or data entry personnel (Herzberg et al., 2011). In the case of dual source systems, a human will typically transfer the information into a disease registry or a clinical trial database. This can be undertaken manually or through automated data extraction systems. In the former scenario, which is by far the most prevalent existing across the healthcare landscape, data quality relies on the motivation and accuracy of the user. Therefore, mechanisms that encourage a user to enter high quality data must be investigated. mHealth applications whilst promising for public health, are not well researched with regards to data quality and challenges around the trustworthiness of the data.

2.4 Enhanced Web-based Solutions and the Move to Complete Virtual Research Environments

As explained, over the last years funding into R&D for new medical entities has increased, but the number of actual developed medical solutions has actually decreased. Web technologies have now evolved to provide more all-encompassing solutions that can increase the amount and quality of data that is collected.

2.4.1 Web-based Frameworks

Ideally, all tools and processes must be developed and used in accordance with GCP Guidelines and all applicable regulations. Quality assurance in clinical research is defined as:

> *"[…] all planned and systematic actions that are established to ensure that the trial is performed and the 'clinical' data are generated, documented (recorded), and reported in compliance with GCP and applicable regulations."* (International Conference on Harmonization, 1996).

To ensure these aims, a variety of Quality Control methods have been developed to assess and improve the quality and integrity of the collected data. One of these methods that assesses the quality of trial performances is (Data-) Monitoring (Campbell & Sweatman, 2002; International Conference on Harmonization, 1996).

Monitoring needs to be an on-going process in Quality Control that continuously aims to decrease inconsistencies. In general, monitoring is a continuous process that can be used to avoid specific errors (Arts et al., 2002). Historically, monitoring has been time-consuming, inefficient, problem-focused and cost-intensive (Baigent et al., 2008). To tackle this, the US Food and Drug Administration (FDA) published guidance documents for efficient (risk-based) approaches to clinical monitoring. This approach recommended a centralized data monitoring approach, that allows to discover critical data and data anomalies and then perform on-site monitoring, thereby allowing local site staff to improve data processes in the selected items (U.S. Department of Health and Human Services et al., 2013).

Arts, Keizer & Scheffer (2002) conducted a literature review about quality control methods for the improvement of data quality in the context of biomedical research. They described procedures that improve the quality of data at the *setup of the registry*, *during data collection* and *general actions for data analysis*. Unfortunately, data cleaning is a method currently established by the database community that has not been targeted to the specific needs of the translational biomedical community (Batini & Scannapieco, 2006; K. Chen et al., 2011). The authors conclude that quality assurance in clinical trials needs to be an ongoing process and not simply involve data cleaning at the end of a given trial. This ongoing systemic process needs to not only ensure that high data quality is collected, it also represents the need for GCP compliant data management and improved data integrity and subject protection (Bhatt, 2011).

Nonnenmacher, Weiland & Strausberg (2007) adapted this framework and developed an infrastructure for quality control in cohort trials. They created a recommendation for a quality score that could report on the quality of the collected information. In comparison to the established quality indicators (structure quality, process quality and outcome quality) from Donabedian (1988), they suggested 24 scores divided into three categories (plausibility, organisation and correctness) (Nonnemacher et al., 2014).

2.4.2 Heterogeneity of Data Management in Biomedical Research

One major challenge with translational research is data heterogeneity (E. A. Zerhouni, 2005; E. a Zerhouni, 2007). Translational research is supported by translational informatics that aims to “enable high-throughput research” (Payne et al., 2009; Payne, Johnson, Starren, Tilson, & Dowdy, 2005). There remain many issues in the diverse landscape of health IT systems with numerous standards and data formats. Of particular importance for the performance of clinical data management systems (CDMS) for clinical trials are GCP Guidelines and standards related to GCP (International Conference on Harmonization, 1996; Ohmann et al., 2011).

One of these documents is the Good Clinical Data Management Practise (GCDMP) Guideline. This industry standard proposes standards for ongoing data quality assurance of CDMS. Among other “best practise” requirements, it suggests calculating error rates, assessing the comparability of data between sites and conducting evaluations on the impact of data quality measurements and associated processes to improve data integrity (Society for Clinical Data Management, 2005).

Unfortunately, hitherto there have been no widely accepted standards for GCP-compliant data management of multicentre trials associated with the underpinning IT-Infrastructure. Results of a survey in a European research consortium showed a significant heterogeneity of CDMS used in clinical trials. The survey showed that in 66 centres, 20 different commercial and 7 open-source CDMS were in use. Furthermore, more than half of all centres were identified as not performing data management according to GCP, FDA or GAMP Guidelines (Kuchinke et al., 2010).

To tackle this, the ECRIN Working Group suggest a general CDMS framework with use of a structured communication techniques comprised of a board of data management experts and discussion groups

(Ohmann et al., 2011). This standard is divided into three parts: standards for IT, standards for data management and general standards (see Figure 2-8). Every group of standards has criteria that are classified as "essential", i.e. they represent the minimal recommended requirements for a clinical trial (Ohmann et al., 2013).

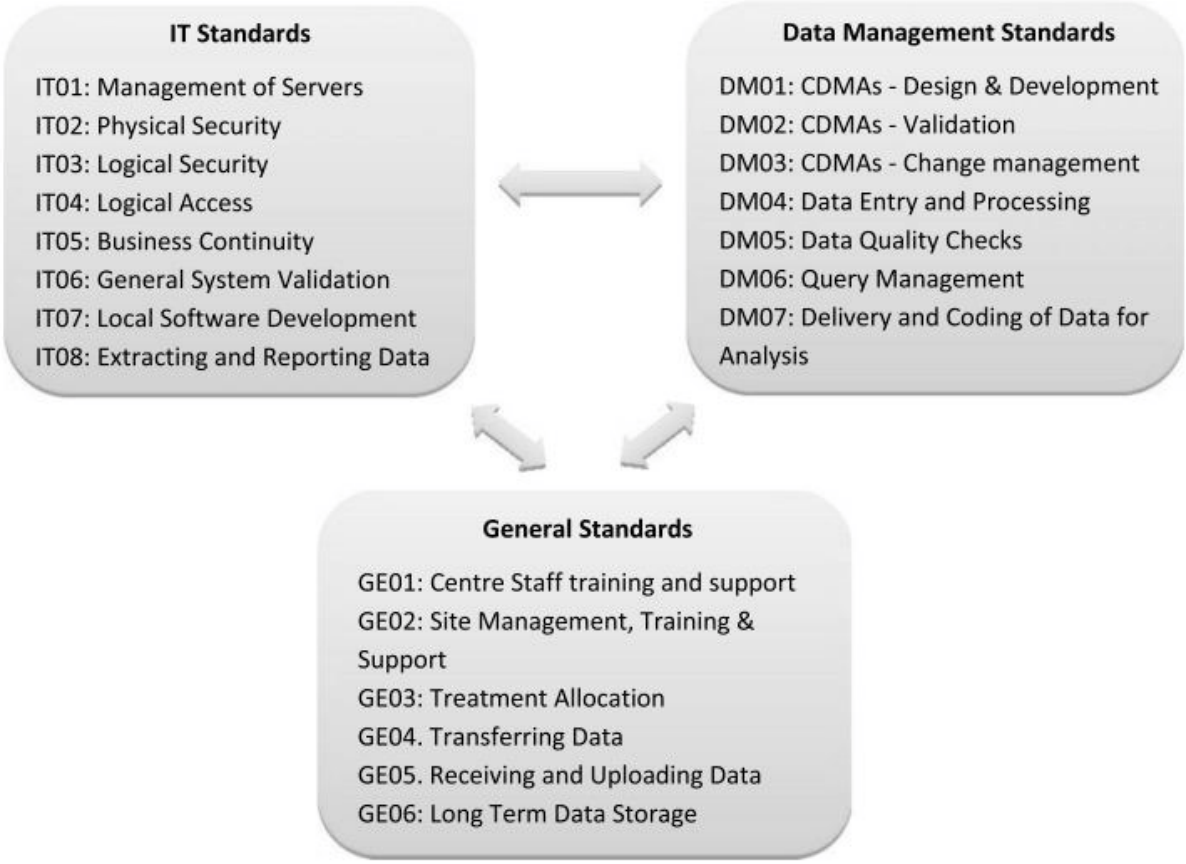


Figure 2-8 List of ECRIN standards (Ohmann et al., 2013)

In essence, one aim of translational informatics is to increase the information flow and knowledge transfer that are caused by, amongst other things, the heterogeneity of data management systems. It was stated that data management systems need ongoing data quality checks. Limited research has been published in this specific area. This suggests that data quality control could be improved through implementing features such as performing feedback in the data management system and encouraging users to improve their data quality and the processes that might impact on their, and hence the overall system, data quality.

2.5 Diversity of Funding Schemes

In the US, Australasia and most other nations the majority of biomedical funding involves single investigator initiatives: the NIH RO1 grant in the US is a prime example (Yin et al., 2015). In 2015, the NIH provided a total of $30 Billion USD for research grants. Most of the NIH budget goes to R01 grants, which fund single investigator initiated initiatives. There are now Clinical Translation Science Awards (CTSA) and National Cooperative Drug Discovery Groups (NCDDG) awards for group efforts, but these are a minor part of the overall NIH budget (<2%) and they are only awarded through NCATS (Collins, 2011). Some smaller NIH institutes are also now moving to partnership awards and providing minor budgets for more cooperative/collaborative-oriented research.

On the other hand, research elsewhere is much more collaboratively oriented. As two examples, Germany is supported by three large sources of funding: the German Research Foundation (DFG); the

Federal Ministry of Education and Research (BMBF), and the EU through the Horizon 2020 initiative. In France, the French National Research Agency (ANR) funds increasingly interdisciplinary and collaborative research projects. In contrast to the US, in the EU there is much more effort on cooperative multi-project and multicentre grants. In 2014, 21% (~$683 Million USD) of the DFG budget was used for Collaborative Research Centres (SFBs), but there are other multicentre funding grants including Clinical Research Groups (KFOs) (DFG, 2016a). Around 35% of this budget is used for Life Science research (DFG, 2016b). The BMBF similarly funds partnerships. The third major source of funding is the EU through the Horizon 2020 initiative. These grants are explicitly targeted to large-scale multinational research endeavours. It should be noted that EU funding has gone up dramatically since the mid 90's from less than 1 Billion Euro to estimated 80 Billion Euro for the package 2014-2020 (Grove, 2011). Despite this growth of funding, many challenges related to the volume, the veracity and the variability of data (often called big data) still need to be addressed by biomedical research networks if they are to achieve a higher overall impact.

To overcome these issues, interdisciplinary research and improved communication (feedback) is needed. Research activities, that provides education for researchers and infrastructures that allow for rapid and accurate knowledge sharing are essential (Wild et al., 2015). It is assumed that international teams and interdisciplinary research enhances the quality of research and ultimately the impact of the resultant conclusions, however this assumption needs to be considered in more detail.

2.5.1 Multicentre clinical collaborations

As mentioned previously, to achieve significant breakthroughs especially in rare diseases, multicentre trials that involve many centres and multidisciplinary teams are required (Gatta et al., 2010; E. A. Zerhouni, 2003). One of the primary benefits of multicentre trials are increased sample sizes that offer increased statistical power leading to improved research outcomes.

However, multicentre clinical trials also face a number of challenges: diverse performance of trials can be caused by ongoing changes of the study protocols or complex protocol designs and impacts on associated data collection (Brandt et al., 2006). Other problems include data propagation; fraud and erroneous randomization or in some cases incorrectly calibrated equipment, which can make the conduct and comparison of studies difficult (Venet et al., 2012). A literature review showed that Randomized Clinical Trials (RCTs) in single centre settings have larger treatment effects than intervention trials in multicentre settings (Dechartres et al., 2011).

As argued, multicentre research depends on tackling data management challenges. With the existence and adoption of data processing standards, the usefulness and quality of collected data can be increased to improve the findings and conclusions of collaborative research. This is especially important in the area of rare diseases such as adrenal tumour research. Rich research-oriented infrastructures – often called Virtual Research Environments (VRE) - allow exchange of patient information to support multiple trials. VREs can now be established across many targeted areas that augment basic databases and support targeted data collection and support the interconnectedness of data related to multiple trials and studies (Sinnott & Stell, 2011; Sinnott et al., 2011).

The technical software solutions that can now be achieved depend on the funding models that allow them to be created, adopted by multiple organisations and centres, and ultimately used to influence research and clinical outcomes. In this context, we consider how diverse funding schemes in Europe are allowing investigators and centres to work more collaboratively and in so doing, having a more positive impact on biomedical research productivity.

One way to measure collaboration and research success is through actual research publications that are generated as a result of data sharing. However, research success is not the only impact,

nevertheless it is often a major factor in how funding is delivered. Successful submitted research articles are often a key determinant in the allocation of funding that subsequently shapes the activities in the clinical, biomedical and indeed many other research domains.

2.5.2 Measuring Clinical Research Collaborations

International societies, research networks and/or individual scientists often use bibliometrics to measure research success, e.g. impact factors of papers that are successfully published. Traditionally, bibliometrics typically offers scores that characterise a journal by calculating the count of citations from other articles that refer to a paper in that journal. Bibliometrics are used to capture an expression of citation quantity not quality of the article/journal, however they are often used as a tool to show the influence of a journal in a particular field of research.

Commonly used and well-established bibliometrics are the impact factor (IF), the Eigenfactor or h-index. The IF is calculated through the average number of citations from articles in a given journal over a 2-year period. The IF shows the impact of a journal in a particular field of research. In contrast, the h-index measures the total number of articles and the associated citations. The h-index is provided by major research index providers including Scopus and Thomson Reuters. Scopus also provides the SCImago Journal Rank (SJR) that measures the prestige of a particular journal. Scopus has the advantage that new journals and journals not listed in more mainstream Thomson Reuters Journal Citation Reports. Different bibliometrics are used for different purposes. Exploring the impact of funding and software solutions on research success implies that an analysis of articles and journals must be performed using more than just one metric (Bollen, Van de Sompel, Hagberg, & Chute, 2009; Bornmann, Marx, Gasparyan, & Kitas, 2012; Falagas, Kouranos, Arencibia-Jorge, & Karageorgopoulos, 2008). It is noted that this is just one aspect of research success; novel treatments, drug therapies or patient management therapies are other clinical outcomes that are obvious alternative measures of research success.

To measure the impact of collaborative research networks, it is mandatory to explore whether the research networks funded by the EU have made an impact. Specifically, an exploration is needed to determine whether the establishment of research networks with support software infrastructures has galvanised the research community, and determine how this has changed on a global scale. To prove this hypothesis, we describe a representative case study focused on the adrenal tumour research community – specifically through the European Network for the Study of Adrenal Tumours (ENSAT[3]). It is argued that investigating the impact of ENSAT in adrenal research, the findings can be applied more generally to other clinical domains, or indeed to other domains and disciplines. ENSAT is an example of a large European network of scientists, clinicians, software engineers and quality controllers that takes advantage of European funding schemes to perform international research. If the research success of ENSAT is significant higher compared to other research initiatives of non-European funding schemes, conclusions and discussions about the general practicability of single centre funding must be initiated. The enhancement of technologies in clinical research assists large research networks to collaborate, exchange knowledge and promote young scientists in the field of research. Data management tools that provide audit and support data entry are general expected by authorities that provide funding for clinical research.

2.5.3 The European Network for the Study on Adrenal Tumours (ENSAT)

Luke et al. (2015) conducted an investigation into collaborative research of an academic institute in the US between 2007 and 2011. Here, the networks of scientists were analysed. It was shown that

[3] www.ensat.org

studies involving scientist with higher numbers of collaborations and more interdisciplinary team activities over time were identified as being more successful. The authors concluded that interdisciplinary collaboration is key to obtaining higher impact of research conclusions by the scientific community. By promoting collaboration between basic scientists and clinical researchers, collaborative research achieves a higher impact than individual researcher-driven or individual centre-driven science (Barker-Haliski, Friedman, White, & French, 2014; Wuchty, Jones, & Uzzi, 2007). To validate this hypothesis, it was suggested that network analyses should be used to evaluate diverse developments in collaborative research (Luke et al., 2015).

Research into collaborative studies and their outputs, has been explored in the field of the Science of Team Science (SciTS). It was identified that multidisciplinary teams are advantaged and able to find conclusions on complex diverse objectives that are often not possible through single investigator/single centre research efforts. The phenomenon of so called “team-based research” or “team science” is supported by international online resources that focus on exploring and improving knowledge sharing within and across research communities. One example of this is the team science toolkit from the National Cancer Institute (NCI) (Vogel et al., 2013).

Over the last 20 years there has been increased emphasis in Europe, compared to the US and other nations, in supporting collaborative, multicentre research that has spurred the formation of international multicentre biomedical research networks. This thesis claims that this has resulted in accelerated translational research productivity in Europe compared to the US and other nations. ENSAT is an example of a research network that is conducted by international researchers who want to make an impact onto adrenal research. Investigating the impact of ENSAT in the context of the adrenal research community will help to gain an insight into the impact of collaborative funding schemes on research success that can potentially be generalised to other research areas.

ENSAT was founded in 2002 with the aim to improve the prediction and management of specific types of adrenal tumour. The network was merged from three existing but largely independent, adrenal tumour research networks in France, Germany and Italy, with research teams from the UK. ENSAT was originally developed for clinical researchers across Europe but has since grown to include member centres in Japan, USA, Brazil, Canada and Australia. The process of the development of the ENSAT VRE in close cooperation with the ENSAT network and the empowerment that it supports has been a key to its global success. Thus clinicians are not obliged to send physical bio-specimens or offer further information on patients other than what is in the registry. This model of collaboration has overcome many of the concerns of the clinical community that are involved.

ENSAT investigates four primary types of adrenal tumours including: adrenocortical carcinomas (ACC), pheochromocytoma and paragangliomas (Pheo/PGL), non-aldosterone producing adrenocortical adenoma (NAPACA) and aldosterone-producing adenoma (APA) - all of which are relatively rare (e.g. for ACC this is 1.5: 1million) and have typically poor survival rates (Golden, Robinson, Saldanha, Anton, & Ladenson, 2009; Kebebew, Reiff, Duh, Clark, & McMillan, 2006). All clinical information entered into the ENSAT VRE is categorized into one of four related sets of tables in a database with four associated (international) working groups (see Figure 2-9).

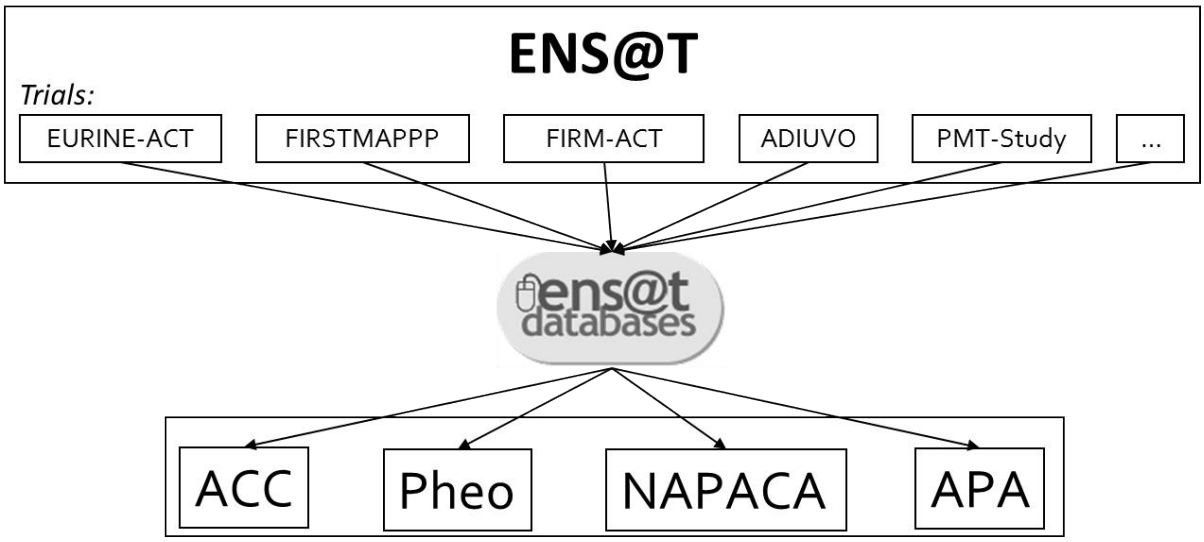


Figure 2-9 Patient Information from ENSAT-trials categorised into work groups

The information technology and data management solution of the ENSAT platform itself is offered through a security-oriented, web-based VRE. This VRE provides a variety of functions for clinical researchers (Stell & Sinnott, 2012). The VRE supports a portfolio of coordinated translational studies (basic science/genetics and clinical trials/diagnostic & treatment) of adrenal tumour patients that can be used to reveal new molecular mechanisms of the growth of these tumour types and provide insight into associated clinical research areas, e.g. their role in hypertension (Beuschlein, 2013).

A key aspect of the ENSAT VRE is the re-usability and classification of the captured clinical data. This is accomplished through centralized data monitoring for data completeness (Glöckner et al., 2015). An audit tool allows for measurement of the completeness of data and record eligibility in the registry. This can be used for research centres data completion rates, associated trial data completeness rates as well as the tumour type data completeness rates.

Based on the above, we consider the impact of collaborative funding with regards to the publication success of adrenal research between 1995 and 2015. It is assumed that only some of the highlighted funding goes to biomedical research and only some to endocrine efforts and an even smaller fraction to the adrenal tumour research domain. Nevertheless, the discovered patterns in the data analysis are compelling and representative, and we argue a key overall barometer to measure the success of clinical research collaborations. This in turn has dependencies on the quality of the data within the ENSAT registry.

2.5.4 Assessment of Research Success

To explore ENSAT as collaborative research network funded by European grant and its impact into adrenal research globally, published work related to adrenal research and health care from MEDLINE was considered. To cover the primary topic of adrenal research the search terms: "*primary aldosteronism*", "*Cushings Syndrome*" (including the various spellings), "*pheochromocytoma*", "*paraganglioma*", "*incidentaloma*", "*adrenal cortical cancer*", "*adrenocortical carcinoma*", "*adrenal cancer*" and for general articles "*adrenal tumour*" (including the various UK/US spellings) were used. The search was restricted on titles between the years 1995 to the end of 2015. All types of publications and languages were included. Reviews papers were not considered as research success. To exclude reviews, all articles with fewer than two authors were excluded from the final analysis.

In total, the search identified 9275 publications that were indexed in MEDLINE. The search was subsequently exported into an XML (Extensible Markup Language) file to use an XML reader library of Python to automate the process of data mining to create a CSV spreadsheet (for source code see Appendix: Python Code for Data Extraction). With this CSV (Comma-Separated Values) file details of the publications were categorized into article name, PubMed identifier (PMID), name of the first and last author along with their institutional affiliation, number of authors, International Standard Serial Number (ISSN) and year of publication. The nation of the author was assigned based on the author institute affiliation. After the exclusion of review papers, 7856 articles were identified for the final analysis.

Here publication success was characterized as the total count and average yearly increase of successful submitted papers in journals with high influence and reputation. The data analysis was based on three journal metrics: the SJR, the h-index and the IF. These metrics were chosen to evaluate the success from multiple aspects: the prestige of the journal (SJR), the productivity of the journal (h-index) and the average number of citations of the journal within a given 2-year period (IF).

To review the impact of multicentre research versus single centre funding on publication success, a comparison of publications and metric trends between North America (NA) and Europe (EU) in the years 1995 to 2015 was conducted. For NA data from the US and Canada was used in the analysis. Mexico and other North/Central American countries were investigated separately.

To confirm the thesis of different regions and related research funding schemes, nations/regions with the highest gross domestic spending (GDP) on R&D in the year 2013 were also separately analysed regarding their submitted publications and bibliometric scores. The identified nations/regions selected were: US, Japan, Italy, France, Germany, Korea, Israel, Nordic Countries (Island, Denmark, Norway, Sweden, Finland), China/Taiwan (OECD, 2016). To show the outcome of related research networks, the identified regions were compared to the publication success of ENSAT. Research articles where first or last author were affiliated with ENSAT were marked as an ENSAT article.

The observed years were clustered over the periods 1995-1999 (baseline data), 2000-2004 (start of ENSAT research network funding in Europe), 2005-2009 and 2010-2015 (following the establishment of the ENSAT research networks and the implementation of the ENSAT-VRE) and the shift of funding in the US vs. EU. The data analysis was categorised by:

(a) general publication overview;
(b) the yearly increase of the SJR per paper per continent/nation;
(c) the yearly increase of the h-index per paper per continent/nation, and
(d) the yearly increase of the IF per paper per continent/nation.

A linear regression analysis for the years 2000 and beyond was used to compare the period 1995-1999. The SJR and h-index data was obtained through the Scopus database[4]. The IF was obtained via the Journal Citation Reports[5]. Bibliometrics were available to 2014. To include publications for 2015 in the data analysis, the articles from 2015 were assigned with the same scores as 2014. This was similar to the SJR and h-index, which could not be obtained before 1999, hence articles published before 1999 were assigned with the same score as those of 1999. Also the IF was not available before 1997, hence articles from 1995 and 1996 were assigned the 1997 IF scores. For a concluding overview of the article analysis, the top 3 papers related to the bibliometric score were also presented.

[4] http://www.scimagojr.com/journalrank.php
[5] https://jcr.incites.thomsonreuters.com/

During the data collection, it was discovered that author affiliations were rarely included in the MEDLINE export, therefore data was reviewed and in the majority of cases author affiliation manually added. This research and the data processing elements could be extended to include an investigation including more than just first and last author affiliation. However, the last author affiliation is often not available via MEDLINE and in many cases just the first author has information given. Furthermore, due to license restrictions, it was not possible to access the full text of all articles and all named authors. This was an issue especially for articles available in other languages besides English where no author affiliation was given. During the manual review of the articles it was discovered that some non-English articles needed special review, e.g. Spanish articles were checked to determine whether their origin was Mexico or Spain, similarly for French articles (Morocco, Switzerland or France), Portuguese articles (Brazil or Portugal) and German articles (Germany, Austria or Switzerland). All non-English articles with no author affiliation were assigned with the nationality of the language used in the paper.

This analysis also included conversations/comments that were indexed in MEDLINE. It is questionable whether a comment has the same weight of influence in the research community as a peer-reviewed article in a highly cited journal.

To assign every article with the SJR, h-index and IF, the ISSN exported from MEDLINE was matched with the ISSN information exported from Scopus and Thomson Reuters Journal Citation Ranking. MEDLINE exports the eISSN (electronic ISSN) or ISSN-L (linking ISSN), however the bibliometric databases provide the p-ISSN (print ISSN) and this does not always match with the general ISSN of the journal. Therefore, where required, the ISSN was manually added to obtain an appropriate journal metric value for every article. Some papers were also discovered that had changed the name and ISSN within the observed period of years.

2.5.5 Discoveries

In the observed period, first authors affiliated with institutions from the USA published nearly 18% of all articles in the field of adrenal tumours. Notably countries with higher R&D per capita like Japan, Korea, Taiwan or Israel scored 2nd, 12th, 15th and 20th. In total 78 countries published 7856 papers within the 21 observed years (see Table 2-3).

Table 2-3 Nations (First Author affiliation) with the most published papers between 1995 and 2015

First Author Nation	Research articles
USA	1376
Japan	1134
Italy	734
France	477
Germany	448
China	390
Spain	350
India	305
United Kingdom	270
Netherlands	202
Rest of the World	2170
Total	**7856**

The EU published 3605 research articles between 1995 and 2015. Asian countries published 2425 and North America published 1526 articles. The Journal of Clinical Endocrinology & Metabolism was the

journal with by far the most successfully submitted and published articles from all analysed journal papers between 1995 and 2015 (see Table 2-4). The highest non-UK/USA journal was Hinyokika Kiyo (Japan) with 83 articles (place 9).

Table 2-4 Top 5 Journals of successfully submitted articles

Journal	Publications
The Journal of Clinical Endocrinology & Metabolism	400
Clinical Endocrinology (Oxf)	164
European Journal of Endocrinology	161
Journal of Endocrinological Investigation	105
Endocrine Journal	102

SJR

In the observed years, NA achieved the highest SJR values per paper per year. In contrast, the EU had the highest regression rate, which indicates that - if the trend continues - the EU will overtake NA with regards to SJR per paper per year in 2025. It is also notable that NA had a very high fluctuation of SJR values in the observed time period (R^2=0.2) (see Figure 2-10).

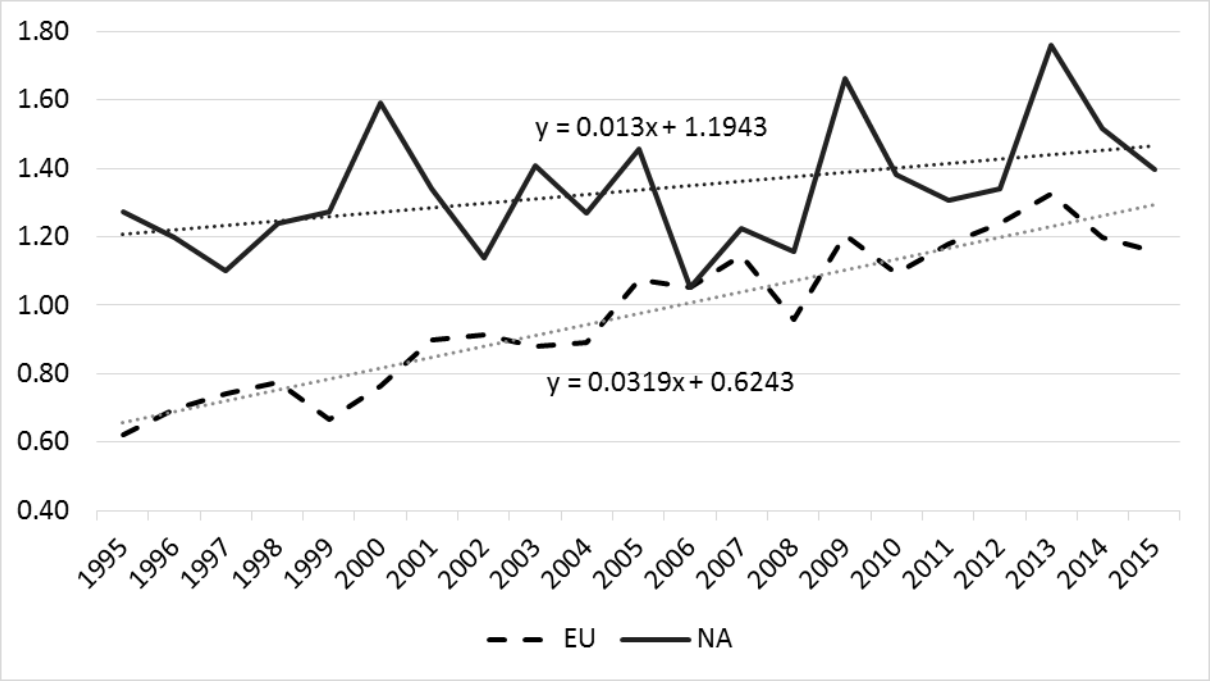


Figure 2-10 SJR per paper per year of EU and NA

The analysis of individual nations and regions with a higher GDP on R&D shows that France, Germany, the USA and Italy have the highest SJR scored per articles per year. The negative regression of Japan is explained by the fact that Japan started with a very high average SJR in 1995 (SJR=0.74). This contrasts to significantly lower values in the following years. ENSAT, as an example of international research cooperation, scored the highest average SJR values together with a steady regression rate (see Table 2-5).

Table 2-5 SJR per paper per year in selected nations/regions

	1995-1999	2000-2004	2005-2009	2010-2015
USA				
SJR per paper per year	1.19	1.37	1.36	1.44
Regression coefficient	-0.01	0.02	0.02	0.01
Japan				
SJR per paper per year	0.60	0.57	0.87	0.75
Regression coefficient	-0.06	-0.02	0.03	0.02
Italy				
SJR per paper per year	0.71	0.98	1.34	1.32
Regression coefficient	0.06	0.05	0.06	0.04
France				
SJR per paper per year	0.76	0.81	1.35	1.78
Regression coefficient	-0.08	-0.01	0.04	0.06
Germany				
SJR per paper per year	0.70	1.24	1.22	1.51
Regression coefficient	0.11	0.10	0.06	0.05
Korea				
SJR per paper per year	0.37	0.86	0.73	0.74
Regression coefficient	0.23	0.11	0.04	0.02
Israel				
SJR per paper per year	1.14	0.82	1.32	1.20
Regression coefficient	0.00	-0.03	0.02	0.01
Nordic Countries				
SJR per paper per year	0.72	0.85	0.87	1.17
Regression coefficient	-0.09	0.03	0.01	0.03
China/Taiwan				
SJR per paper per year	0.56	0.67	0.86	0.90
Regression coefficient	0.09	0.02	0.03	0.02
ENSAT				
SJR per paper per year	1.45	1.42	1.69	2.05
Regression coefficient	0.25	0.03	0.03	0.04

In 2014, the Journal *Nature Reviews Genetics* had one of the highest SJR sores in the biomedical domain. Journals with higher SJR are *Ca: A Cancer Journal for Clinicians* (SJR 2014: 37.4), *Annual Review of Immunology* (SJR 2014: 28.6) and *Nature Reviews Molecular Cell Biology* (SJR 2014: 24.3). Adrenal articles with the highest SJR between 1995 and 2015 are shown in Table 2-6.

Table 2-6 Top 3 articles with the highest SJR Score

Place	Article	FA Nation	Score
1.	Qin, Y., … Dahia, P. L. M. (2010). Germline mutations in TMEM127 confer susceptibility to pheochromocytoma. *Nature Genetics*	USA	22.4
2.	Comino-Méndez, I., … Cascón, A. (2011). Exome sequencing identifies MAX mutations as a cause of hereditary pheochromocytoma. *Nature Genetics*	Spain	20.5
3.	Goh, G., … Lifton, R. P. (2014). Recurrent activating mutation in PRKACA in cortisol-producing adrenal tumors. *Nature Genetics*	USA	19.6
	Assié, G., … Bertherat, J. (2014). Integrated genomic characterization of adrenocortical carcinoma. *Nature Genetics*	France	19.6

h-index

NA and EU showed a decrease of h-values over the observed period. The EU kept the h-values relatively constant over the observed years. The predicted intersection of the EU and NA regression will be the year 2019, i.e. when the EU will have a similar h-index. Similar to the SJR analysis, h-values show a drastic fall between 2006 and 2008 (see Figure 2-11).

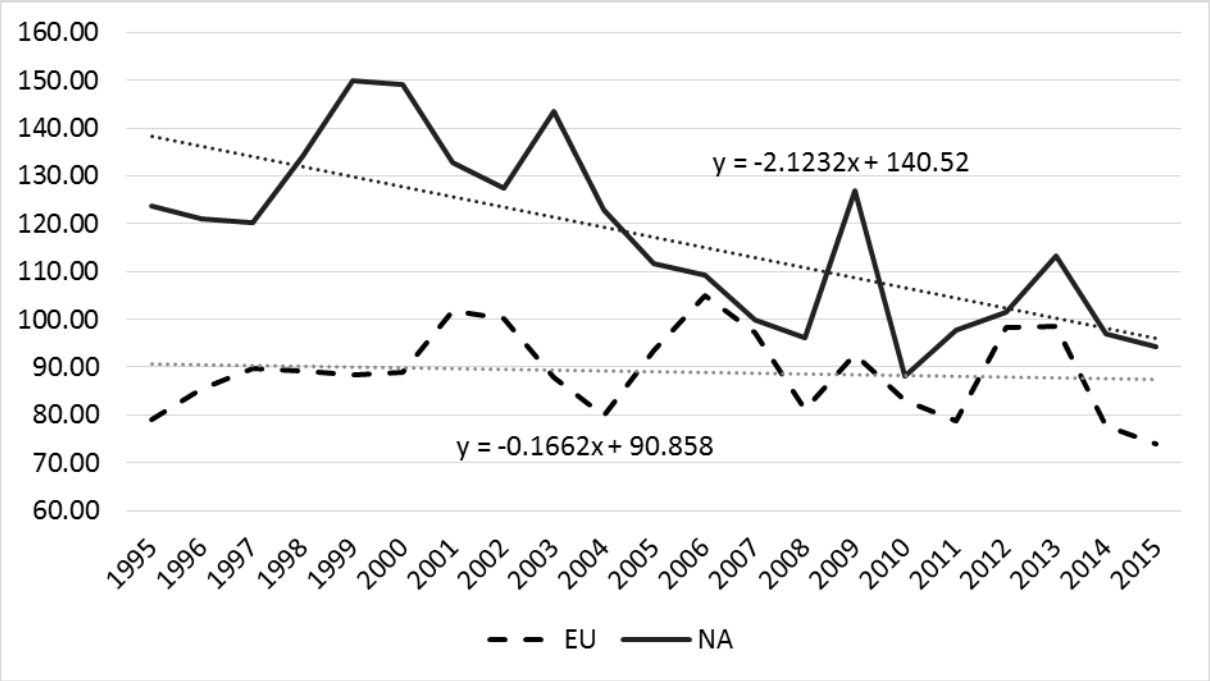


Figure 2-11 h-index per paper per year of EU and NA

Similar to the EU vs NA evaluation, the general h-index downward trend can also be observed in the national/regional analysis. France and Germany scored higher h-values than the USA, while France also achieved an upward trend regarding the h-index in the observed year periods (see Table 2-7).

Table 2-7 h-index per paper per year in selected nations/regions

	1995-1999	2000-2004	2005-2009	2010-2015
USA				
H value per paper per year	124.98	138.59	112.54	97.27
Regression coefficient	5.46	2.19	-1.00	-2.01
Japan				
H value per paper per year	76.09	67.33	62.62	57.41
Regression coefficient	-5.38	-2.33	-1.56	-1.09
Italy				
H value per paper per year	91.93	113.22	113.33	90.36
Regression coefficient	8.64	4.19	1.91	-0.30
France				
H value per paper per year	88.75	88.43	124.94	123.10
Regression coefficient	-9.24	-1.68	2.26	2.18
Germany				
H value per paper per year	89.41	121.00	113.46	113.42
Regression coefficient	15.39	6.19	2.84	1.49
Korea				
H value per paper per year	63.90	98.13	65.87	45.06
Regression coefficient	32.35	10.18	1.46	-1.43
Israel				
H value per paper per year	145.99	74.27	126.63	67.14
Regression coefficient	6.53	-9.58	-1.86	-3.73
Nordic Countries				
H value per paper per year	97.26	90.34	79.13	93.69
Regression coefficient	-12.27	-2.84	-2.58	-0.62
China/Taiwan				
H value per paper per year	62.28	78.19	71.75	65.59
Regression coefficient	11.51	2.60	1.00	0.01
ENSAT				
H value per paper per year	166.81	151.23	136.84	138.31
Regression coefficient	29.76	0.69	-1.93	-1.47

The journals with the highest h-index in 2014 were *Nature* (h-index 2014: 890), *Science* (h-index 2014: 851) and the *New England Journal of Medicine* (h-index 2014: 757). The best three papers with the highest h-index between 1995 and 2015 are shown in Table 2-8.

Table 2-8 Top 3 articles with the highest h-index Score

Place	Article	FA Nation	Score
1.	Baysal, B. E., … Devlin, B. (2000). Mutations in SDHD, a mitochondrial complex II gene, in hereditary paraganglioma. *Science*	USA	851
	Sato, Y., … Ogawa, S. (2014). Recurrent somatic mutations underlie corticotropin-independent Cushing's syndrome. *Science*	Japan	851
	Hao, H.-X., … Rutter, J. (2009). SDH5, a gene required for flavination of succinate dehydrogenase, is mutated in paraganglioma. *Science*	USA	851
	Cao, Y., … Ning, G. (2014). Activating hotspot L205R mutation in PRKACA and adrenal Cushing's syndrome. *Science*	China	851

IF

In the first 6 years of the 21st century, the average IF of all published NA papers dropped steadily. However, it was noted that the IF went up in the following years leading to their peak in 2009 with an average IF of 3.9 per paper per year. The EU significantly increased the IF from 2004 to 2008. Following this period, the EU held this trend and achieved a slight increase in 2011 onwards. If the trend continues, the EU will overtake NA in 2019 (see Figure 2-12).

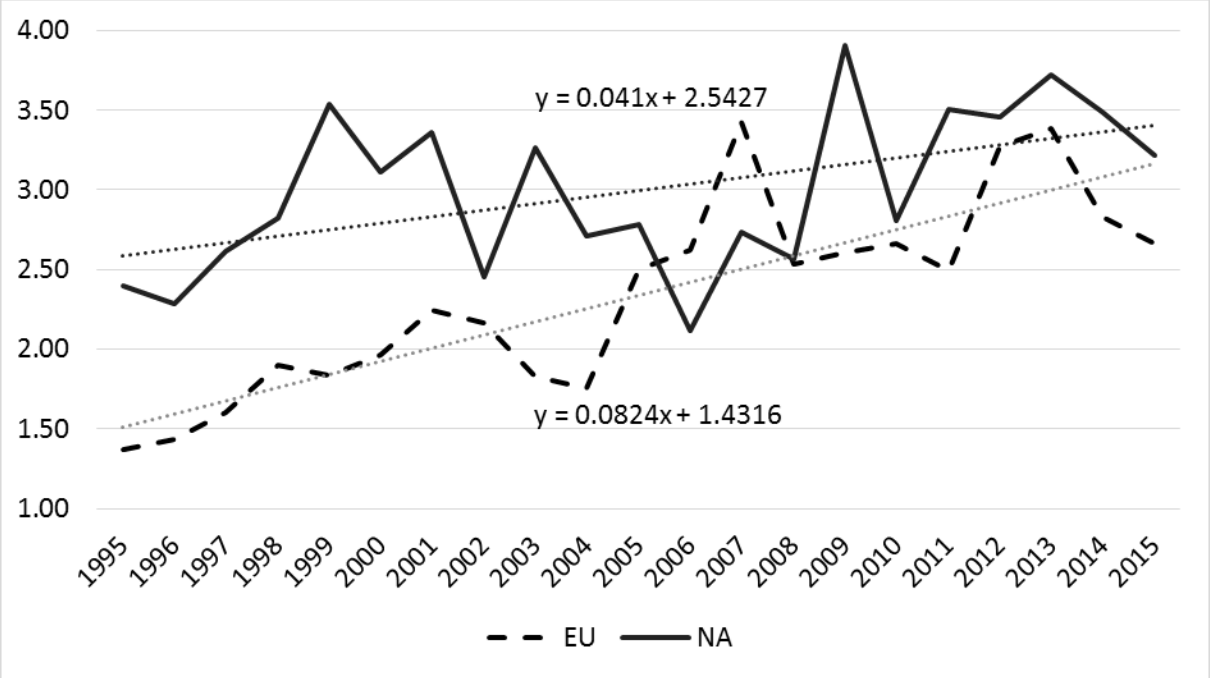


Figure 2-12 IF per paper per year of EU and NA

Like in the h-index evaluation, France and Germany have the highest IF scores, followed by the USA. The USA had a rapidly increasing trend between 1995 and 1999, but only a gentle increase from 2000 onwards. In 2005-2009, France, Italy, Germany and Israel scored significant higher IF scores than the USA, however Israel could not continue the high IF trend between 2010-2015 (see Table 2-9).

Table 2-9 IF per paper per year in selected nations/regions

	1995-1999	2000-2004	2005-2009	2010-2015
USA				
IF value per paper per year	2.55	3.00	2.94	3.30
Regression coefficient	0.20	0.08	0.05	0.04
Japan				
IF value per paper per year	1.19	1.26	1.32	1.68
Regression coefficient	-0.11	-0.01	0.00	0.03
Italy				
IF value per paper per year	1.69	2.40	3.60	3.02
Regression coefficient	0.25	0.13	0.16	0.09
France				
IF value per paper per year	1.84	1.82	3.74	4.69
Regression coefficient	-0.08	-0.02	0.16	0.19
Germany				
IF value per paper per year	1.72	3.11	3.47	4.07
Regression coefficient	0.39	0.24	0.17	0.15
Korea				
IF value per paper per year	0.79	1.75	1.65	1.14
Regression coefficient	0.67	0.25	0.11	0.02
Israel				
IF value per paper per year	2.08	1.67	4.73	1.56
Regression coefficient	0.62	0.06	0.25	0.01
Nordic Countries				
IF value per paper per year	1.72	2.27	2.42	2.96
Regression coefficient	-0.08	0.13	0.08	0.08
China/Taiwan				
IF value per paper per year	1.10	1.30	1.50	2.03
Regression coefficient	0.26	0.06	0.05	0.07
ENSAT				
IF value per paper per year	4.30	3.48	4.17	5.11
Regression coefficient	1.08	-0.01	0.02	0.08

The journals with the highest IF in 2014 were: *CA: A Cancer Journal for Clinicians* (IF 2014: 144.8), *New England Journal of Medicine* (IF 2014: 55.87) and *Chemical Reviews* (IF 2014: 46.57). Table 2-10 shows the highest IF scored in the period between 1995 and 2015.

Table 2-10 Top 3 articles with the highest IF Score

Place	Article	FA Nation	Score
1.	Beuschlein, F., … Allolio, B. (2014). Constitutive Activation of PKA Catalytic Subunit in Adrenal Cushing's Syndrome. *The New England Journal of Medicine*	Germany	55.9
	Teo, A. E. D., … Brown, M. J. (2015). Pregnancy, Primary Aldosteronism, and Adrenal CTNNB1 Mutations. *New England Journal of Medicine*	UK	55.9
3.	Assié, G., … Bertherat, J. (2013). ARMC5 mutations in macronodular adrenal hyperplasia with Cushing's syndrome. *The New England Journal of Medicine*	France	54.4
	Baudry, C., Paepegaey, A.-C., & Groussin, L. (2013). Reversal of Cushing's Syndrome by Vandetanib in Medullary Thyroid Carcinoma. *New England Journal of Medicine*	France	54.4

Conclusions

The results clearly show that the publication success of the EU and other single nations/regions like Korea, Nordic Countries, China/Taiwan, Israel will likely challenge the bibliometric scores of the USA in the not too distant future. The USA still has very high bibliometric values that have largely remained constant since 1995. On the other hand, countries like France, Germany and Italy have shown a drastic increase in their bibliometric scores with higher regression coefficients than the USA, especially over the last few years. Also these countries have all increased their baseline scores significantly since 2000. The findings support the argument that this is caused directly by the paradigm shift in more collaborative funding schemes provided by the EU and individual European national funds (such as the ANR, DFG and the BMBF).

The evaluation of publication success also indicates that research networks enhance the form of adrenal research and hence translational research worldwide. In our example, ENSAT will surpass the USA and NA with regards to publication success based on articles published in highly influential and heavily cited journals. Even though ENSAT started with higher scores in 1995-1999 compared to other nations, the regression analysis showed that these scores improved in comparison to the USA.

Furthermore, the results show that after the establishment of the ENSAT VRE in 2010, the publication success has improved greatly. This was caused by the ongoing and strategic improvement of the ENSAT patient registry – a key part of the ENSAT VRE and more importantly by its adoption by the wider adrenal tumour research community. With the focus on data quality improvements, further trials with new objectives are increasingly being identified thus consolidating the data quality and the grants and activities that are now possible through a larger critical mass of patient data and associated research opportunities. These subsequent grants build on and consolidate the collaborations and the community of researchers and clinicians that are engaging in ENSAT.

It is recognised that the used search terms don't cover the whole research field of adrenal tumours. However, with the focus on targeted terms associated with adrenal tumours the majority of published papers are covered and hence we would expect to see a similar trend using other related search terms.

The assessment has shown that research funding supporting collaborative networks to enhance biomedical research can and indeed has made a clear difference, at least in the adrenal tumour domain. However, that evaluation of research success requires consideration beyond the analysis of publications and their associated bibliometrics. To tackle the major impediments to translational research depends upon financial support for multidisciplinary and multi-organisational activities. Such

investments should ideally be aligned with established societies and networks. The ongoing tracking of research success and associated impact should also be increasingly considered. A key aspect of this that has been successful in ENSAT has been the realisation and adoption of the VRE and the associated data management tools that are used for continued monitoring and improvement of data quality, which has to be taken into account in every research evaluation (Hicks, Wouters, Waltman, de Rijcke, & Rafols, 2015). It was possible to determine which sites are truly collaborating and sharing data, and this information is made available across the ENSAT network. Such transparency is essential to facilitate research collaborations, especially those involving collaborations with often competing organisations.

Besides this, interviews with NIH researchers have revealed that the crucial elements for team success and effectiveness are focused on setting up the correct roles for each individual researcher in a team and showing the overall purpose of research. This in turn depends upon the objectives and transparency in the research. One of the most important elements in collaboration is trust. Trust in individuals that they will work with the same motivation and work load to achieve the same goal as those of collaborating team (Bennett & Gadlin, 2012). In our example, transparency and trust are supported through direct feedback in the ENSAT VRE that enables the research to continually improve.

This research has shown that there is a notable increase in influential research articles that have been published and indexed in MEDLINE by EU and ENSAT researchers between the years from 2000 to 2015. To continue to increase the publication and research success, it is essential to continue to support such international research networks. Furthermore, all networks should leverage web-based e-Infrastructures as typified through the ENSAT VRE to strengthen research findings and support new hypotheses. Collaborations, especially those that are facing challenges of big data depend upon multiple teams with mutually beneficial areas of expertise that can work on common research problems.

Since 2010 multicentre research funding that has been invested into ENSAT has galvanised international research in the area of adrenal tumours. Evaluation of the publication metrics substantiates this claim. Increasing numbers of patient records and information, combined with multiple features to manage the data continually improves the usefulness and quality of data. Rare diseases often struggle to achieve a critical mass of data and coordination of international research community. The ENSAT community and associated VRE is certainly the exception to the rule in this regard. The ENSAT network has also received follow on funding as part of the Horizon2020 ENSAT-HT[6] that will continue to build upon and support the ENSAT research community.

New dimensions to the work of ENSAT are the involvement and support of patients and patient support groups. To this end, the ENSAT VRE is growing with a range of new mobile applications that allows patients to enter their own data. This will offer new dimensions that will help better understand the daily challenges and issues of patients with adrenal tumours and their treatment and management.

Given all of this, one would assume that ENSAT has excellent high quality data. As we shall see, the capture and improvement of data quality continues to be an ongoing process. We show in later chapters how technology can be used to help support this process.

2.6 Chapter Conclusions

This chapter has systematically explored the many challenges associated with the capture of clinical data in disease registries to support clinical trials and studies, and thereby improve healthcare as a whole. We focused on the many dimensions of data quality that can impact on the overall usefulness

[6] http://www.ensat-ht.eu

of data entered into such registries. A key part of this is implicitly connected with the motivational aspects of the protagonists that are involved. This in turn impacts upon the funding models that are associated with the clinical collaborations, e.g. if researchers are paid to enter data then one assumes that the data will be of better quality.

We have systematically explored the research impact of the ENSAT registry to support clinical research into adrenal tumours and the international network that has been funded to establish and use the registry. We have demonstrated how such collaboration drives research output and how this can subsequently inform and improve clinical trials through the statistical power that increased amounts of data can imbibe. Data quality remains a major challenge however, especially when multiple centres are coordinating their efforts through web-based international collaborations.

The following chapter explores the extent that direct (face-face) feedback to the data entry personnel can improve the quality of data entered into disease registries such as ENSAT. If we can improve the quality of data in such a manner, then we hope to demonstrate that web-based (automated) feedback can also improve the quality of data in disease registries. This is the focus of subsequent chapters.

3 CASE STUDY OF DIRECT DATA QUALITY FEEDBACK

The previous chapter provided insights into quality assurance regulations in clinical research, data quality measurements and data quality feedback. Key to the investigation of the research hypothesis is that feedback can improve the quality of data in clinical data registries. Before exploring web-based approaches for automated feedback, a precursor investigation is whether feedback works at all. This chapter describes the empirical basis by which we demonstrate that we can measure data quality and understand why low data quality can occur through direct, face-to-face feedback with data personnel involved in entering data into clinical research registries.

For this exploration, a case study was performed within the ENSAT-CANCER consortium. Prospective evaluation of data quality was performed for two multicentre diagnostic trials EURINE-ACT (Evaluation of Urine Steroid Metabolomics for the Differential Diagnosis of Adrenal Cortical Tumours) and PMT-Study (Prospective Monoamine-producing Tumour Study). A diagnostic trial has the aim to improve current diagnostic procedures (also called gold standards tests) that can include finding new biomarkers or procedures or evaluating the sensitivity and specificity of particular approaches. In the case of diagnostic trials in cancer research, this means that newly developed tests will have the best possible true positive rate and best possible false negative rate with regards to the evaluation of whether the patient has an adrenal tumour or not.

The following chapter introduces the background in relation to adrenal tumours and the empirical work that was performed for the exploration of data quality assessment within ENSAT-CANCER. These assessments involved an initial centralised data monitoring (data quality evaluation of EURINE-ACT), a survey to explore commonly used data processes in the participating trial centres and a report about findings from six face-to-face on-site monitoring visits.

3.1 European Adrenal Research

3.1.1 Adrenal Tumours

It is assumed that three in 100 of all 40-year-old humans carry an adrenocortical tumour. The majority of all adrenal tumours are discovered incidentally during the diagnosis for another condition. For this reason, the tumour is often called an ‘incidentaloma’. The prevalence rate of incidentaloma results in a categorization of benign tumours as a common disease (Arlt, 2009; Kapoor, Morris, & Rebello, 2011).

Due to hormonal inactivity and their typically slow size growth, the majority of incidentaloma remain undetected during an individual’s lifetime. However, in rare situations, this tumour can be hormonally active and the tumour tissue can secrete adrenal hormones. In around 1-2 per 1 million healthy humans, such tumours can have a malignant character, causing back pain because of the rapid size growth, with infiltration of adjacent tissue and potentially the spread of cancer cells in other organs (metastasis) (Fassnacht & Allolio, 2009). The treatment often depends on the location, the patient themselves, e.g. their demographic details, and the tumour behaviour including for example whether the tumour is secreting hormones or not.

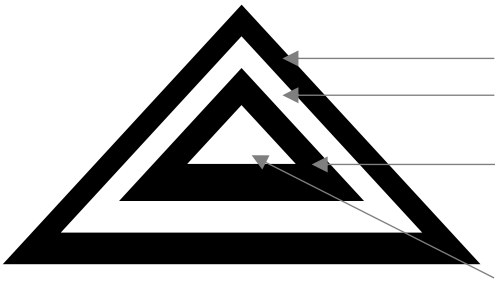
To understand the rareness of adrenocortical cancer, it is useful to consider the prevalence rates. Considering the population of Melbourne (4 million people), around 40,000 people could have an incidentaloma, while just 4 people in Melbourne could have an adrenocortical cancer. Other adrenal tumours like malignant pheochromocytoma have even lower prevalence rates.

This low sample size requires not only a multicentre approach for collaboration, but also an international effort to recruit sufficient numbers of patients. It is often the case that statistically significant evidence of new therapies or diagnostic tools can only be shown through large sample sizes (Gatta et al., 2010). The ‘Levels of Evidence’ provided by the Centre of Evidence Based Medicine

(CEDM) located at the University of Oxford consider that only the results and conclusions of randomised controlled trials (RCTs) or systematic reviews have enough power to be transferred into clinical care and influence clinical guidelines (Phillips et al., 2009). However for RCTs, large populations are required and the correct trial performance is mandatory. The patient sample size (cohort) and study more generally can also have ethical issues regarding the performance of RCTs, which challenge clinical care and patient management. One example is the FIRST-MAPP trial[7], the first RCT for malignant pheochromocytoma. Here the success rate of the cancer drug Sunitinib, an oral treatment for adrenal cancer was tested. Through the trial, the randomisation step required that some patients with malignant pheochromocytoma would not receive appropriate treatment for the tumour.

Whilst adrenal tumours can be relatively common diseases, some forms, especially malignant tumours, are very rare. Given this, clinical trials have to be performed on an international base and meet a range of requirements to provide conclusions that can be rapidly transferred into clinical care. In such contexts and given the previous discussions on data quality, infrastructures that support data management have a very high impact on the trial performance. Every collected data item needs to meet appropriately high standards. One standard is defined by the GCP and relates to the collection of high quality data. In multicentre trials, external auditors typically need to understand how rare tumours can be diagnosed. This can lead to the definition of compulsory data items that have to be collected during particular trials. To show how high data quality can support adrenal research and how data inconsistencies can be improved, the general pathophysiology of adrenal tumours needs to be understood. We summarise the most pertinent aspects here. A richer discussion on this is available at Arlt (2011).

The adrenal is an endocrine gland on each pole of the kidneys. The gland consists of two hormone-secreting regions: the adrenal cortex and adrenal medulla. The adrenal cortex is differentiated into zona glomerulosa, zona fasciculata and zona reticularis. Each zone produces specific hormones that include glucocorticoids, mineralocorticoids, androgens and catecholamines.



Zone	**Hormone Example**
Adrenal Cortex:	
Zona glomerulosa	Mineralcorticoids (Aldosterone)
Zona fasciculata	Glucocorticoids (Cortisol, Corticosterone, Cortisone)
Zona reticularis	Androgens (Dihydrotestosterone (DHT))
Adrenal Medulla	Catecholamines (Adrenalin, Noradrenalin)

The secretion of the hormone aldosterone influences blood pressure. An endocrine tumour in the zona glomerulosa increases the aldosterone production, which causes primary hyperaldosteronism (or Conn Syndrome) and consequently causes high blood pressure (Beuschlein, 2013). High levels of cortisol can be caused by a hormone-secreting tumour in the zona fasciculata. Not all causes of Cushing's syndrome (or hypercortisolism) are adrenal tumours, however there are specific diagnostics that can lead to the indication that higher cortisol levels are caused by an adrenal tumour. Typical symptoms are central adiposity, proximal myopathy, striae or hirsutism in women (Arlt, 2011). Hyperandrogenism can be caused by a tumour in the zona reticularis, which can lead for example to

[7] https://clinicaltrials.gov/ct2/show/NCT01371201

polycystic ovarian syndrome (PCOS) in woman. The symptoms of PCOS include menstrual disorders, infertility or high levels of male hormones that can lead to acne and hirsutism (Teede, Deeks, & Moran, 2010).

The adrenal medulla secretes catecholamines including adrenalin, noradrenaline and dopamine. Hormone-secreting tumours in the adrenal medulla are called pheochromocytoma. To diagnose a pheochromocytoma the gold standard test is the collection of 24-hour urine, where the levels of metanephrine and normetanephrine (metabolites of adrenaline and noradrenaline) are measured. When these levels are higher than the defined reference intervals, a pheochromocytoma is suspected. With a cortisol test, which should reduce levels of catecholamines in the blood, the tumour in the adrenal can be confirmed or excluded (Därr et al., 2014; Eisenhofer, 2004).

In all cases detailed information on the patients themselves (including their phenotypic information) needs to be captured to contextualise the biomedical information related to the adrenal glands and subsequent tests that can be applied.

3.1.2 ENSAT

In 2002, the European Network for the Study of Adrenal Tumours (ENSAT) was founded with the aim to improve the prediction and management of specific types of adrenal tumours. In particular, the community focuses on the tumour types: adrenocortical carcinomas (ACC), pheochromocytoma and paragangliomas (Pheo/PGL), non-aldosterone producing adrenocortical adenoma (NAPACA) and aldosterone-producing adenoma (APA) - all of which are relatively rare, e.g. for ACC the annual prevalence is 1.5 per 1 million population. Malignant adrenal tumours have typically poor survival rates (Arlt et al., 2011; Eisenhofer, 2004; Golden et al., 2009). To improve diagnostics and treatment of malignant adrenal tumours, the ENSAT-CANCER project[8] was funded as part of the EU Framework Program 7 (FP7) initiative in 2011. The web-based data management systems were delivered through a security-oriented, web-based virtual research environment (VRE). This VRE provides a variety of functions for clinical research (Stell & Sinnott, 2012; Stell, Sinnott, & Jiang, 2010). The VRE supports a portfolio of coordinated translational studies (basic science/genetics and clinical trials/diagnostic & treatment) of adrenal tumour patients that is used to reveal new molecular mechanisms related to the growth of these tumour types and provide insight into associated clinical areas, e.g. their role in hypertension (Beuschlein, 2013).

The ENSAT-CANCER VRE was built using a standard n-tier web application setup with a MySQL back-end database, business logic programmed in Java rendered to a JSP front-end. The front-end web application was hosted in a Tomcat container – with various libraries used to provide additional feature support (Apache POI, iTextPDF, Guava). Parameter rendering was achieved using a separate database listing the available parameters for the central forms and the subsidiary “one-to-many” forms implemented for treatments, biomaterials and follow-up. The VRE included a variety of features that enabled collection and validation of high quality data:

- consolidated databases reflecting the international community needs and consensus on adrenal tumour data;
- consolidation of international security policies on access and use of data that meets international ethical considerations;
- CRUD capabilities for search and export queries;
- interfaces for specific clinical trials (e.g. EURINE-ACT);
- freezer and aliquot management;

[8] www.ensat-cancer.eu

- labelling and barcode support with manifest sample tracking allowing biomaterial shipments and tracking between centres;
- association with other related studies including cross-referencing and candidacy for other studies as a collaborative feature including re-use of registry data;
- resource pooling and network collaboration, and
- statistical analysis interfaces with re-usability in mind.

By October 2016 the ENSAT VRE had collected biomedical information from 91 research centres across Europe and internationally on more than 10,000 patients. This included clinical annotations and stored specimens in biobanks (see Figure 3-1). At this time, the ENSAT community ran a multitude of clinical trials from phase I to phase IV studies across the diagnostic and treatment continuum. EURINE-ACT was one example of a diagnostic trial, which aimed at the development of tools for accurate differentiation between malignant and benign tumours to improve the early detection of recurrence of ACC and to better understand the biological processes that take place that lead to hormonal excess in adrenal tumours. To receive a prediction of the likelihood of an adrenal tumour, biochemical profiles of all relevant steroids were measured from 24-hour urine. The differentiation from malignant and benign tumours was carried out through learning vector quantization (LVQ) approaches (Arlt et al., 2011). The VRE also provided access to Cloud-based resources and high performance computing facilities (Sinnott & Stell, 2011; Stell & Sinnott, 2012).

	ACC	Pheo	NAPACA	APA	Total
Records	2541	2758	2940	1809	**10065**
Biosamples	5134	1998	7460	5759	**20351**
Clinical Annotations	29024	14898	8392	5652	**57966**
Annotations Per Patient (Mean)	11.42	5.40	2.85	3.12	**5.75**
Biosamples Per Patient (Mean)	2.02	0.72	2.53	3.18	**2.02**
Active Centers	54	60	35	23	**91**

Figure 3-1 ENSAT-CANCER Registry Summary (October 2016)

Given the scale and scope of ENSAT, data quality assurance methods are an essential component of the registry.

An additional data quality module was developed that correlated the user online behaviour and patient record quality in the context of specific studies including EURINE-ACT. The EURINE-ACT study had specific standard operating procedures on the registry data that needed to be monitored by the presence of certain forms, e.g. availability of biomaterials. By combining information from the registry data and the meta-information drawn from the log analyser, a survey of the completeness and accuracy of information within the study was undertaken and a “Data Quality Score” (DQS) assigned that offered an important metric when interpreting the results of the study. These metrics used captured data that represented the needs of the end-user scientists, however it was recognised that a balance between usefulness of data and requirement for excessive data input was needed. Thus not all data was needed for particular trial objectives. The DQS accuracy utilized the comparability, timeliness and validity of the entered values.

For the assessment of data completeness within the ENSAT-Registry, a graphical interface was developed to show researchers the overall completeness of the data that was available in the registry.

To support this, all records were assessed separately and results typically displayed as bar charts. Figure 3-2 shows an example graph for pheochromocytoma records from a particular centre.

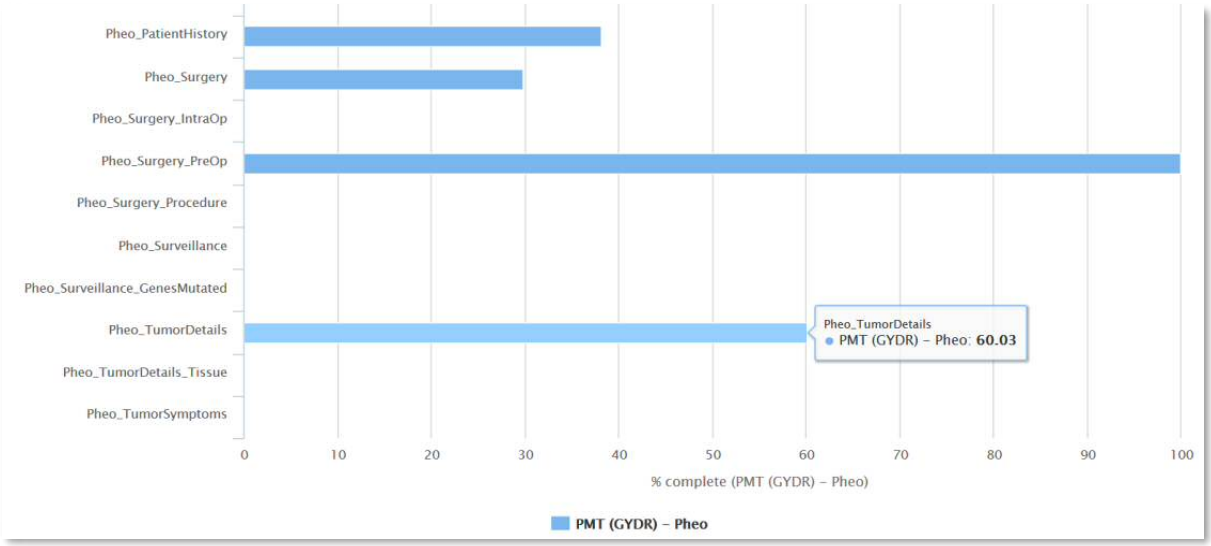


Figure 3-2 Data completeness bar charts in the ENSAT registry

3.1.3 PMT Study

Headaches, dizziness and sudden sweating that occur without physical exercises can be signs of high blood pressure. The reason for high blood pressure (HBP) is rarely investigated by general practitioners, due to the fact that the treatment of HBP is less expensive than the diagnosis. In 0.2-0.4% of all cases, the reason for such symptoms is a pheochromocytoma - a metanephrine producing tumour. Metanephrines are metabolites of the catecholamines adrenalin and noradrenaline. The PMT-Study (Prospective Monoamine producing Tumour) is a diagnostic trial that investigates such tumours to find biomarkers that can improve the diagnosis for pheochromocytoma (Reisch, Walz, Erlic, & Neumann, 2009). The primary objective of PMT is to identify new and improved biomarkers for the therapy and diagnosis of pheochromocytoma and paraganglioma (PPGL). Another goal is to stratify different groups of PPGL with the help of biochemical profiles. In the majority of cases overproduction of hormones causes secondary hypertension (Darr et al., 2012).

The PMT Study is a full phase 1 to phase 4 trial performed within the ENSAT consortium. Data collected in the study is captured in the PMT-eCRFs that are linked directly with the ENSAT-CANCER registry. Clinical information collected in PMT can be transferred into the ENSAT-Registry (and vice versa). In addition to the collection of demographic and disease related information, all patients undergo a standardized blood withdrawal and 24h urine samples are also collected. The PMT registry includes around 500-1100 data items per patient (depending on whether they enter phase 2-4). Many of these data items, e.g. demographic data, originate and are transferred directly from the ENSAT-CANCER registry. The PMT Study enrols all patients with a current finding or history of pheochromocytoma, but also patients with secondary hypertension (hypertension not caused by an identifiable underlying causes). Of specific interest to the PMT Study are those patients where the cause of their hypertension is through possible pheochromocytoma. A new Horizon 2020 project ENSAT-HT is focusing specifically on this topic.

A total of 2400 subject were planned for enrolment into the PMT study protocol. The trial design was split into four phases. All patients that meet the inclusion criteria enter the trial in phase 1. Here, a diagnostic screening for increased plasma and urine metanephrines was performed. If the test was negative, a hormone-secreting tumour in the adrenal medulla was excluded and the patient would be

called again after 2 years to verify the test results (Phase 4a). In case the metanephrines were strongly increased, a pheochromocytoma was likely and imaging procedures to detect the tumour were conducted (Phase 3). If the extraction of the tumour tissue was possible (benign pheochromocytoma) the patient would undergo surgery and regular follow-ups would be scheduled (Phase 4b). However, there is also the chance that the results of Phase 1 are only slightly increased, hence a differential diagnosis in form of a clonidine suppression test needs to be performed to confirm the suspicion of a pheochromocytoma (Phase 2). In case of a positive clonidine suppression test is the patient assigned to phase 3, otherwise a pheochromocytoma is excluded and the patients is assigned to phase 4a. Figure 3-3 shows the flow chart associated with the PMT Study including the statistical power analysis of subjects needed to meet the criteria for all phases and hence to realize the study objectives.

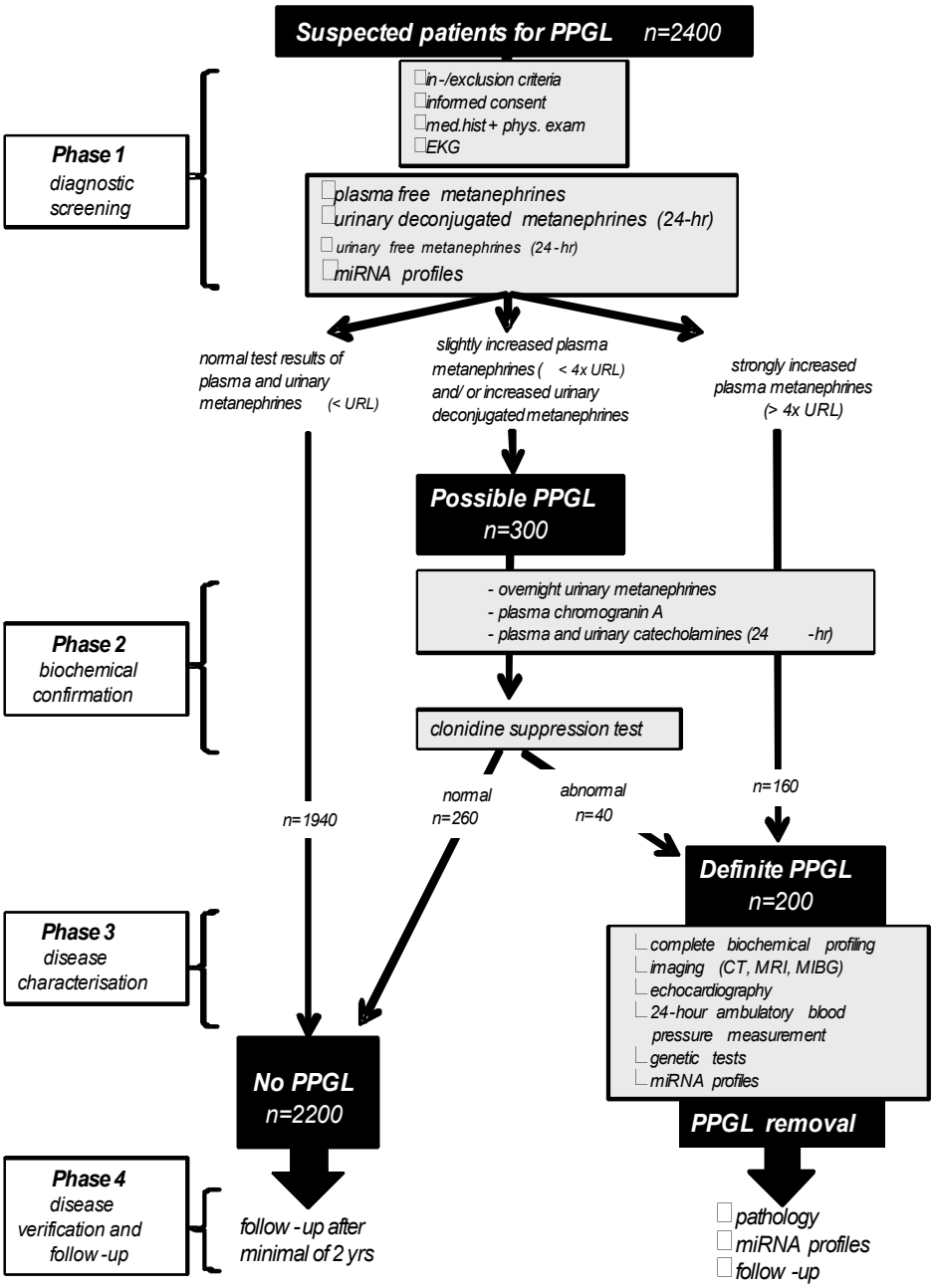


Figure 3-3 Flow chart PMT-Study

PMT Centralised Monitoring

Centralised monitoring is essential for quantification of data quality and the associated empirical work. Centralised monitoring has been a key focus of the PMT Study. The majority of all quality indicators typically need to be adjusted when evaluating data quality in clinical (cancer) registries. The assessment considered in the PMT Study included *comparability* of metanephrine values and the *validity*, *timeliness* and *completeness* of these data sets within the ENSAT research registry.

PMT Comparability

Crucial data items for the PMT Study include the results of blood tests that prove the existence of metabolites from metanephrines, normetanephrines and methoxytyramine (MTY), a metabolite of dopamine. To determine the normal range of such blood results, blood was drawn from a healthy cohort (n=300). Because of the importance of the correctness of these values in relation to the diagnosis of PPGL, a comparison was performed between the blood test results from healthy patients and those with suspected PPGL. A study within the healthy cohort discovered that some centres had several outliers for MTY (see Figure 3-4).

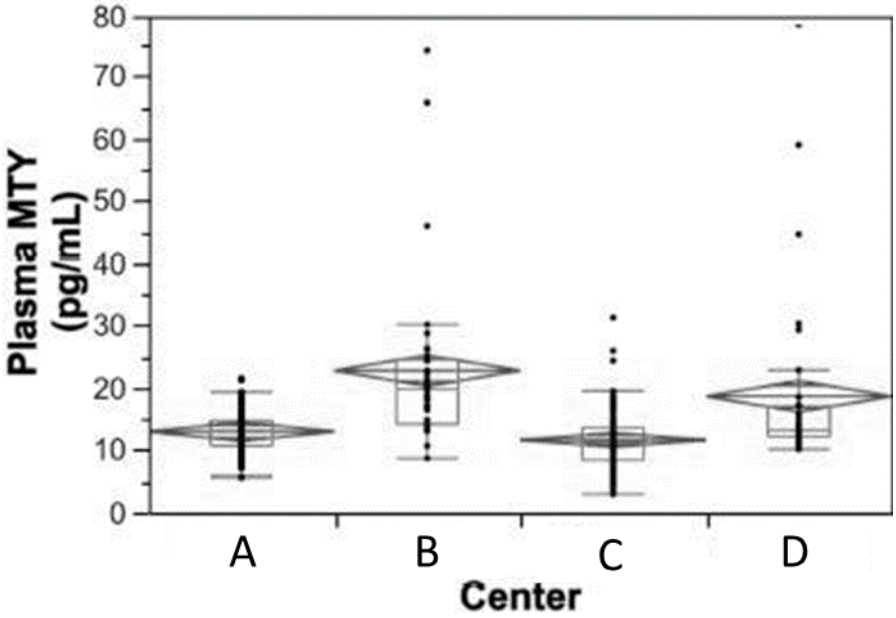


Figure 3-4 Outliers in a healthy cohort for blood tests in the PMT Study

Figure 3-4 was reported to the local investigators who subsequently refined their processes in how they performed blood sampling in accordance to the SOPs.

As shown, Centre A is hosting the principal investigator of the PMT Study and hence provides 'clean data'. All other centres show a larger range of values with three to four statistical outliers. These inconsistent results were reported to the centres and it was discovered that Centre B, Centre C and Centre D were not following the same Standard Operation Procedures (SOPs) for blood tests and subsequent processing of samples for presence of metanephrines (see Table 3-1).

Table 3-1 Summary SOPs for Blood Withdrawal (BW)

SOP Blood withdrawal	
…	the patient is 30min before and during the BW in a supine position;
…	the patient has fasted 12 hours before the BW;
…	the patient has not taken paracetamol within the last 4 days;

…	the patient has not taken tricyclic antidepressants within the last 2 weeks;
…	the patient was not physically active before the BW, and
…	the patient has no renal insufficiency.

The trial-related SOPs for blood withdrawal clearly states that the patient has to rest for 30min in the hospital before blood is extracted. Furthermore, the patient has to fast for at least 12 hours before the sampling and the blood has to be withdrawn in a supine position (Därr et al., 2014). It was reported that some patients arrived with a bicycle and the blood withdrawal was performed immediately after their arrival in the research centre. During an onsite monitoring visit, it was discovered that every centre had different hospital related standard procedures for their routine blood sampling collection processes (e.g. not in supine position or not after 30min rest).

PMT Validity

Data validity consists of three major quality indicators: *completeness, accuracy* and *plausibility*. While completeness provides information as to whether the data item exists in the data set, data accuracy is often used to apply special ranges or values that are expected, e.g. has the patient consented for the study or do they meet the inclusion criteria. Plausibility assesses whether data items are logical, e.g. 'female' patients are unlikely to have "prostate cancer' in their records. For the PMT Study, just one plausibility check was performed. Specifically, if the patient was enrolled into the study protocol with a history of PPGL, i.e. they had known genetic mutations or therapy resistant (secondary) hypertension, then the associated information in the eCRFs must include these details.

In the PMT Study, centralised monitoring of all Phase I patients require that 45 data items were defined as compulsory. As shown in Table 3-2, centres with separate research units (Centre A and Centre C) achieved lower error rates. It was further discovered that Centre B and Centre A created the majority of their PMT records in the ENSAT registry (Centre B 51%; Centre C 55%). The transfer of such records into the PMT Study eCRFs resulted in data loss. Furthermore, it was identified that some data items (identification and genetic information) were discovered in the ENSAT registry, but not in the PMT Study eCRFs.

Table 3-2 Results PMT data validity

	Centre		
	Centre A	**Centre B**	**Centre C**
Records	85	59	91
Completeness errors	8	31	15
Validity errors	5	13	3
Plausibility errors	-	2	1
Total error percentage	**0.34**	**1.73**	**0.46**

PMT Timeliness

The evaluation of timeliness for the PMT Study focused on the period between data entry and date of consent of the patient. In clinical research, the patient needs to be fully informed of all aspects of their participation in a trial and that this is voluntary in accordance to GCP Guidelines and the Declaration

of Helsinki. Furthermore, the patient should be able to ask questions and the possibility to withdraw from a study at any time without any disadvantage in the clinical care that they receive. The first visit of the patient, the initial date of data collection and when consent was obtained are important time-points to be captured. The PMT registry has two different data items that allow for the calculation of the period between data collection and actual data entry. One assumption that was explored was that the longer the time between the first patient-visit and data entry, the more inconsistencies would appear in the eCRFs.

The period of data collection (date of consent) and data entry of Phase I data of all patients in the different research centres was calculated. Results showed that the centres with a lower error rate entered the data within the first four days after the patient was initially seen. This confirms the hypothesis that the longer the time between data entry and data collection, the more errors are likely to appear in the eCRFs (see Table 3-3).

Table 3-3 Results PMT data timeliness

Centre	Total time in days	Total records	Average days of data entry after data collection (Standard deviation)	Total errors in % (see Table 3-2)
Centre A	224	57	3.9 (53.26)	0.34
Centre B	588	59	10.0 (20.74)	1.73
Centre C	184	91	2.0 (10.44)	0.46

The results show, that Centre A needed an average of 3.9 days to enter data after the data collection, but with a very high standard deviation. In contrast, Centre C needed 2.0 days with a low standard deviation. Compared to the validity, centre C also had a very low rate of total errors in their data set (0.46 %). It is noted that for a small sample size a large variance is expected.

Completeness of the research registry

The primary objective of this assessment was to provide an insight as to whether the research centre enrolled all patients available in their region. This calculation was based on an experimental approach, since checking the registry completeness would require that other comparison registries were assessed to draw conclusions. This assessment explored whether centres enrolled a balanced group of patients or only specific groups of patients. As a result, a comparison of inclusion criteria was performed as shown in Figure 3-5.

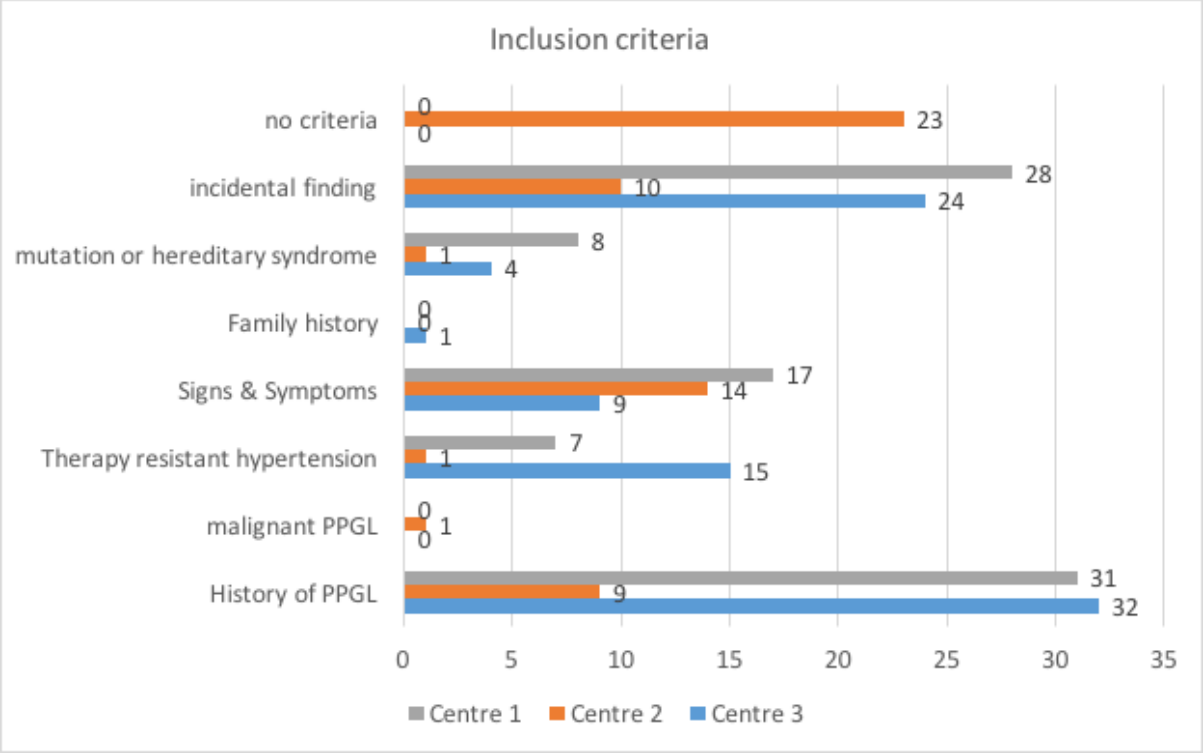


Figure 3-5 Comparison of inclusion criteria in the PMT Study

The analysis shows that 23 patients from Centre 2 were entered with no inclusion criteria. All records transferred from the ENSAT Registry to the PMT database. Centre 1 and Centre 2 recruited the majority of their patients via information gathered during routine care (history of PPGL). Also both research centres enrolled patients with incidentalomas (benign, non-hormone secreting adrenal tumours). The results show that there was no general trend in the subject recruitment. This might be due to the fact that Centre 1 and Centre 3 are both specialist adrenal tumour centres and researchers have their own lists of patients with known pheochromocytoma.

In summary, registry completeness, timeliness, validity and comparability are data quality dimensions that should be measured in clinical research registries. To draw significant conclusions more patient information is needed for assessment of registry completeness. In discussions with ENSAT researchers, an overall score or values that show the data quality for one specific study was requested by both the principal investigators and the clinical research fellows. This score provided an arithmetic mean of quality indicators that impacted directly on the future research results of this trial and the ability to translate the results of the study into a clinical setting.

It was also identified by the researchers, that another important improvement was in relation to modifications to the web-based PMT Study user interface to support the re-use of data and increase the data completeness. Due to the multiple forms that are mandatory or non-mandatory for the diagnostics of PPGL, users need to easily determine whether the form has been completed (green), is incomplete/empty (red) or only partially complete (yellow). This information has subsequently been supported via a coloured bullet point in the record list (see Figure 3-6).

Figure 3-6 PMT Study Completeness Visualisation for Phase 1 Patients

Exploration of the evaluation of data quality and modifications of the user interface of the PMT-Study provides important results for further investigation. The remainder of this chapter considers the evaluation of data quality for the EURINE-ACT study. Here a data quality score was calculated to motivate the researcher to increase the data quality when entering compulsory data into the study specific eCRFs.

3.1.4 EURINE-ACT Study

A systematic centralized data monitoring study was undertaken to review the data completeness and data accuracy of the EURINE-ACT study using data derived from the ENSAT-CANCER registry. This case study was supplemented by on-site monitoring visits involving six selected ENSAT centres to better understand the potential reasons for data inconsistencies and overall data quality.

In a clinical context, there can be many reasons for incomplete information that can only be found through a local (face-to-face) investigation of data processes (on-site monitoring). To gain further insight, interviews with the investigators and the local data managers need to be conducted with regard to low data quality and potential reasons for incomplete data sets. In total, eight on-site monitoring visits involving 8 local investigators and 7 data managers responsible for ENSAT registry data entry were undertaken in 2014.

In these visits, researchers provided insights into their work, which typically involved combining routine clinical care with biomedical research. Researchers from Italy and France identified that there was no governmental funding for research on benign diseases. Studies such as EURINE-ACT thus provided no direct funding (salary) for researchers. This is unlike ENSAT-CANCER, which had direct EU FP7 funding. Therefore, the motivation to enter research data after 8-10 hours routine care a day is important to understand. Furthermore, based on the researchers that were interviewed, the majority were young female clinicians at the start of their career. Their private lives (family), and continuous need for residency or sub-speciality training impacted directly on their availability and motivation for additional research work. A second problem that appeared during the structured interview was, that

data managers typically were unaware of the actual objectives of the trial. Rather, they were just tasked with entering (incomplete) data by the lead investigator at that site without fully knowing the reason for doing so.

To tackle these problems as part of the face-to-face meetings that were conducted, every staff member was introduced or reminded about the study protocol and the aims and the outcome of the trial. This was done to show the bigger picture of the study objectives. The initial results of a first data quality evaluation were presented with suggestions on how to face the data challenges of each of the visited centres.

Table 3-4 shows selected compulsory items from the ENSAT dataset and the EURINE-ACT eCRFs. Here ACTH (Adrenocorticotropic hormone) is a hormone that is measured during a Dexamethason Test (DST), a test that checks if an adenoma is hormone-secreting (Nieman, 2010). The dataset also includes information on whether imaging was undertaken or not and if yes, more related characteristics were required to be entered. It is recommended to perform a DST for every patient including those patients with suspected ACCs (Terzolo, Bovio, Pia, Reimondo, & Angeli, 2009). The data set showed the example where the data manager entered the value 'Not Done' for ACTH. Such an entered value can be defined as complete. However, ACTH is an important criterion for external judgment about the behaviour of adrenal masses and can provide additional information regarding routine care. Similarly, CT Density 'Not Done" is not clinically useful information and should generally be avoided.

Table 3-4 Example Calculation of NAPACA Quality Score

Item	Value	Completeness	Accuracy
ACTH	Not Done	✓	✕
24h Urine Cortisol	[Select…]	✕	✕
Imaging Form	<exists>	✓	✓
Imaging Method	CT	✓	✓
CT Density	Not Done	✓	✕
Ki 67	1%	✓	✓
Weiss Score	Not calculable	✕	✕
Score		**5/7 = 0.71**	**3/7 = 0.43**

In the given example in Table 3-4 the record for the selected items show a completeness score of 71% and an accuracy score of 43%. Using an arithmetic mean of both values, a basic DQ score can be calculated through the steps outlined below. It is important to note that for this result, every item has the same importance. Compulsory items are represented in the Record Eligibility Score (see below), which should be part of the data accuracy feedback.

Report of Data Completeness

Data completeness determines whether all necessary clinical trial information is entered into the study-specific eCRFs. It does not measure/record whether the entered data is actually correct. Rather it gives information regarding the rigour of performance of data collection and the assiduousness of data entry personnel at particular research centres. When data completeness is low, critical gaps should be identified and resolved in the data collection and data entry processes. This can for example, require improvements in local training requirements.

In the first completeness calculation, only the most important imaging, pathology and identification items of the EURINE-ACT eCRFs were considered. The analysis was performed for ACC and NAPACA records separately.

The data completeness of ACC records showed a lack of pathology information. It is assumed that every patient without metastasis will undergo surgery and therefore will have information about the tumour tissue. A simply plausibility check was added, if the item 'Distant metastasis' is 'No', then the Weiss Score and Ki67 must have a value entered. Of the 193 patients without metastatic spread of cancer cells, only 68% of pathology information was entered. Nearly all records showed no information about imaging (see Figure 3-7).

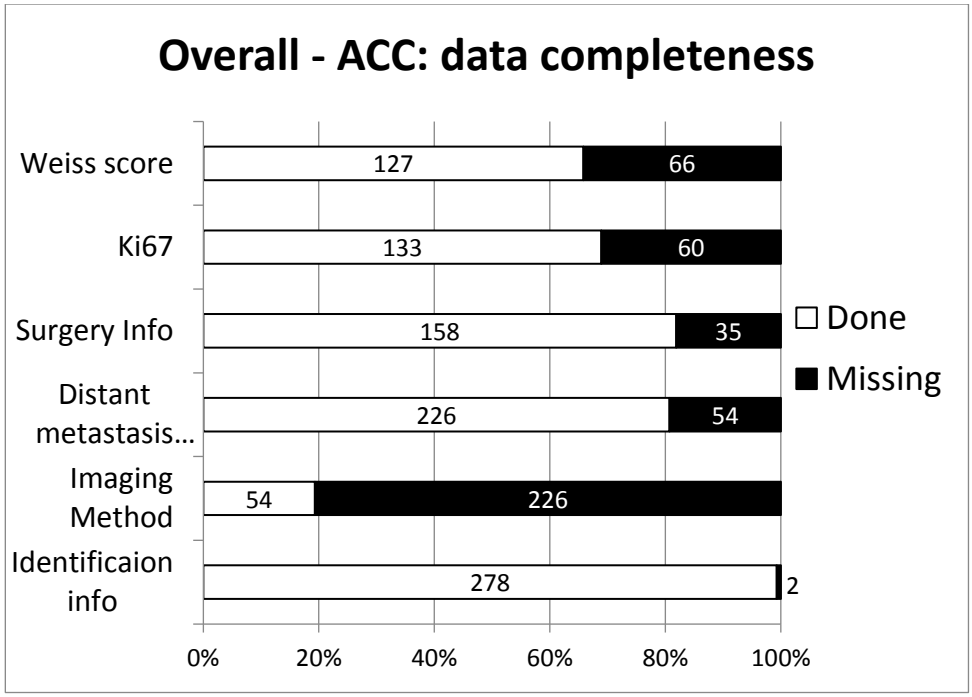


Figure 3-7 Overall ACC Data Completeness

Benign lesions must also have imaging information. It is the case that all NAPACA tumours are detected and ultimately classified as benign – this is the specification of a NAPACA tumour. The entered values must therefore reflect this decision, e.g. so that an external reviewer can make a similar judgment based on the entered characteristics. Compared to ACC records, NAPACA records showed a better overall completeness, but mandatory items regarding the detection of the tumour were still missing in every third case. The data completeness assessment indicated that in nearly 30% of all patients, no DST was performed or it was not entered (see Figure 3-8).

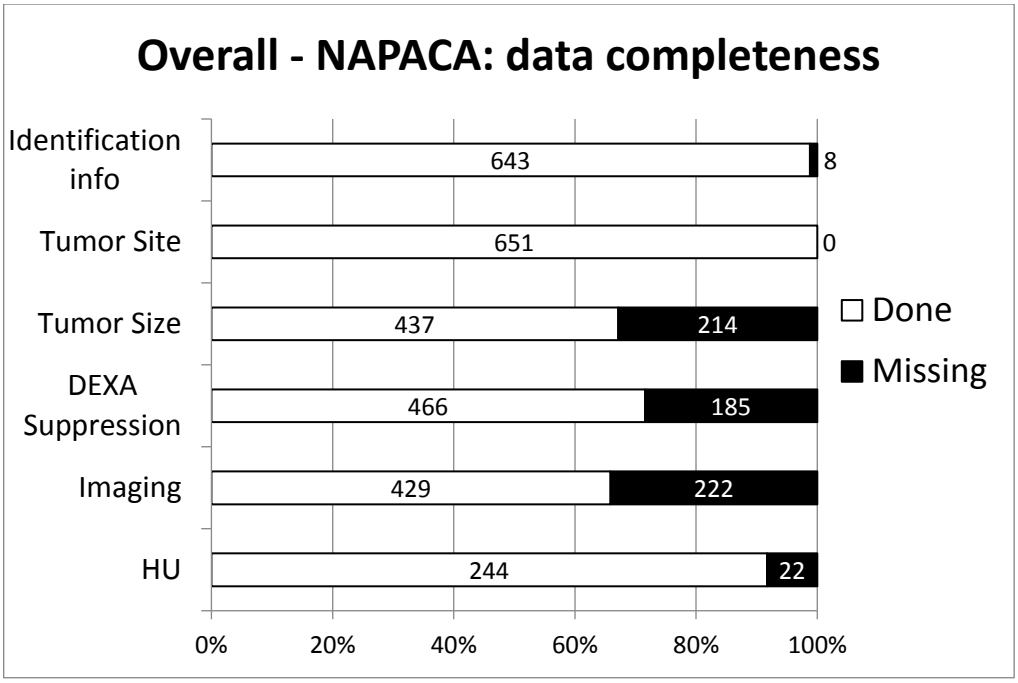


Figure 3-8 Overall NAPACA Data Completeness

Report of Data Accuracy

Each monitoring visit includes an interview with staff members that are responsible for performing data entry. This is followed by source data verification steps and the direct witnessing of local biomaterial sampling for EURINE-ACT. For every site visit, one and a half days were scheduled to assess local processes and organisational structures (Donabedian-model). For the outcome quality, all EURINE-ACT records were checked as to whether they could be confirmed as EURINE-ACT. For this, inclusion and exclusion criteria were checked. NAPACA and ACC records were assessed separately. For NAPACA the records had to meet the following five criteria:

- biomaterial was received in the investigator site;
- imaging reference standards were entered;
- tumour growth was lower than 20% within 6 months;
- in case of surgery, the Weiss score and Ki 67 were checked, and
- imaging follow up information was included.

An additional four mandatory NAPACA registry items were also monitored: tumour size, DEXA Suppression Test, Imaging information and if CT information was entered, then the Hounsfield Unit (HU) must be reported.

ACC records had to meet the following four criteria:

- the timestamp of the biomaterial collection (sampling) was required;
- whether the patient received chemotherapy or mitotane had to be documented;
- checks were required on the biomaterial availability in the biobank, and
- the patient follow up status had to be checked.

The ACC items were also checked with regards to tumour size and if the patient had no metastasis, then the entered surgery and pathology information was required to be marked as compulsory.

Once a range check was applied on every single compulsory data value, an accuracy score could be calculated. In addition to the data accuracy calculation required for EURINE-ACT, it was necessary to check whether the record met the eligibility criteria for the study. Patients with adrenal tumours that participate in clinical trials are entered into the ENSAT registry. Rules apply on these patient (registry) records, whether they meet the recruitment criteria for EURINE-ACT.

As outlined in Figure 3-9, once a record has been flagged as NAPACA, the biomaterial must be received and stored in biobank. If this condition is fulfilled, then it needs to be checked as to whether an imaging characteristic was given in the patient (registry) record. Imaging characteristics are mandatory for every patient with an adrenal mass and hence information about the existence of a CT, MRI, PET or X-Ray must be given (Blake, Cronin, & Boland, 2010). The next check was whether the patient underwent surgery. If this is the case, it is important to identify whether this record was marked as an Adrenal Cortical Adenoma (ACA) or not. In the case of an ACA, it is mandatory to record information about the Weiss score and the associated Ki67 must be entered and be correct. These pathology scores serve as an external review indicator that are used to determine whether the patient has a benign or malignant adrenal tumour (Arlt, 2011). If all conditions are met, the record can be confirmed as eligible for EURINE-ACT. If one or more of these criteria are not met, then it might be due to the data entry person not having this information at the time of data entry, or their lack of rigor or a variety of other possible causes. The resolution of such data quality issues is essential to understand and processes instigated to identify and resolve such future issues.

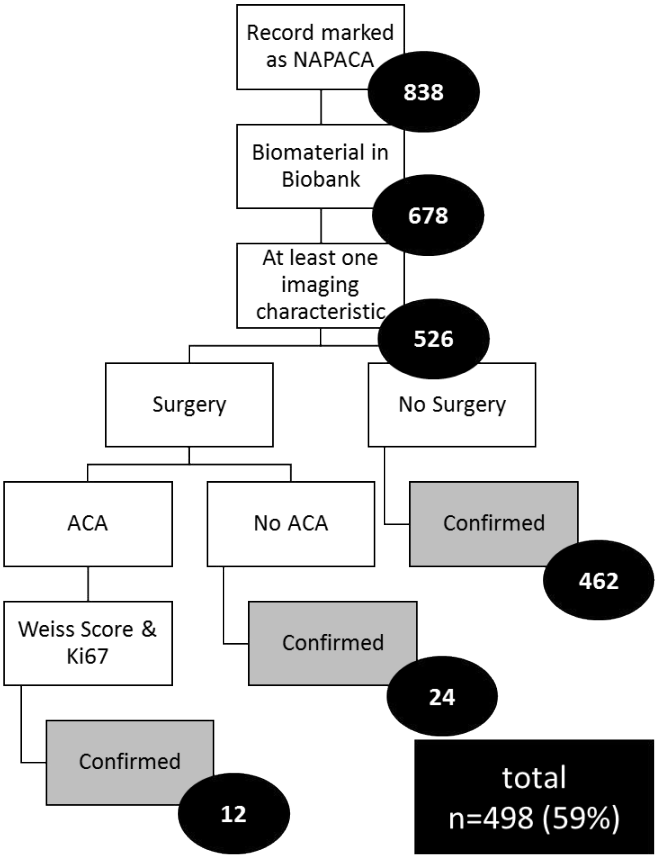


Figure 3-9 NAPACA Eligibility calculation

After the calculation it was discovered, that just around 59% of all entered NAPACA records could be confirmed as eligible for EURINE-ACT. However, the majority of records did not have information about imaging characteristics. Additional problems also appeared in patients who underwent surgery, e.g. there was no information about pathology given.

Also considered in the accuracy score were *plausibility checks*. As an example for malignant (ACC) tumours, it was checked if the patient with selected "No Metastasis" underwent surgery. It is highly recommending that every patient with a malignant tumour (diameter >0.5cm) should undergo surgery, to prevent spreading to other organs. If no surgery and/or pathology information was entered, an error was given to this record.

Report of the Data Quality Score

Data quality in this chapter was considered as the combination of data completeness and accuracy. In later chapters we consider other aspects of timeliness and comparability that should be included in this calculation. A threshold must be also defined for every individual score to detect conspicuous and non-conspicuous values. In the present version, a total score was added in the feedback, so that the site could identify whether they achieved a score below or above the average (compared to other centres). Centres with less than 10 NAPACA records where not considered in Table 3-5. The displayed DQS is the mean of data completeness and record eligibility.

Table 3-5 Community Feedback NAPACA DQS (Oct 2014)

Centre	Records	DQS
PLWW2	92	.97
GRAT	116	.96
ITTU	71	.93
GBBI	140	.89
NLEI	17	.88
CRZA	56	.83
GYWU	40	.68
GYMU	84	.61
BGSO	51	.43
GYBN	96	.41
FRBO	18	.37
TOTAL	838	.73

It was discovered, that the majority of the records achieved a completeness of above 60%. In contrast to this, the accuracy for two centres (n=2) with less than 10 entered records were calculated to be 0%. It was assumed, that when sites want to actively contribute, they mark the records as trial related, but

they do not always add any other study items, i.e. trials that they are not involved in. This underlines the need for training of local staff and processes in data quality management, but also need for better data management tools.

3.2 Monitoring Visits

Clinical monitoring can be based upon centralized monitoring or decentralised (on-site) monitoring visits. Such visits are used to assess predefined quality indicators that can have an impact on the overall data quality. Nahm (2012) described six major components that are mandatory to inspect during an on-site visit:

- Capable processes;
- Appropriate tools;
- Personnel management;
- Project management;
- Documented work processes, and
- Consensus on data quality requirements.

For example, the assessment of 'personnel management' includes checks on staff training and relevant documents, job descriptions, employee feedback through to overall performance evaluations. To identify issues and inconsistencies in data quality processes, quality initiators need to be defined. For the PMT and EURINE-ACT clinical studies on-site monitoring visit indicators focused on all data process levels that could have an impact upon data quality. For every level a quality indicator was assigned (see Table 3-6).

Table 3-6 Quality indicators for on-site visits

Data process level	Quality dimension (Donabedian model)	Quality indicator (Deming cycle)
Data collection	Structure	Source documents
		Documentation (SAEs)
		Biobanking and storage
		Staff training
		Collection in accordance to SOPs
		External reports (Pathology/Surgery)
		Funding schemes
		Infrastructure of research unit
		Recruitment methods
Data capture	Process	Access to devices to enter data
		Time of data capture
		Hardware and software issues
		Method of data capture
Data analysis	Outcome	Data quality

On-site visits were performed in three German PMT participating research centres (Wurzburg, Munich and Dresden). For the EURINE-ACT study, international monitoring was demanded and the on-site visits included Wurzburg, Munich, Paris, Birmingham, Padua, Turin, Florence and Athens. To allow fair comparison of these centres, the national health services for routine care and research policies as well as their associated funding schemes was considered. Every hospital has its own SOPs for routine care. Similarly, every European country has different approaches with regard to data capture and

processing in hospital systems including the adoption of technologies and heterogeneity of IT solutions. In this context, we consider how the monitoring visits could be compared.

3.2.1 PMT On-site Monitoring Visits

Wurzburg and Dresden both used paper-based (non-digital) sources for data capture and patient information management. Similarly, all trial related information of patients is stored on separate paper documents in clinical trial-specific folders. Munich however uses digital clinical data management systems provided by the hospital as their primary information source. When data items are requested for a particular clinical trial or study such as PMT, the user who enters the data into the PMT eCRFs has to search the hospital patient records for the relevant data items and enter them (copy/paste) into the PMT eCRFs. Furthermore, all centres have a separate folder (paper/digital) where they store the trial master file and related documents like ethical approval and forms for Serious Adverse Event (SAE) reporting. All three centres have an assigned data manager (research assistant) to support the actual capture and entry of data into PMT eCRFs.

It was identified that the data managers in Wurzburg and Munich have multiple (seven) other trials where they are expected to enter data, whilst the data manager in Dresden was just assigned data entry needs for the PMT, EURINE-ACT and ENSAT registry. The research unit in Munich also specialised in several endocrine disorders (secondary hypertension, Diabetes type I & II, Conn syndrome, osteoporosis, Cushing's syndrome, PPGL, ACC, Incidentaloma). This site is different to the research facilities of Dresden and Wurzburg where the focus is primarily on disorders of the adrenal gland.

The recruitment type and technical forms that are used is also important for data quality assessment. For comparability, a homogenous group of patients is important criteria for good clinical research. All centres have a long history of adrenal research and have data that they have collected – often over extended time periods. Often this is through Excel spreadsheets for former patients with a history of PPGL. Such spreadsheets are a key source of recruitment for registries and studies (subject to consent). Furthermore, in every research centre, all clinicians conduct endocrine-related outpatient clinics. Through such clinics they know first-hand, the inclusion and exclusion criteria of particular trials and if patients they see match the criteria for particular studies that might arise. A third possibility for recruitment type can also be through high values of metanephrine in central biochemical analysis laboratories of the hospital. Due to security regulations and the approach taken within ENSAT-CANCER for aggregation of data, a direct search of people that match the PMT recruitment criteria is typically not permitted across all of the hospital clinical data management systems. Such heterogeneity of systems and processes in individual healthcare facilities is common and remains a challenge for improved overall health care.

In conclusion, all centres can be considered as having equal conditions with regards to structural quality. All centres have experience with adrenal patients and have several possibilities to recruit patients. Special assigned data managers are available that have similar data entry training. Furthermore, every office where the patient is seen has direct access to the PMT eCRFs. Only Munich uses a direct manual transfer of data from the electronic hospital records to the PMT registry. This can have impact on the data quality, but a more detailed analysis will be described in chapter 4.

3.2.2 EURINE-ACT On-site Monitoring Visits

For the EURINE-ACT on-site monitoring visits, the experiences from PMT were adapted and a general SOP monitoring protocol was developed (see Appendix I). In addition to the PMT structure and data processing assessment, a source document verification (SDV) of 3-5% of the records was performed. This sample size is in accordance to European recommendations (ECRIN, 2010).

Centralized data monitoring provided an overview of the data items with low data completeness in the EURINE-ACT eCRFs. Pathology and surgery information was rarely entered. Therefore, all data processes with regards to tumour tissue sampling and analysis was assessed. Every visit was started with a presentation to local investigators and associated research staff regarding the centralized monitoring results. Information about the overall trial objectives was also discussed. After the presentation a guided tour was performed. This included tours of the wards where the patients were enrolled into the protocol and the laboratory where biomaterial was aliquoted and analysed. The most important discovery was that all patients that have a suspected tumour in France are hospitalised for a one or two overnight stay. This has the added advantage for sampling of overnight urine – which is key to the needs and demands of the EURINE-ACT study.

Another important outcome identified was that different countries have different hospital SOPs and funding structures. In France, only malignant tumours are funded by the government, whilst in Germany, a general funding scheme exists for the treatment and support of all tumour types. In Italy all users that enter the data are young female clinicians that have routine care from 8am to around 6pm. It was identified that they enter data into the registry only after their day-jobs. It was also discovered that their research related work was also not financially covered. Therefore, only limited individual motivation was discovered through personal interviews. It was identified that data inconsistencies and low completeness rates of surgery and pathology information were caused by diverse reasons. For example, the centre in Paris conducts research that assesses tumour tissues and directly interacts with the researchers who enter data into the registry. In all other cases, pathologists and associated laboratories were not located in the hospital and instead they received (via post) the tumour samples for analysis.

3.2.3 Conclusions

In conclusion, visiting the centres provided detailed feedback and insight into the practices and processes related to data quality. The auditors gained insights into local processes, how data inconsistencies had developed and the day-to-day reality of data collection and data entry. A primary issue stemmed from the data processes of external providers and stakeholders that analyse biosamples. A secondary issue was the motivation of the data entry user and the realities of data entry often being in addition to the day job (mostly unfunded). While funding and the combination of routine care and research is typically overly cumbersome for the scientist that has to care for patients and trial subjects, their staff also and the roles they are actually funded for do not always reflect the needs and demands of good quality data entry into international disease registries and trials such as PMT and EURINE-ACT.

It was identified that the visits were overwhelmingly beneficial because of the background presentations about EURINE-ACT and highlighting of the results of the centralized monitoring. Experiences of other centres could be shared and small problems in the local data process could be resolved. The on-site monitoring shone a spotlight on day-to-day challenges facing all sites with regards to data entry.

In this context, it is essential to understand the role of motivation and the way in which technology can help in improved data quality capture. Staff members who are overworked and underfunded can easily be demotivated, if they see that they are performing less well than other centres. The roll out of technical solutions was thus carefully orchestrated to ensure that all aspects of data entry processes and reasons for potential issues were understood.

3.3 Other Examples of Data Quality Assessment from other Studies

Data quality feedback is not a new approach in the scientific literature. The novelty here is to apply selected quality indicators to motivate researchers to increase the overall data quality in disease

registries and eCRFs to increase the usefulness of the data. Key factors to understand are what quality indicators are best for particular research environments and how can they be optimally reported to researchers.

Data quality reports should be continuously updated and fed back to data entry personnel and the research (clinical) collaboration protagonists. Ideally, tools that are used by everyone should be adopted, e.g. mobile phones or laptop/PC. By far the most important ubiquitous technology is the mobile phone, which provides a model for always-connected research and the ability to receive targeted notifications. Therefore, an approach for data quality reports should ideally focus on notifications, email reminders or alerts. Ideally data management systems should be accessible via mobile applications (apps) in accordance to data security standards.

Little research has been published in the field of feedback reporting and its efficiency. Stausberg et al. (2006) described feedback as ”periodic reports about data quality provided to those […] doing the input” (Stausberg, Nonnemacher, Weiland, Antony, & Neuhäuser, 2006). A literature review published by de Lusignan in 2005 listed ten characteristics for the improvement of data quality in primary care:

- Motivation of professionals to have a positive attitude of their structured computer data;
- Working with lead clinicians receptive to evidence-based quality improvement initiatives;
- Respect for the “clinical judgement” (phronesis) of experienced clinicians;
- Using informatics as an enabler of quality improvement;
- Using education as an appropriate change agent;
- Data quality feedback using parameters with a positive predictive value and high sensitivity;
- Personally provided feedback, by a skilled facilitator, within the workplace;
- Professionally led programmes, supporting local clinical champions;
- Alignment with national, evidence-based, quality improvement programme, and
- Financially incentivised data of higher quality. (Simon de Lusignan, 2005)

In all of these areas tools are needed that support clinicians that allow to keep them continuously aware of the importance of data quality with focus on the implementation of professional data management systems that meets standards and improve data interoperability (Simon de Lusignan, 2005).

Feedback systems in CDMS should support quality improvement purposes. There are a variety of feedback mechanisms published in the medical literature, that provide quality feedback during data entry and via email reminders (Herzberg et al., 2011; Porcheret et al., 2004). In the literature feedback also refers to the term ‘reminder’, that informs the user about missed documentation, displays summary reports and provides real-time/immediate feedback (Nair, Newman, Peterson, Wu, & Schwid, 2010; Wurst, Lamia, Schlundt, Karlsen, & Kuhn, 2008). Summary reports can be delivered to the research community that count as a reward that encourages the researcher to enter higher quality data. It was also discovered that passive cues during data entry increase the compliance with data entry more than hard stops (Nielsen, Peschel, & Burgess, 2014).

The effectiveness of such processes in the domain of clinical trials with manual data entry and dual source systems (which is the situation in the majority of cases) has not been extensively investigated. This research has focused on a simple and generalizable method to improve the data quality by proposing to distinguish feedback into *real-time feedback*, *investigator feedback* and *community feedback*. In combination with suggestion/reminder features, the research has explored whether feedback can improve data quality in research registries. To give a lay description of the three feedback mechanisms we identify:

- For data entry users - *real-time feedback*, a user receives feedback during the data entry about the entered data and whether it meets the minimum research requirements,
- For local investigators - *investigator feedback*, where all information about the current usefulness of data is given to reveal potential data inconsistencies of the captured data from the given sites, and
- For all researchers - *community feedback*, where quality reports are provided to the research community to encourage the staff to collect higher quality data than other sites.

Each sort of feedback will report its own selected type of quality dimensions. The feedback collectively can serve different purposes to increase the overall data quality and user motivation.

3.4 Demonstrating that Data Quality Increases through Feedback

To determine whether data quality improves based on encouragement of quality assurances, a final calculation was performed three years after the first centralised monitoring. Here, an investigation of data completeness of all phase one records was performed. The first evaluation was performed 2 years after the start of the study protocol (October 2013) and the second evaluation performed 3 years after the first calculation (October 2016). Several data items were changed during the three years, which resulted in a new data completeness calculation for October 2013. It is noted that not all research centres were included in the analysis, since they have no reference value, or they did not participate in the PMT-Study in October 2013.

Results showed that two centres (including the centre of the principle investigator – GYDR) decreased their data completeness (DC), but entered more than 200 records in the registry. Whilst the overall data completeness of PLWW is below the average, they increased their data completeness slightly and entered 597 records within the three years (see Table 3-7).

Table 3-7 Trend of data completeness (DC) of the PMT Study

Centre	Records Oct 2013	Records Oct 2016	+/-	DC Oct 2013	DC Oct 2016	+/-
GYDR	91	291	200	0.88	0.81	-0.07
GYLU	2	36	34	0.39	0.83	0.43
GYMU	71	263	192	0.61	0.76	0.15
GYWU	92	313	221	0.84	0.50	-0.33
ITFL	9	9	0	0.30	0.30	0.00
PLWW	503	1100	597	0.58	0.61	0.04
TOTAL	768	2143	1375	0.64	0.66	0.02

This evaluation of PMT provided an explorative case study to gain insights into data quality evaluation. Feedback was given just to the principal investigator (investigator feedback), who planned quality control procedures to improve data quality. An example was given with the assessment of outliers in a healthy cohort. This calculation detected that GYWU, GYMU and PLWW were not performing the blood withdrawal in accordance to the study SOP. If regular feedback to the research community and to the principal investigator works, we should see a similar trend calculation for EURINE-ACT.

To assess the efficiency of the performed case studies, the data quality of EURINE-ACT was monitored for 6 months. In April 2014, the first quality report was sent to the research community via email. This score was developed in March 2013 and implemented in EURINE-ACT in November 2013. One year later, the ENSAT-Consortium met for a general Assembly in Nice and results and trends of the DQS were presented. For this purpose, a trend/efficiency of the DQS was calculated. Here two quality dimensions were calculated separately to build an arithmetic mean in the form of an overall DQS. First

it was investigated how the completeness, accuracy and overall quality scores changed through the year.

Table 3-8 Trend of Quality Scores

	Apr-14	May-14	Jul-14	Oct-14
Data Quality	.60	.64	.69	.73
Data Completeness	.76	.82	.83	.86
Data Accuracy	.45	.46	.55	.59

As shown in Table 3-8 the overall data quality increased. After the implementation of a data quality score, the overall quality of record information increased by 13%, while the accuracy changed by 14% and DC grew slightly by 10%. The slow increase in data completeness was caused by adding more data items over the time period of the implementation. The ability to strike the balance between overall completeness and quality should be aimed at encouraging good behaviour of site staff entering clinical data. The more items included in the data completeness calculation, the increase in effort required and the danger of potential lower data accuracy scores. Alternatively mandating that all data is compulsory and has to be completely accurate is often difficult to enforce: many registries and studies can contain thousands or tens of thousands of data items per patient. A balance is thus needed. In this review of ENSAT centres a spectrum of data quality was discovered. It should be noted that only centres with a data quality score over 50% showed interest and asked for more support in how they might improve their data entry processes.

To more closely investigate whether data quality issues appear in reporting or other data processes, an error count per sub-record (diagnostic, imaging and pathology) was conducted (see Figure 3-10). Counting errors or calculating an error rate is an established method in quality assurance in clinical trials to discover problems in data processes (Rostami et al., 2009). It was revealed that imaging and diagnostic problems could be resolved by improvements to data quality. However, pathology errors did not decrease over the given timeframe. One major reason why reports about Weiss Score or Ki 67 could not be entered by local teams was due to funding problems and missing centralized pathology reviews. Furthermore, it was often not possible to obtain these values retrospectively.

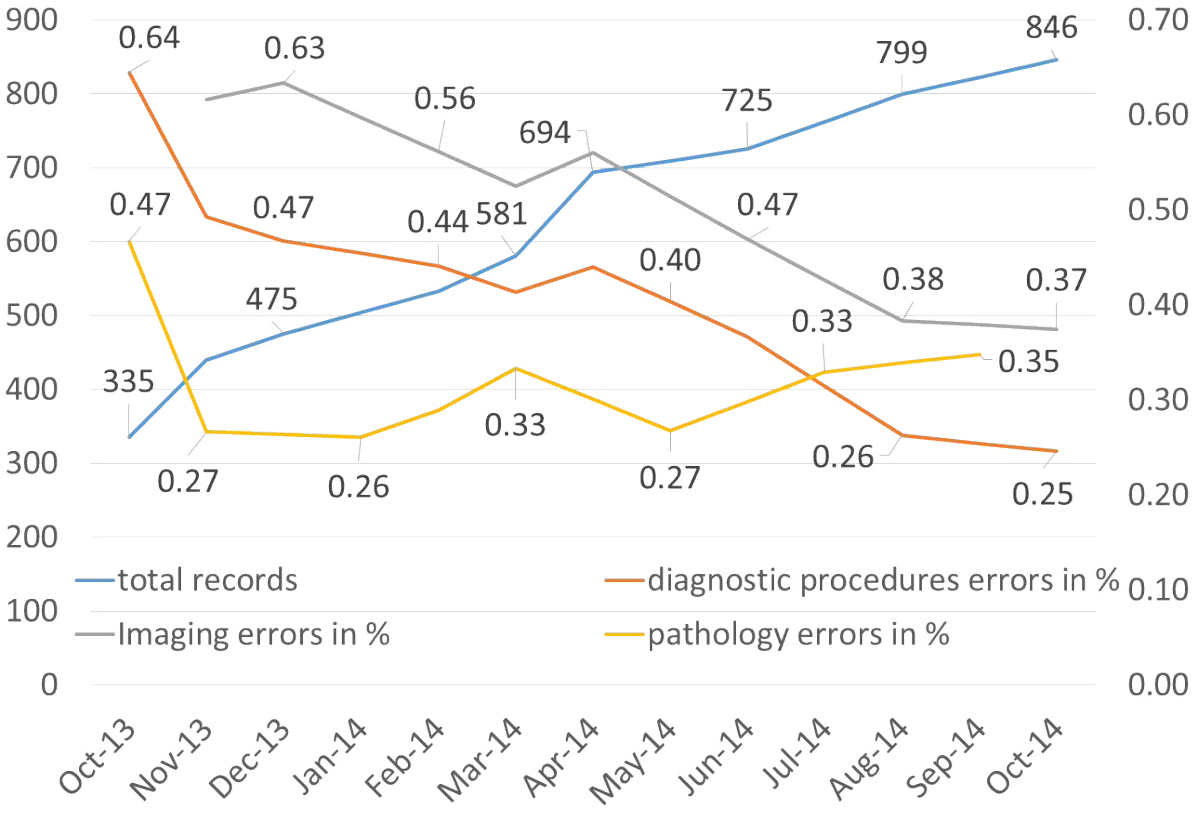


Figure 3-10 Trend of selected sub-records

In conclusion and importantly for this thesis, EURINE-ACT showed a data quality increase. Feedback on a regular basis to the investigator encouraged the majority of sites to improve their data entry practices and data entry processes. This improvement was the result of a combination of feedback, monitoring visits and interactions via email. Thus we have established that direct/manual feedback is one way of improving data quality. However, such improvements are expensive. This is especially the case with site visits in international settings. In the rest of this thesis we thus explore whether such improvements can be achieved by automated (web-based) feedback.

Additionally, given the above case study, whilst community feedback encouraged some hospitals and users to increase the collected data quality; tools for user motivation and improving data quality were not fully established. That is, whether all three combinations of feedback were fully effective could not be shown by EURINE-ACT since real-time feedback was not available given the nature of the EURINE-ACT study.

3.5 Chapter Conclusions

The evaluation of data quality in the area of European adrenal research has provided important insights for the development of a data quality score. Researchers agreed that the data within the ENSAT registry needed to be more useful for other research initiatives. Therefore it was proposed that increasing the data quality with ongoing data quality feedback would increase the usefulness of data to facilitate other clinical trials. A data quality score that quantifies the usefulness of information entered into disease registries or eCRFs of a clinical trial is highly beneficial. We discovered a significant impact in both investigator and especially community feedback with regards to data quality. Community feedback needs to address intrinsic and extrinsic motivational factors of data entry users. Additionally, we identified that the development of a record eligibility score was important to show the community compulsory trial items. Presentations and discussions related to this score received

considerable interest in regards to data quality in clinical data management systems from the ENSAT community at international meetings. Quantifying quality is an important method for quality control of clinical trials.

The analysis undertaken in this chapter was based on centralized monitoring augmented by on-site monitoring visits. We have demonstrated that face-to-face feedback works and has been quantifiably demonstrated, however traveling to all centres, giving presentations, interviewing and training researchers is both very costly and time consuming. On-site monitoring can motivate the users to improve the performance in regards to information processes. However as shown, the motivation did not last very long. Again as shown, on-site monitoring efforts can help to discover just 5-10% of general data quality inconsistencies. The majority of all errors can be detected via centralised monitoring tools. Issues about travel costs, spent time and short lasting motivation leads to the need for other models. In the remaining chapters we describe the extent that web-based applications that give automated feedback to the data entry user, the investigator(s) and the research community involved in the study collaboration can be used to replace the human-driven and expensive local on-site monitoring efforts.

4 CHALLENGES TO SUPPORT AUTOMATED FEEDBACK

This chapter explains the challenges that automated feedback needs to address in data management systems. Most importantly are criteria associated with end users and their motivations for data entry. To understand this, a survey was performed within the ENSAT consortium to gain insights into motivational factors that could influence the quality of data entered into the ENSAT systems by data entry and related personnel. As noted previously, data quality is crucial for registries that ultimately need to provide useful data for current and future clinical trials and researchers with diverse research interests. An assessment of users of the ENSAT registry was performed with specific focus on determining whether data completeness of special tumour types for registry data was better (more complete) than those associated with ENSAT-CANCER related clinical trials such as PMT or EURINE-ACT. Independently, from the result, methods of data quality reporting have to be addressed to understand how users can be motivated, e.g. by leaderboards, graphs or dashboards, and ensure that any kind of feedback is not demotivating. Finally, this chapter discusses how research registries with data quality feedback can improve the data quality and thereby improve the quality and conduct of registry-related clinical trials.

4.1 Survey

In order to prepare for EURINE-ACT onsite monitoring visits, a survey was conducted to identify any causes that might influence the data quality during data entry motivation. The survey was developed with the open-source GoogleForms platform. This application provided the option of a personalized survey development without subscribing to a website or company. The final survey was easily accessible to all users that have the survey link, even without registering with Google or having Google accounts. The users were also able to access the service on all platforms (Android, Windows, Mac, iOS) and on all general bowsers (Chrome, MSIE, Firefox, Safari). As such, all participants were able to submit responses.

An email was sent to all active ENSAT-members, with an invitation to participate in the survey to investigate their local data entry processes. The survey link was also provided to all users on the internal ENSAT registry start page (after the user login). It was assumed that the persons that answered the survey within the first few days were highly motivated to improve the ENSAT registry. The survey was anonymous, i.e. the names of the respondents were not requested or given however the centre that they are data entry personnel for was required.

A specific identified bias of the survey was the possibility that a user could submit a survey more the once or that other non-ENSAT individuals could have completed the survey if they discovered the link. This could have been countered with the possibility that every user had to log into their Google account (or set up a new Google account). The former option was not considered since it was identified that the login could be a barrier of users completing the survey. Also due to the presumed high motivation of every researcher in the ENSAT consortium it was assumed that there are honest and submit just one survey. It was also felt that any non-ENSAT user would be unlikely to want to complete the survey.

4.1.1 Evaluation Method

The survey was conducted with multiple choice and open questions to verify the cause-effects displayed in Figure 4-11. The primary purpose for the survey was to answer questions related to potential low data quality in clinical institutions or indeed understand why certain institutions have higher data quality. There are several hypotheses that were tested through this survey. Specifically, we considered that data quality could be influenced by:

(1) The individual's personal characteristics such as their age and gender;

(2) The specific individual role in the project and whether they are explicitly funded for data entry or not;
(3) Whether the centre is a clinical trial centre or a participating trial centre and is thus aware of the importance of data quality;
(4) If the centre is participating in registry-related clinical trials or not and if so, whether they have to enter data into clinical trials-related eCRFs;
(5) The type of source documents (paper or electronic);
(6) The time length of data capture;
(7) The time between data collection and data capture, and
(8) Whether the same data had to be entered multiple times in other registries.

The survey was concluded by two open questions focused on the individual's personal motivation for using the registry and recommendations about features that should be added to the registry to help improve the data entry processes and ultimately improve data quality.

To encourage the user to complete the survey, where appropriate, the questions provided default suggestions to elicit user responses such as "it is my job" or "I want to cure cancer". In case these default answers were unchanged, the responses were ignored in the final analysis for the survey. Open questions were analysed in a recursive process using an approach based on thematic text analysis (Braun & Clarke, 2006).

4.1.2 Survey Results

The survey was analysed after 28 days. In total 220 ENSAT participants received the first email requesting them to complete the survey. On day 19 after the initial announcement, a reminder was sent to all participants. In total 45 responses from the participants were submitted, giving a response rate of 19%. The peak response periods of the submission were the two emails that asked for participation (day 1 & day 9) and the first Monday (day 4) when the survey went live (see Figure 4-1).

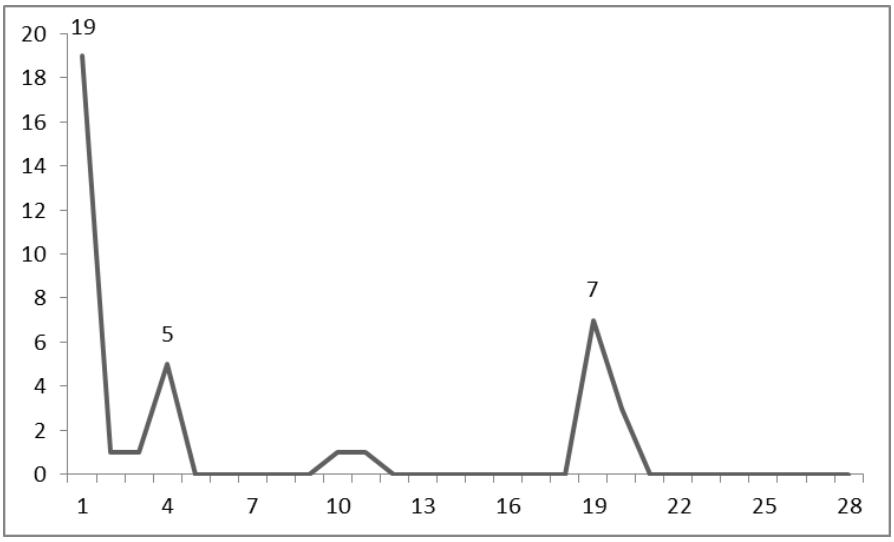


Figure 4-1 Timeline of Survey responses

The largest groups that submitted completed surveys were clinicians, investigators and research assistants (see Table 4-1). It was identified that the largest group of ENSAT registry data entry users were female clinicians between 30 - 49 years of age (n=9).

Table 4-1 Overview of Survey Participants

	Female	Male	Total
Age			
16-29	5	0	5
30-49	19	12	31
50 and older	5	4	9
Role			
(Local-) Investigator	4	5	9
Research Fellow	2	2	4
Clinician	12	9	21
Lab Assistant	0	0	0
Research Assistant	7	0	7
Nurse	2	0	2
Data Entry Assistant	2	0	2

The participating centres that submitted the most surveys were Munich, Wurzburg and Florence (see Table 4-2). These centres were also visited during the EURINE-ACT monitoring visits. All three centres were identified as highly motivated to enter high data quality and involved in the conduction of the majority of clinical and biomedical trials associated with the ENSAT research network.

Table 4-2 Participating Centres (number of submitted surveys)

GYMU (6)	GYWU (3)	ITFL (3)	NLNI (3)	GYMV (2)	BZSP (2)	SBBE (2)	ITTU (2)
FRPA1 (1)	FRST (1)	GBBI (1)	GBDD (1)	GRAT2 (1)	GYBN (1)	GYBN2 (1)	ITBR2 (1)
ITMI2 (1)	FRLY2 (1)	ITTU2 (1)	NLEI (1)	NYBE (1)	PTCO (1)	CRZA (1)	SPMA (1)
SPSV (1)	TRIZ (1)	GBED (1)	GYDF (1)	PLKK (1)	PLWW (1)		

EURINE-ACT and ADIUVO are the trials where most survey participants confirmed their ongoing involvement and contribution. Both of these trials recruit the majority of their patients from the ACC component of the ENSAT registry, while EURINE-ACT also has a large control group with NAPACA patients that are recruited from the ENSAT registry. Other trials of relevance include PMT and FIRST-MAPP, which have a focus and demand on data quality for patients with NAPACA and Pheo/PGL respectively. In the case of PMT, NAPACA data is used as a control population to the ENSAT-registry. Importantly PMT supports bidirectional data transfers, i.e. data from the registry can be used (imported) into Phase 1 of the study, and other data collected throughout the trial can be fed back into the registry. The survey confirms the proportions of records as discussed in the adrenal research background, i.e. NAPACA, Pheo and ACC have almost equal amounts of patients in the ENSAT registry and these are used by and associated with the largest number of multicentre trials in the ENSAT research consortium (see Figure 4-2).

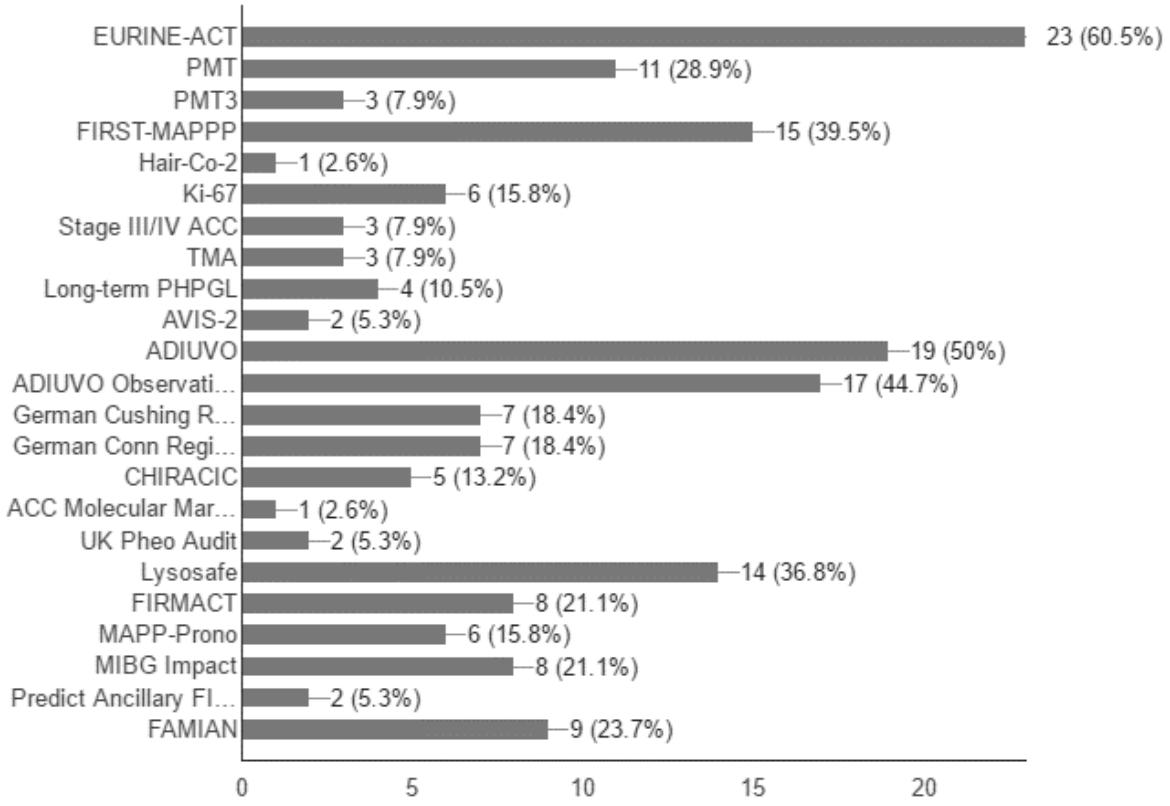


Figure 4-2 Overview of current ongoing clinical trials associated with ENSAT (October 2016)

The final analysis of the survey data should ideally provide a correlation between the data completeness in the registry and the survey answers. For this, survey questions were ranked. The answers were scaled between 1 and 3, where “3” indicates the worst possible way to receive good (complete) data quality and “1” shows that the environment provides the optimal approach for data entry and ensuring best data quality (see Table 4-3).

Table 4-3 Ranking of Survey Questions

Rank	1	2	3
Question 1 (Q1): *"Where do you access the ENSAT-Registry and enter patient related data?"*	*"in the office, where you meet the patient"*	*"in an office on the ward/unit where you meet the patient"*	*"in an office in your research facility", "at home"*
Question 2 (Q2): *"What is the Source of the Information you enter in the Registry?"*	*"Electronic Medical Records"*	*"Paper Based Records"*	*"you know the patient and enter the data out of your mind"*
Question 3 (Q3): *"In your opinion, how long does it take for you to enter a new patient record into the registry (with clinical information)?"*	*"under 10min"*	*"10-30min"*	*"more than 30min"*
Question 4 (Q4): *"In your opinion, how long after the patient visit do you enter the data?"*	*"within 30min after the visit"*	*"on the same day"*	*"within the week", "later"*
Question 5 (Q5): *"How many registries are you entering the patient data?"*	*"1"*	*"2"*	*"more than 2"*

Survey exploring data completeness confounders

Figure 4-3 compares 'good conditions' and 'bad conditions', where 'good conditions' consist of answers with rank 1 and 'bad conditions' answers with rank 2 and 3. After ranking the answers (Q1-Q5), every centre with more than 10 ACC records was matched to their own achieved data completeness levels at the time of the survey (see Figure 4-3). The results show that there is only a relatively insignificant correlation ($p<0.05$) with the number of patients in the registry. This indicates that centres that enter patient data in more than one research registry generally achieve lower data completeness levels. Despite the p values of 0.08 and 0.11, a trend is seen in the time of data entry and the age of the individual undertaking the data entry. Thus the longer the data entry time and the younger the data entry user, the lower the data completeness levels that were observed. There was no correlation discovered based on t-tests related to the gender, time of data entry (how long after the data was originally collected collection was the data entered in the registry), nor any relation regarding how many trials the centre participates in or whether the source documentation, i.e. the original medical record, was electronic or paper based.

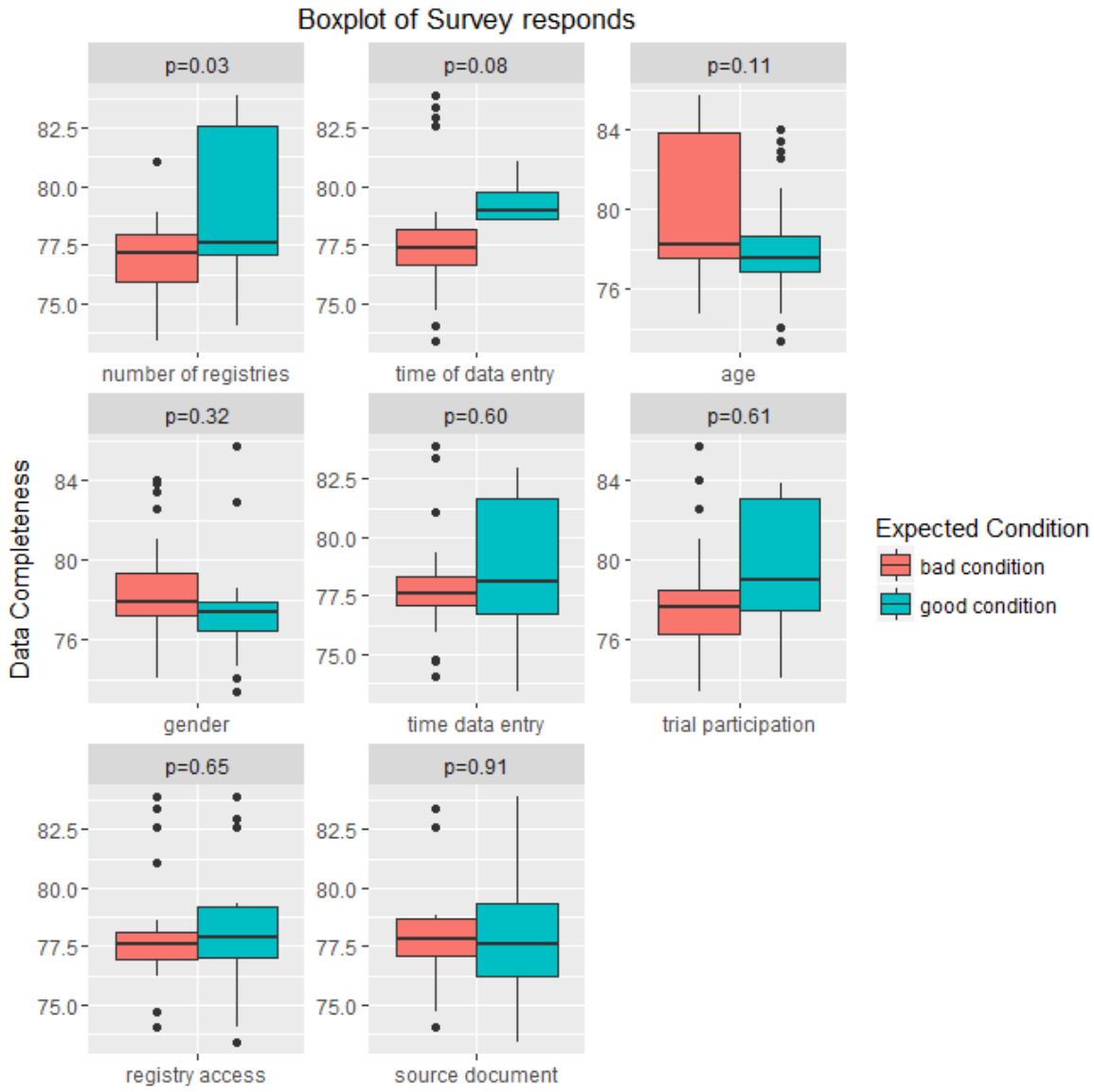


Figure 4-3 Correlation of a range of structural confounders for the ENSAT network

The survey was performed within ENSAT and so the results may not be generally applicable to other studies and trials. Furthermore, whilst 45 participants responded, this was a relatively small samples size. In summary, the results show that there is little or no direct correlation of data completeness and associated environmental factors and related confounders (R^2=0.007). Even in the analysis of individual questions related to data completeness, there was no apparent correlation (R^2 between 0.02 and 0.48). Nevertheless, the survey as a whole provides insights into ENSAT and importantly systematically explores various quality control considerations that need to be incorporated to improve data completeness and data quality more generally in clinical registries.

Assessment of Data Processes

Possible issues related to analysis of confounders for data completion could be recall biases of the participants about the perceived and the actual length of time for data entry. A second bias could be the truthfulness of the response, e.g. whether the participant knows the trials objectives or not may seem to imply a lack of expertise and domain knowledge. During the initial on-site monitoring audit of the PMT-Study that took place in two different centres, three clinicians and two data entry

assistants admitted that they did not know the objectives of PMT and/or EURINE-ACT studies. As discussed in Chapter 3, linkage between the knowledge about trial objectives and the reasons and motivation for entering data was discussed with participants. Insight into the specific trial objectives clarified why the study was being conducted and it is presumed, motivated the data entry user to participate in entry of better quality clinical data. The survey showed that only members of the centres at Munich and Athens identified that they did not know the specific objectives and intended outcomes of the trials in which they were participating. These centres were also participating in several trials and studies including PMT, FIRST-MAPPP, MIBG Impact and Lysosafe, hence it is quite possible that the data entry personnel would not be fully aware of a given specific study needs, demands and the intended outcomes.

The survey provided an overview of individual data processes. Specifically, these included questions related to where data is entered? What source is used for the data entry? How long it takes to enter the data and when the data entered is actually entered into the ENSAT registry? Answer choices where given in a hierarchical order so that if the first criteria did not meet the actual process, the next criteria might match. As an example consider Q1, which focuses on where data entry takes place. Is the data entered either in the same room where the patient is met, or somewhere on the same floor of the research facility or some other more remote location, e.g. at home. If the first possibility does not match, the survey participant should understand the hierarchy of the possible answers and choose the place where the clinical data entry actually takes place.

To present an overview of the data processes that are currently taking place, a graph was developed in accordance to the general data processes described in chapter 3. Case studies showed, that EDC methods are less complete but more accurate to manual data entry (Arts et al., 2002). The survey investigated the manual data transfer from electronic medical records (EMR) and paper-based records to the ENSAT registry. More data entry users transferred information from existing EMR (electronic medical records) compared to paper-based records to the ENSAT registry. Using EMR as the source document achieves a slightly higher data completeness level in ACC records (+1%). The average time for data entry and data capture after the patient visit was calculated based on the mean of the answers. Data entry users that use paper-based records enter that data around 17 days after the original information collection and the data entry user need 5.1 minutes less time for data entry compared to the electronic medical record centres (see Figure 4-4).

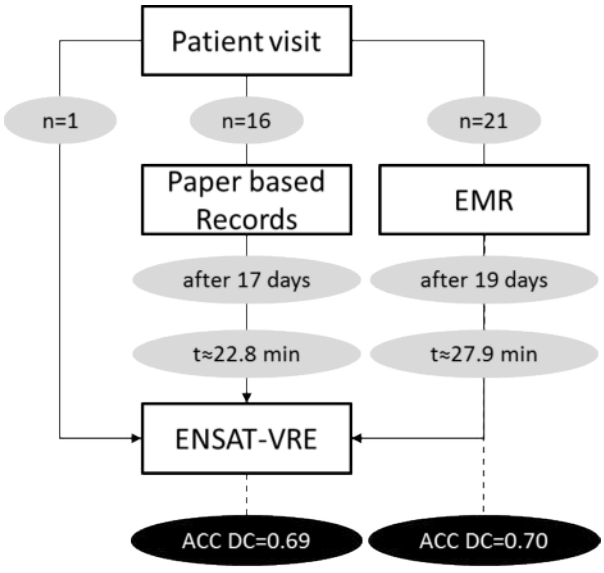


Figure 4-4 Data Process Assessment

Given these results, both methods deliver an approximate equal overall data completeness score for ACC records. This phenomenon is also confirmed by combining ACC and NAPACA records (both sources achieve an overall data completeness level of 74%). From this it can be concluded that the type of source document has no significant influence on data quality based on manual data entry. The use of the average mean for the period of time between the time of information collection and time of the data entry in the ENSAT-registry, allows for a comparison between both types of document sources. Data entry users that have EMR as their source document, enter the data around 2 days later than users with paper-based records. This provides an indication that the availability of paper-based records motivates user for earlier data entry. This is somewhat surprising; however, one possible explanation might be that the users are keen to move from paper-based records as soon as possible.

Open Question – Motivation

In 2012, interviews that took place during on-site monitoring visits for the PMT-Study showed that co-authorship and/or recognition at international ENSAT meetings motivated users to enter better quality data into the registry. Questions related to this topic and the incentives that might be used to improve data quality were thus included in the survey.

Results of the open questions were analysed through qualitative text analysis. The answers were categorised through an iterative process. For this, all 45 answers were divided into three categories. Each category was analysed and further decomposed to reassess and harmonise the categories. As an example, one participant entered: “It’s necessary to improve the research in these rare diseases”. This text entry was first categorised as “progress in rare disease research”. After the decomposition, the text was categorised as “support researcher/science/network in rare diseases” because the statement matches a larger category and a single statement on its own has no significance in the overall text analysis of the survey.

Another example was the statement that “research is necessary for my patients”. First this was categorized as “importance for patients”, but ultimately re-categorised to “improved medical standards and treatments”, because the patient centres view was also shown in improved medical therapies. Some answers also provided multiple motivation factors. A second researcher verified these categories independently.

After the final iteration of the text analysis, ten categories were identified (see Table 4-4). Generally, multicentre research focuses on collaborative research goals, however as can be observed here, the primary focus of the survey participants was related to the patients themselves. Multiple answers identified the need for better care, treatments and standards for patients with adrenal tumours. Other answers focused on knowledge improvements and on international collaboration that can only realistically flourish through collaborative adrenal research. It was identified that international collaborations could have a far deeper impact with co-authorship of research articles for data entry personnel, especially for publications with high impact. However, the term ‘co-authorship’ was not used by all participants. This may be due to the fact that many data entry personnel are not academically motivated or indeed judged by their publication track record with regards to their career progression.

Table 4-4 Motivation for Data Entry based on text analysis (subgrouping)

Category	Count of Answers
Patient focused (improved medical standards and treatments)	11
Knowledge focused (learning about endocrine cancer)	9
Collaboration focused	8
Data mining focused	7
ENSAT focused	6
Support of science	6
Study participation	3
Support young researcher	1

Just a single researcher from Sao Paulo, Brazil, identified the importance of supporting young researchers in their career. Specifically, the researcher motivation for using the ENSAT registry was to build “new avenues for future young investigators”.

The second researcher identified 6 categories and went through 3 iterative cycles of decomposition. Similar categories were defined, while the majority of answers were labelled with the term ‘collaboration’ (see Table 4-5).

Table 4-5 Motivation for Data Entry based on text analysis (grouping)

Category	Count of Answers
Collaboration	18
Treatment	9
Knowledge	6
Profession	5
Research	4
Support	1

Open Question – Recommendations

For many the most important motivation is the collaboration itself, which provides outcomes for patients and for researchers. The advancement in treatments and procedures are also important motivational factors to enter data into the registry.

The last survey questions asked for general recommendations, e.g. how could the registry be improved? In principle, these features reflect the actual needs of the community and what would be required to improve the data quality in the registry. As such they are fundamental to this thesis for improvements to processes and procedures for improved quality of data. A qualitative text analysis was performed but somewhat limited since 20 of the 45 responses entered “no”, or “not yet”. It was noticed that 9 of the answers were given by clinicians, compared to research assistants and principal investigators that have clear expectations what features should be added to improve their own research needs and demands.

All other responses comprised individual answers with different categories. Participants that did enter recommendations typically requested special features like support for web-based comments, easier login or logout, easier data entry including improvements to the user interface and support for web-

based chat. However, there was no general consensus on these recommendations and no categories or community-wide agreed enhancements could be determined.

It should be noted that only one principal investigator (from Munich, Germany), asked explicitly for the implementation of quality control into the registry. Another participant asked for e-mail reminders for compulsory trial related data. Before the survey was sent to the participants, the motivation question was expected to show consensus on the need for data quality assurance. Despite wishes for various new features, improved data quality could not be identified as an explicit demand from the community through the survey. These insights from the community feedback were subsequently used for a more structured assessment in accordance to data quality indicators and especially those that could lead to lower data quality. All answers and a google developed summary of the survey can be found in Appendix II.

4.2 Quantifying Motivation

Research in the field of motivation and data entry focuses on human behaviour in use of computer systems. Key terms that are used in this context include motivation, participation, engagement and contribution. Motivation can often include other criteria like funding of research staff members, authorship, regulations and external barriers in data collection. Tools and feedback processes can directly improve the motivation of the user, but it is recognised that these are not likely to be the full list of external influences. A perfect web-based user interface is unlikely to convince a researcher that has no vested interest in using the system.

Motivation is a key driver of a human that 'moves' them to something. We can make a distinction between intrinsic and extrinsic motivation (see Table 4-6). While intrinsic motivation emerges from the individual's own challenges, enjoyments or interests – extrinsic motivation is given by external influences to the person. The driving factors could be grades, rewards and pressures. Internalisation of motivation has the highest value, because the human can self-regulate and self-motivate for a given activity (Ryan & Deci, 2000).

Table 4-6 Taxonomy of human motivation (adapted from Ryan & Deci, 2000)

Amotivation	No intention to contribution
	External Regulation
Extrinsic Motivation	Introjection
	Identification
	Integration
Intrinsic Motivation	Enjoyment

Intrinsic motivation is the main driver of user behaviour to use an information systems (H. M. Huang & Liaw, 2005). During the data entry process, users need insights into the contribution they provide that are associated with their data entry. A case study discovered that feedback about data contribution led to an increase in the motivation levels of participants in the research objectives (Rashid et al., 2006).

In addition to the factors that can increase user motivation, another approach for the improvement of data entry focuses on dynamic entry forms. The form changes due to a learning process by a probabilistic models and adapts the data entry forms accordingly. This method orders items during data entry and re-asks items if their input looks inaccurate and reformulates or adjusts answers to the plausibility of the entered values. This data-driven data entry approach purports to support an avoidance of data cleaning and reduces the expense of double data entry (K. Chen et al., 2011).

Another important aspect of (de-)motivation is the form of data entry related to the interface. This refers particulate to a clear interface, the ease of use and especially if and how double data entry (DDE) must be performed as part of quality control (Wang & Strong, 1996). In regards to monitoring purposes, DDE is considered to be the gold standard for record validation. However DDE incurs huge time, effort and increased monitoring costs (Day et al., 1998). One way to minimise these costs is to randomly selected records that should be re-entered or values that appear to be incorrect during the data entry (Brandt et al., 2006; K. Chen et al., 2011; Ohmann et al., 2011). Quality tools like tracking systems can also be applied (Tudur Smith et al., 2012). Research has shown that automated form processing with "check boxes, numerical codes and no dates" increases the quality of data (Paulsen et al., 2012).

In summary, motivation is a key factor for data quality. Without motivating the user, data entry and research contributions can be drastically reduced, especially over the time period of a trial or study. It is therefore imperative to keep the community continuously aware about the level of data quality of data that is being entered to assess and increase the motivation of data entry personnel.

4.2.1 Increasing Motivation for Participation

Any given multicentre clinical trial depends on the information that is entered in the trial registry by the researchers and data entry staff. However, a fundamental question is, what drives every person to entry data into a repository? As discussed previously, a key aspect of increasing data entry quality is to understand what motivates users that enter data. Ideally data entry should provide an intrinsic value for the users and clearly show the impact that data entry can have for significant conclusions in biomedical research. Research has shown that a combination of intrinsic values with gamified activities, like rankings, badges or user levels increases the motivation for participation. Examples are Khan Academy, RibbonHero, HealthMonth, EpicWin, Recyclebank, FoldIt and StackOverflow (Deterding, 2012). Even popular resources and applications like Amazon, Ebay, LinkedIn and Duolingo use elements of gamification to increase user interests (Herger, 2014).

As an example, StackOverflow is an online Q&A (Question & Answer) platform for programmers. It has three main criteria: everyone can ask questions, everyone can answer questions and users can vote for the best answers. Every user receives points and badges for special interactions on the site. With this earned 'reputation', the user gets increased access to special actions like editing or deleting posts. Users are placed in a 'league' to compete with each other to increase their overall reputation. A case study investigated the user activity related to earned badges and the results showed "strong empirical evidence […] of gamification in stimulating voluntary participation" (Cavusoglu, 2015).

StackOverflow is an example of an online community that relies on the help of users to increase the quality of software development by programmers. This aspect can be transferred to the biomedical domain, where the research consortium can motivate clinical and biomedical researchers to increase their own quality to achieve a particular trial objective.

Gamification is an approach to apply game elements in non-gaming contexts. It was shown, that the interaction of online users can aid in the development of building enzymes, motivate participation and improve collaboration to achieve a goal (Fernandes et al., 2012; Khatib et al., 2012; McGonigal, 2011; Reeves & Read, 2013). Cloud-based software companies provide services that use aspects of gamification that benefit from application of gaming mechanics on user activities. Established examples include Badeville and Bunchball (Deterding et al., 2011). McGonigal (2011) describe four fundamental elements that a game and gamified applications require: a goal, rules, feedback system and voluntary participation (McGonigal, 2011). Examples of methods that are used to increase the motivation of users in online communities include points, leader boards and league tables,

achievement/badges, levels, stories/themes, clear goals, feedback, rewards, progresses and challenges (Hamari et al., 2014).

The benefits of gamification are not well studied in a biomedical content. It is proposed that three criteria are related to the success of gamified elements: (a) the social environment (planned behaviour), (b) the nature of the system (either utilitarian or hedonic) and (c) the user involvement (cognitive or affective) (Hamari et al., 2014). A case study investigated effects of gamification related to age, sex and usage. It was discovered that gamified tools do not affect age, but women are more aware of gamified mechanics than men. Furthermore, women admire the social interaction of these tools and appreciate this method as more 'playful'. It was shown that the motivation over time fades and users between the age of 20 and 29 show a stronger loss of motivation in using such services than those aged 30 and beyond. This research also considered aspects of networking that showed a positive correlation between usage time and newly acquired contacts (Koivisto & Hamari, 2014).

To summarize, gamification is not a well-studied process in the domain of biomedical informatics. It provides motivation to increase scores of a user in a domain, but this motivation is not permanent. Furthermore, gamification alone cannot increase data quality. The environment must provide suggestions and self-improvement strategies to gain maximum benefit. This in turn requires a creative and dynamic data management system coupled together with intrinsic motivation factors for the participants that are involved.

4.3 Data Quality in the ENSAT Registry and associated Studies

An important aspect of data entry motivation for clinical registries such as ENSAT is whether the researcher is also expected to enter data into clinical trials systems (eCRFs) for general use or to meet specific trial objectives, i.e. is there an expectation of double data entry. Clinical research registries need to be driven by data quality, however they are research oriented. It is important to distinguish between data needed for research purposes and data used for clinical care. Ideally there should be no difference, but this is typically not the case and different steps and processes have to be adopted.

In a research setting, a multidisciplinary board is typically needed to guide the discussion related to which data items can be used for the life cycle of the research initiative. The ENSAT registry has large sets of data with all information needed related to specific tumour types. The registry has also grown to include data sub-sets required for individual trials that are ongoing across the research consortium. Multidisciplinary groups involving experienced researchers, research assistants and software engineer discuss the items that need to be included and excluded in the core data model. Importantly, being involved in a trial can directly impact on the quality of the data in the registry itself (intrinsic motivation). To understand the impact of this, we consider the ACC component of the ENSAT registry and the EURINE-ACT clinical study.

A comparison of ACC and ACC EURINE-ACT records showed that trial data subsets don't have more information entered than the whole registry. We consider here the records from Germany used in the registry, noting that German centers have an especially long history of ACC records and associated expertise in large adrenal research and hence are assumed to have high motivation for entering high data quality. In total 1152 ACC records were assessed in the ENSAT registry, which included 288 records affiliated with EURINE-ACT. In comparison to ACC records, EURINE-ACT has more information entered in three of the seven record types (Biomaterial, Follow up, Mitotane). Biomaterial and Follow up information is a crucial part of clinical research and therefore more information is entered in this record type. Other parts of ACC records that are not affiliated to clinical trials such as EURINE-ACT, have more information in Diagnostic Procedures, Tumor Staging, Surgery and Pathology (see Table 4-7).

Table 4-7 Incomplete items (errors) per record from ENSAT ACC and EURINE-ACT

	ACC	EURINE-ACT	+/-
Diagnostic Procedures	5.10 (1152)	7.10 (288)	-2.00
Tumor Staging	4.10 (1152)	5.22 (288)	-1.12
Biomaterial	167.85 (426)	160.49 (281)	7.36
Surgery	22.16 (1037)	22.29 (260)	-0.13
Pathology	81.47 (1087)	85.89 (260)	-4.42
Follow Up	485.19 (1081)	476.69 (263)	8.50
Mitotane	92.51 (798)	92.38 (222)	0.13

Around 860 ACC non-EURINE-ACT records have more information about pathology and surgery entered. This information was entered without a specific research objective, i.e. it is provided to other researchers to allow them to conduct further trials with this data. Quantifying data quality is also an important tool to show user contribution and the benefit and impact of entering data on behalf of the research community. A key aspect to consider is how best to provide this feedback?

4.4 Visualization of Data Quality Feedback

The research group of Timmermans published an adaption of a centralized data quality monitoring method in clinical research and health registries (Timmermans et al., 2016). Statistical methods for data quality assurance have to be used for centralized data monitoring to aid in the detection of data inconsistencies through the help of statistical methods (Venet et al., 2012). Venet and other collaborators developed a centralised statistical monitoring (CSM) tool for statistical monitoring applied to research trials (SMART). Timmermans et al. (2016) adapted this tool for the evaluation and report of data quality in clinical trials. Although the analysis discovered no inconsistencies that had an impact on the outcome of the trial, the method and the research used was novel. They categorised data errors into four groups:

a) Data inconsistencies related to reporting;
b) Data inconsistencies related to data tendency;
c) Data inconsistencies related to visit-to-visit evolution, and
d) Data inconsistencies related to days (see Figure 4-5).

To report data errors, the authors used bar charts with calculated p-values. Only extreme values were investigated and further quality assurance processes were conducted to investigate the origin of these inconsistencies. An overall score (“Data inconsistency score”) was calculated for every centre to report an overall data quality indicator to local researchers (Timmermans et al., 2016). Using scores and bar charts displaying extreme values should support both community feedback (data quality scores) and investigator feedback (reports to conduct additional investigations to improve overall quality).

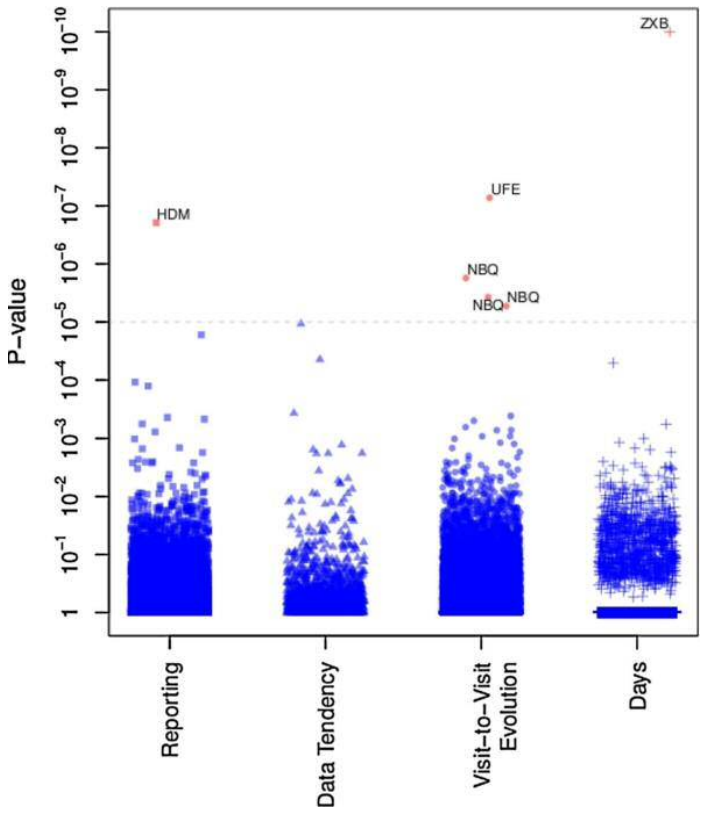


Figure 4-5 SMART p-value calculation with pseudonymised centres (Timmermans et al., 2016)

Puttkammer et al. (2016) performed a qualitative/quantitative case study investigating data quality priorities from health stakeholders. In the quantitative analysis the researchers discovered 13 mandatory data quality indicators for data completeness (n=7), accuracy (n=5) and timeliness (n=1). For every indicator they defined a rule set to support the decision as to whether the items were inconsistent. While completeness covered the primary interests of the stakeholders, e.g., if the item existed, accuracy rules were defined for age, sex, visit dates, medication plans and lab results (range tests). Timeliness was calculated for late data entries (3 or more days after the patient visit). While the analysis results were very specific for developing countries, the researcher concluded that "A dynamic, interactive "DQ dashboard" […] could bring transparency and motivate improvement" (Puttkammer et al., 2016).

Another case study used a survey and a literature review (of articles from 1985-2014) to define "strategies for fraud and misconduct detection" (Knepper et al., 2016) in the area of industrial research. The study was primary conducted to show the performance and resources of TransCelerate, a non-profit organisation providing tools to improve translational research. The survey identified that site monitoring was still the most commonly used method to detect quality issues. Statistical analysis and/or centralized data monitoring were less of a focus of the survey. The authors (who worked for a large pharmaceutical company) discovered that the most used tactic to discover data inconsistencies was to detect extreme values in data trends (outliers) (n=11 of 47 answers), followed by detecting duplicates and confirmations of patient existence (n=8). In the literature review, the author discovered primary motivations (important for the community feedback part of this research), was on "prestige and financial rewards", whereas academic researchers were primarily driven by publishing high impact research articles (Knepper et al., 2016).

With regards to data quality feedback, one group implemented data quality feedback into a food questionnaire-based trial. They distinguished between data input feedback and data analysis feedback. This data quality tool was dynamically adjustable for each user and they could define their

own data quality requirements with assigned weights (Weber et al., 2015). Another group investigated the increase in data quality in a randomised controlled trial. The improvement of health information was aided through motivational short message service (SMS) and by the combination of motivational SMS and data quality feedback. The authors discovered no differences between the two groups, but they noted that the study was underpowered (Joos et al., 2016).

4.4.1 Implementing Data Quality Reports

Considering the breadth of data quality dimensions, a reporting framework should ideally support the calculation and reporting of completeness, accuracy, timeliness, comparability, record eligibility and the calculated overall quality from those scores as they pertain to and support biomedical research (see Table 4-8). All dimensions have different purposes, yet collectively they are fundamental to improving the overall quality of data and importantly, the processes that are used to improve data quality.

Table 4-8 Suggested Data Quality Dimensions related to the Feedback type

Quality Dimension	Real-Time Feedback	Investigator Feedback	Community Feedback
Completeness	X	X	X
Accuracy		X	
Timeliness		X	
Comparability	X	X	
Record Eligibility		X	X
Overall Data Quality		X	X

Reporting Feedback

Generic data quality tools should consist of several feedback considerations and support a range of features. Specifically, we identify and consider support for data quality considerations of the user, the principal and local investigator(s) and the research community (see Figure 4-6). *Real-time feedback* can be applied during data entry and is suitable for the data entry user to allow them to gain insights to the actual quality of the entered data. *Investigator feedback* can include items that are generally reported after monitoring visits and that are possible to report upon in comparison to previously entered data. The *community feedback* considers leaderboards of users and their associated centres in relation to their entered data and the achieved data quality. These tools should give advice related to items and processes that could or should be improved. This can be reported in several ways: it can be reported to the user via email or during the login as a daily task list; it can be regularly presented to the investigator to obtain an overview of items with low data quality, so that processes to improve the items in general or for specific trials can be initiated. The approaches can also benefit from reminders and notifications of progress in data quality as a whole to ensure that where benefits have been observed (increases in data completeness) then the users are actively encouraged.

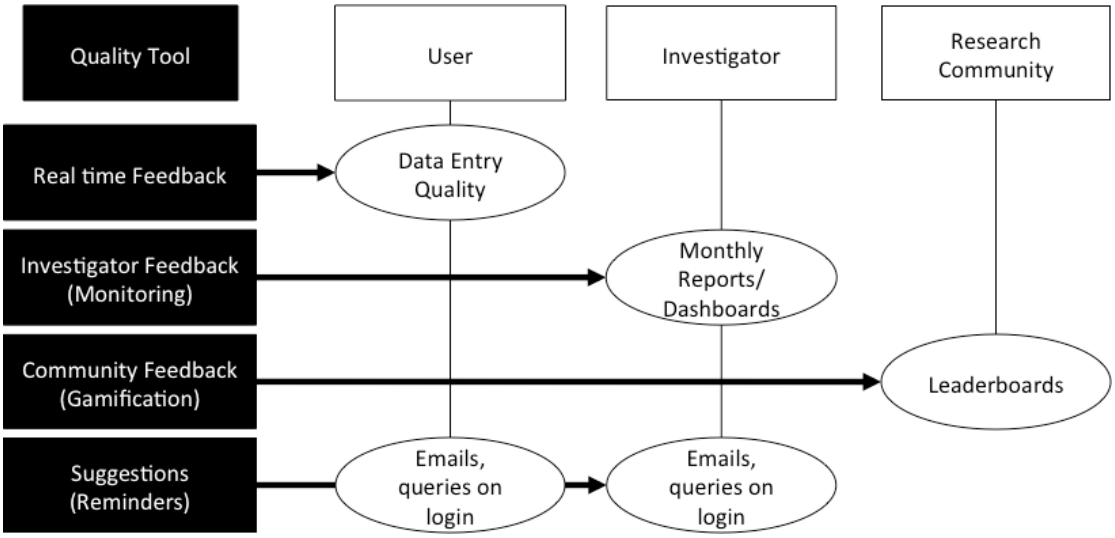


Figure 4-6 General Data Quality Tool Architecture

Real-time Feedback

Identifying errors or data quality issues in real-time data entry implies that every data field must have precisely defined units and ranges. Real-time feedback should provide completeness and quality values related to the form/record that is created and/or updated. This information can be augmented with average quality values of the user and/or the community. A case study showed that real-time feedback doesn't improve data quality immediately, but improvement was only seen in combination with reminders and ongoing training (Nair et al., 2010). In another study, real-time feedback was described as a "check for data integrity", where collection inconsistencies, data errors and "illogical entries" were detected during data entry (Shiloach et al., 2010).

In summary, real-time feedback must include:

- Current value of quality;
- Comparison of quality with other users;
- Errors in data fields, and
- Reminders of missing compulsory data.

All these features must be combined with clear data field definitions. Real-time feedback influences the user interface presented to data entry personnel. During data entry the actual data completeness can be reported in real time with reminders and messages related to compulsory data not being entered and/or statistical feedback regarding the completion of the data that has been entered. A patient calendar can also serve as a reminder that follow-up, examination or tests are expected and/or, have potentially been missed. Consortium-wide networking and trial/consortium related news feeds can also be implemented to encourage users to regularly log in to the database and enter data.

Investigator Feedback (Monitoring/Audit Platform)

Investigator feedback can include quality reports that are periodically sent to research centres with the aim to inform them about their achieved data completion and accuracy. These reports can include information about record eligibility, patient follow-up times, specific information about treatment (dose, time and interval) all of which can represent the main objectives of the registry or trial. Every

category can be reported with a score and the data threshold that must be achieved (Sandman et al., 2006).

Investigators receive feedback after auditing or monitoring events. However clinical on-site monitoring visits can cost up to 30 percent of the total cost of a clinical trial (Davis, Nolan, Woodcock, & Estabrook, 1999). Therefore, Investigator Feedback should ideally include:

- Record eligibility for the trial;
- Overview of (Serious-) Adverse Events (SAE/AE);
- Current status of trial objectives and associated endpoints, and
- Information about staff efficiency (data entry/time/data quality).

Investigator feedback consists of information about data quality that is provided to the principal and/or local investigator associated with a particular centre. To reduce costs this report may consist of results related to possible investigations that would normally be carried out during an on-site monitoring visit. Through such reports, the investigator can determine how best to improve local data processes. Little research has been done in this area, but investigator feedback is critical for the reduction of costly on-site monitoring activities. This feedback needs to display critical data and processes that lead to low data quality. Such reports should:

(a) Assess and display the eligibility of all trial marked records. This eligibility can be calculated by a range of inclusion and exclusion criteria as well as through investigator defined mandatory items. Records that do not meet the minimum trial requirements should ideally be provided in a list with the information that needs to be entered (Suggestion list). These values will be compared with the community.
(b) Calculate the recruitment ratio and the aim of required subjects for the study;
(c) Provide an overall comparison of achieved data to study endpoints (if applicable);
(d) Display inconsistencies of demographic data and other data, e.g. laboratory results of new patients and compared them to other centres;
(e) Check if every record was marked with patient signed informed consent;
(f) Provide statistics for staff members (activity, quality, quality trends) for training purposes;
(g) List any adverse events in the trial (all centres);
(h) List any missed follow-up events, and
(i) Performance differences between the sites.

Those reports should be provided to the investigator on a regular basis. This can be achieved through several mechanisms, e.g. through a website, via the registry or simply via email. The information requires that detailed logging information and associated metadata are collected, e.g. showing overall statistics of the amount of time that the data entry personnel spent entering data. These suggestions can include FDA/GCP standards and requirements as well as recommendations from the literature. Instead of using only lists or tables, graphs and charts can also be used. For example, the report of performance differences may include a RICC vs. R^2 plot. By clicking on the specific item a box-plot appear or table may appear that reports side-by-side differences (see Figure 4-7 & Figure 4-8).

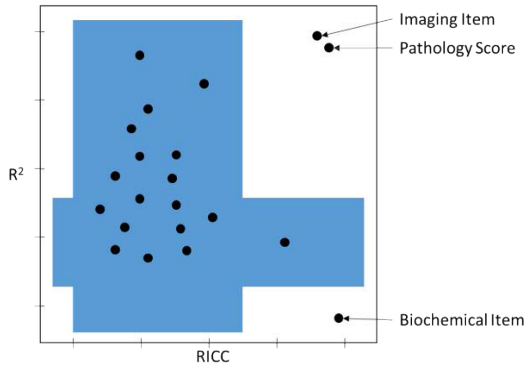


Figure 4-7 Example report of item related performance differences (adapted from Wynants et al., 2013)

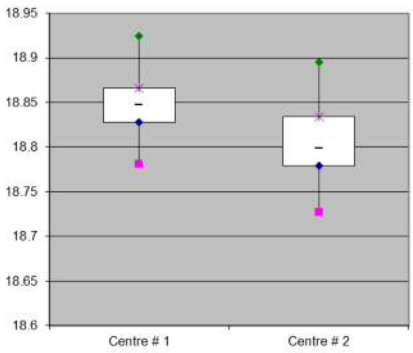


Figure 4-8 Example report of side-by-side comparison of an item (adapted from Guthrie et al. 2012)

Community Feedback (Motivation/Competition Platform)

Reports to the whole research community aim to update all researchers regarding accomplished data quality results, trends and issues that appear during the course of a trial or establishment of a registry. Reports should ensure that the goals of the work as a whole are transparent to all participants. Reports should also empower users to increase their knowledge about the study and the problems that may arise. These are common aspects of transformational leadership that "inspire followers to change expectations, perceptions, and motivations to work towards common goals" (Burns, 2003).

Motivation is key in this type of feedback. For the collection of high data quality, as noted, motivated users are the best way to increase the efficacy of the clinical databases. Therefore, it is highly recommended to use aspects of motivation related to user interfaces and reports. Staff members need to have an insight of the trial, including its overarching objectives as well as ways to understand and tackle their problems. Community feedback must therefore:

- Empower every research/user to understand the collaboration objectives including potential record eligibility for clinical studies and related issues;
- Motivate users to increase their own and/or their centre performance, and
- Motivate users to increase participation and improve the overall quality of data – even if the data that is to be entered does not directly benefit them.

Community feedback can also be construed as 'naming and shaming'. In a gaming context, users compete with others. The possibility of a research assistant to appear on a leaderboard in the first position and to be recognized by the research community could be an intrinsic motivation factor that encourages them to participate in a registry and enter high data quality including increased user activity. Every user who is responsible for data entry in an international research consortium needs to contribute with the highest possible quality of data. To achieve this, gamified applications can be supported in the web-based research environment. Four main elements of such a system need to be considered in the biomedical domain (see Table 4-9).

Table 4-9 Elements of Community Feedback in Biomedical Research

Goal	Support biomedical research through high quality datasets;
Rules	Enter data of the highest quality and sustain this over time;
Feedback System	During data entry, statistics should be available with associated encouragement and/or benchmarking with other personnel/centres, and
Voluntary Participation	The user should be able to decide whether they accept the challenge to compete with others or not.

The primary goal of this feedback is to assess why and how data quality is actually measured. These aims are to improve research using quantifying errors of datasets (assessment); to design actions to reduce errors (plan); to improve the quality of the dataset (do); to evaluate the inconsistencies (check) and ultimately to achieve a higher overall data quality (act). This is an iterative process related to the PDCA quality cycle that can be used to fulfil quality control aspects (see Figure 4-9) (Deming, 2000).

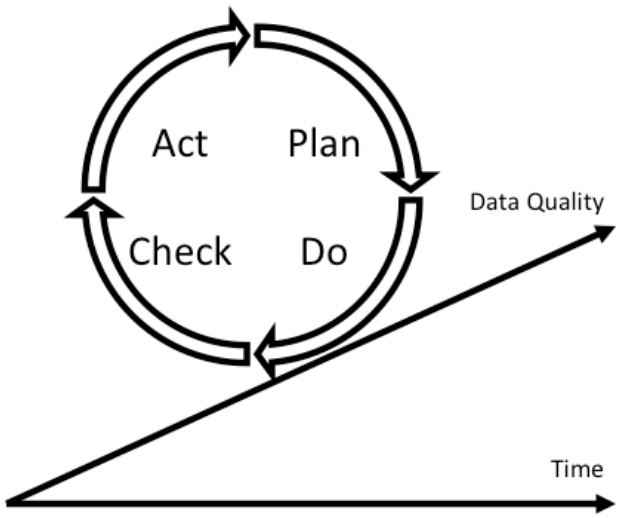


Figure 4-9 PDCA Continuous Improvement Cycle (adapted from Deming, 2000)

Rules are defined that allow the user to "unleash creativity and foster strategic thinking" (McGonigal, 2011). This creativity emerges out of the will to achieve a better place on a scoreboard or to receive a higher user level in the research environment. It was noted in chapter 3 that some users simply cannot achieve high data quality. This can be due to a variety of factors: funding issues; local processes that lead to incomplete clinical information, e.g. lack of imaging facilities or lack of patients at that centre. As such, such factors should be considered when scoring data quality. These scores will be activity and quality improvements or rankings where troubling items (e.g. pathology and imaging information) can be excluded from the user competitive rank.

To consider the design of community leaderboards, it has been identified that six aspects should be considered: centred views, filtering, aggregation, multiple leaderboards, points and metadata (Ventrice, 2013). Reviewing user statistics can be embarrassing and confusing. Therefore, users have to see their scores in the centre of the interface or at the bottom of the screen so that an immediate comparison with other users and centres is possible. This provides the opportunity to establish how much of an improvement is needed to overtake other researchers in the leaderboard for example. Ventrice (2013) also suggests multiple scoreboards, related to teams (centres) scores and individual scores (see Figure 4-10).

Place	User	Centre	Score
1	User #3	Centre A	.98
2	User #2 (you)	Centre A	.97
3	User #4	Centre C	.81
4	User #1	Centre B	.79

	User	Centre	Trials

Figure 4-10 Example of a leaderboard that covers a simplified view of the user and the scores in different boards (individual, team and community)

For voluntary participation it is important to establish the motivational affordances for particular studies and trials. This motivation could be co-authorship on papers; involvement in grants; clinical decision improvement and/or recognition as part of the job description that could have an impact on the individual's career. A further incentive is community recognition, e.g. through rewards/prizes honouring the contribution made at collaborative meetings.

4.5 Continuous Data Quality Assessment

As described previously, European funding schemes have accelerated clinical research. ENSAT has benefited greatly to other adrenal research initiatives worldwide through its collaborative funding model. With collaborative funding, challenges of clinical data quality arise. To identify as early as possible any low data quality issues, primary indicators for data inconsistencies must be identified. A range of commonly accepted quality indicators was developed by Avedis Donabedian in health care (Donabedian, 1988). These include structure, process and outcome quality.

To map a range of possible causes for low data quality, a cause-effect diagram (Ishikawa diagram) for data processed within EURINE-ACT was developed. The starting point for the analysis is the information delivered by the patient to the individual who captures the data for research purposes either by pen and paper or electronically. Primary data quality issues can stem from several areas: environmental issues, by human, by hardware or data management software.

It was discovered that the main issues are often the human resource (see Figure 4-11). Other common problems are time and funding as well as low motivation to enter medical related information into clinical databases outside of the immediate healthcare setting.

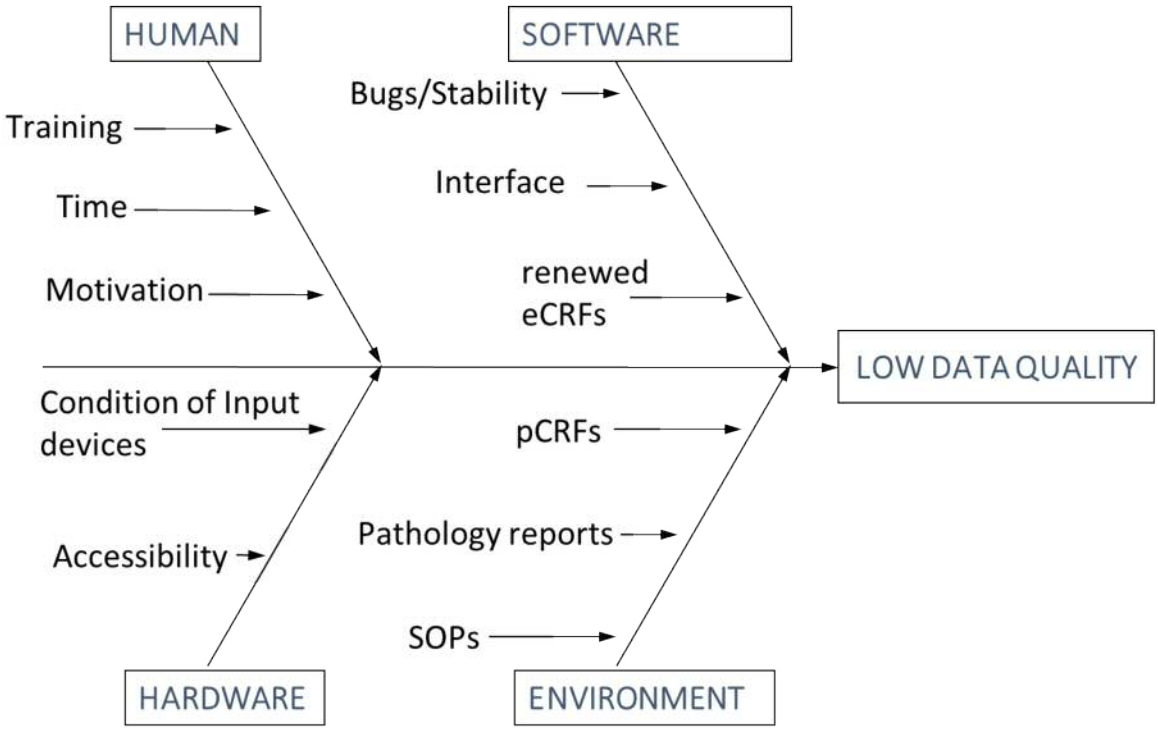


Figure 4-11 Cause-Effect Diagram for Causes of Low Data Quality (own figure)

This research is conducted using qualitative and quantitative research methods. The central concepts of the method are based on the PDCA Cycle (Plan-Do-Check-Act) or Deming-Cycle, which is well established in health care quality assurance (Deming, 2000; Taylor et al., 2014). This process is used as a quality control method and supports continuous improvement of the tools and processes that should be adapted (see Figure 4-12). Such a concept is often mandatory due to the heterogeneous processes that are adopted during translational research (American College of Medical Quality, 2010).

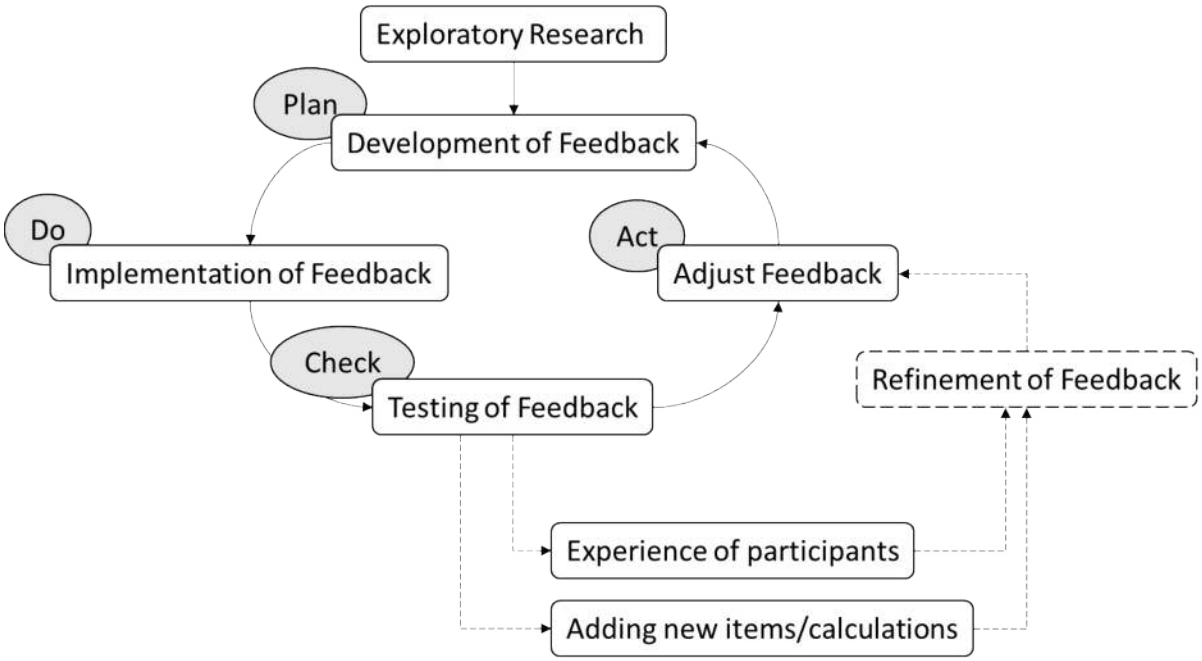


Figure 4-12 Quality Assurance Cycle

To explore possibilities for improvement of data quality in the biomedical and translational domain, a detailed literature review was performed in the initial phase of this research. The general search was performed using Google Scholar and Medline Database with the words “evaluation of data quality” and “medicine” between the years 2000 to 2013. The year 2000 was used because translational medicine was first identified in the literature in this year (Sung et al., 2003). During the review process, leading questions were:

- What methods are used to improve data quality in the translational domain?
- What requirements need data management systems for quality assurance purposes especially in clinical trials and studies?
- What dimensions of data quality are currently discovered and used?
- Are there any published documents that elaborate on the effectiveness of data quality feedback in academic research?
- How can researchers be motivated to enter higher quality data in an external research data repository?

To supplement the search, papers referenced by other papers and authors known by reputation were considered and included in this review.

To illustrate the needs for data quality and the importance for feedback, a case study was performed. The EURINE-ACT trial within the ENSAT Consortium was used to gain insight into translational research and to provide feedback about the requirements on research data sets and data quality for principal and local investigators. Initially all compulsory data was used to calculate a data completeness value. This value was reported to centres. A range of rules where applied to the calculation, including plausibility checks and a novel eligibility confirmation check for adrenal tumours. With these values an overall Data Quality Score (DQS) was calculated and reported to every centre.

After the first reports, feedback from every site was analysed and the DQS adjusted accordingly. Each month, new mandatory items were added to the completeness score to encourage every site to enter higher quality data. In addition to this process, on-site monitoring visits for six selected centres was conducted.

In the trial, two of the proposed feedbacks were introduced. *Community feedback* aimed at encouraging the study site to improve their captured data and to be *better* than other sites with regards to data quality. *Investigator feedback* should provide an overview of captured values of the centre and allow a comparison with other centres to determine, if values are comparable or if inconsistencies exist. The effectiveness of both feedback mechanisms were analysed to evaluate difficulties and possible improvements (see chapter 3.1.4).

The overall data quality (data quality score) calculated for EURINE-ACT is a specific quality score for adrenal cancers. It is mandatory to show the usability and effectiveness of quality scores across other biomedical communities. The PDCA method is an iterative process and the cycle restarts and the experience gained in EURINE-ACT can be applied to other studies and trials. Here, all feedback types were analysed, including whether it was possible for user motivation to enter high quality data to be increased.

4.5.1 Proposed Structure of Quality-driven Research Registries

Research platforms in biomedical research typically don’t need long-term motivation since studies run over a given time window, e.g. a few years based on the research funding. Furthermore, since data collection already happens during routine care, the ongoing motivation to enter data into research databases is important. It could be argued that data entry should not be for enjoyment and simply

part of the formal job description. In many (most) cases however, research funding is inadequate for full time data entry personnel.

While reporting feedback, it is important to keep users motivated. Therefore, it is advisable to implement feedback features in stages. There are two different scenarios of data quality feedback requests: either at the onset of a new study/trial or during the conduct of a particular study/trial. Data feedback at the end of a trial is not considered to be feedback. Rather it is more a retrospective evaluation of data quality.

Continuous data quality assessment during an ongoing trial is essential. A researcher or an EDC method enters/feeds data into the study database ideally with data quality tools giving immediate feedback to the researcher with reports that can be generated for the community and investigator (see Figure 4-13).

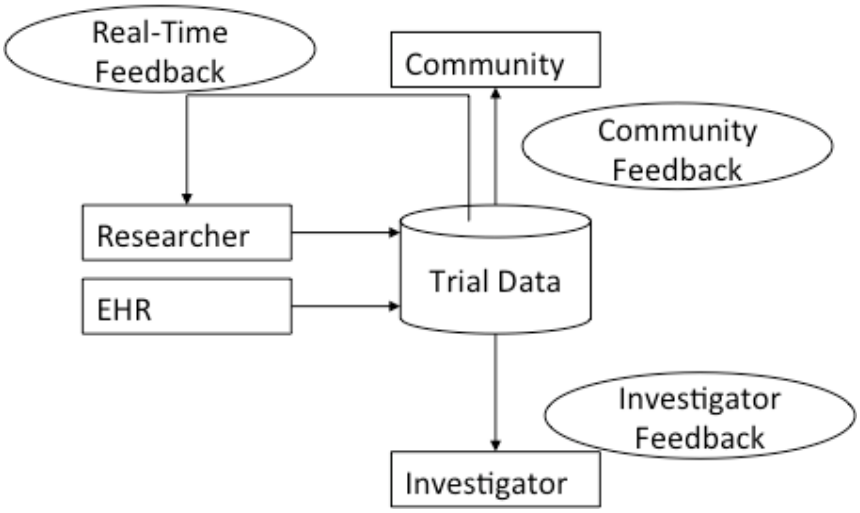


Figure 4-13 Quality-Driven Data Quality Feedback during a trial

Another purpose of data quality reports is to support the onset of new trials and studies, where the investigator defines the study objectives and compulsory and optional data items. This information is entered into the clinical database and eligibility scores should be identified and subsequently captured. This includes how many records fit the study and what records need to be updated with information to be confirmed as eligible for the trial (see Figure 4-14).

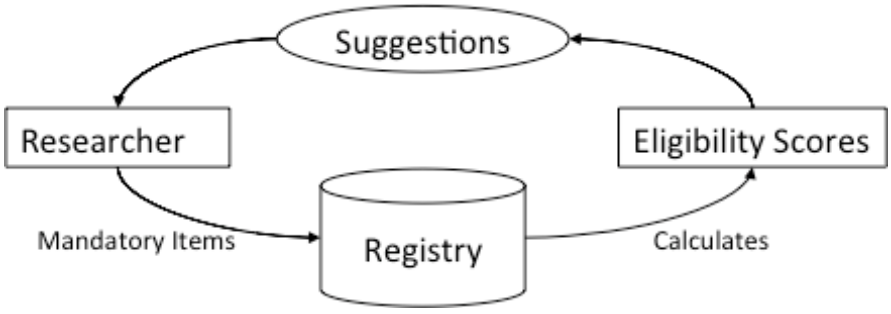


Figure 4-14 Quality-Driven Data at the onset of a new trial

4.5.2 Development of Automated Data Quality Feedback

A data quality-reporting framework is essential for exploring and analysing the results of quality control methods and quality dimensions that need to assessed and adjusted with continuous user feedback to optimally motivate data entry personnel.

In this situation, the trial is initiated by a principal investigator responsible for defining compulsory and optional data items that need to be collected during the trial (as part of the clinical trial protocol). During the set up phase of a clinical research trial, queries should be run to ascertain the quality of the data in the registry. This quality should encompass how many mandatory data items are required (according to the clinical protocol) and how many new items are required (in the eCRFs to be developed for the trial).

As part of this process, the data manager performs the evaluation of data quality in accordance to centralized monitoring guidelines. As discussed, general accepted quality dimensions are data completeness and data accuracy (validity) of data in the registry. Additionally, as shown in EURINE-ACT, a record eligibility score can be calculated to show the investigator, how many records can be confirmed as eligible for the trial. The next step is the selection of reporting methods, i.e. how the data manager reports the data quality to the data entry personnel, principal and local investigators and the research community.

There are several tools and features mentioned in the previous chapters that are useful for these steps. Data entry users can receive real-time feedback during data entry, as well as possible query lists with suggestions regarding what data items are missing in records to support their eligibility for a trial. The investigator receives regularly quality reports in the form of a dashboard page of the registry and/or clinical study with an overview of trial performance and if the primary items are collected in accordance to clinical trial SOPs (if applicable).

Finally, the research community can be set into a competition with one another, where individuals and centres gain rewards by obtaining high quality values or entering many patients over defined time periods. These extrinsic motivational factors can encourage data entry users to be better than others. At the end of the iterative cycle, the data manager will assess if the data quality has indeed improved. If the data quality improves the feedback works, if the data quality does not improve, the loop starts again with a discussion of data items in an expert group between the principal investigator and data manager (see Figure 4-15).

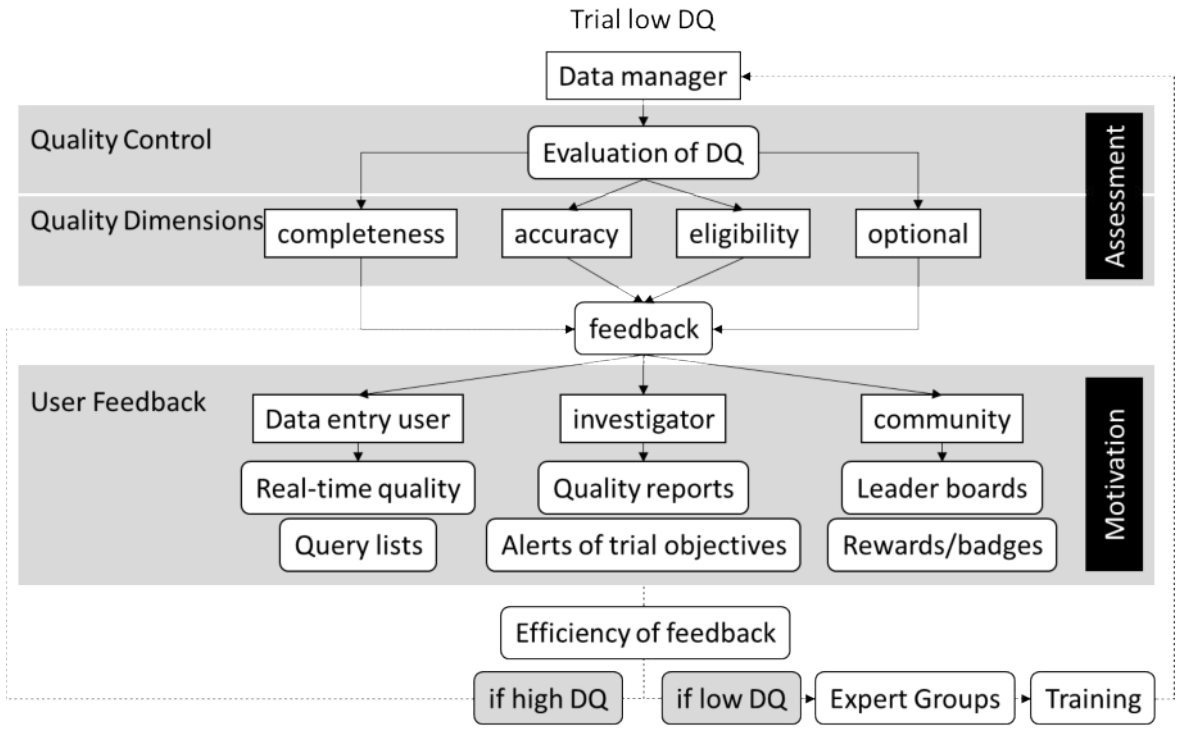


Figure 4-15 Automated data quality feedback loop

This feedback loop will be supported with notifications that can be delivered via a range of technologies. This can be email notifications of missing compulsory items or when trial milestones are reached for example.

One other dimension underpinning this work is the notion of trust and honesty. When awards are made through ranking and leaderboards, it is possible that end users may be pushed to enter the highest quality data on the registry and where the data itself is manufactured. In this thesis and work as a whole, we treat data entry personnel as honest citizens that work within organisational frameworks that such falsifying of data would result in severe reprimands. Nevertheless, encouraging data quantity and quality should be directly augmented with data auditing capabilities to ensure that such issues do not arise.

4.6 Chapter Conclusions

The previous chapter has showed a practical and theoretical approach for systematic evaluation and improvement of data quality. It identified that reporting data quality and data completeness are an ongoing process that should ideally motivate the involved personnel. While PMT and EURINE-ACT showed an improvement of data quality through investigator and community feedback, this does not indicate that data feedback alone is the cause of that trend. There can also be performance bias and hence monitoring visits should not be replaced completely. Rather this approach can be an aid in the reduction of monitoring visits to improve data quality.

It is also known that for the continuous improvement of data quality, ongoing training of data entry personnel should be mandatory. To clearly demonstrate the impact of feedback, it is necessary to investigate whether there is indeed a positive effect of feedback on data quality. This can be achieved through observational studies. The following chapter provides a study design to investigate whether automated feedback can indeed improve data quality. For this a stage-wise research framework will be developed, that has to be implemented into the data management system. This framework has to assess intrinsic motivation (self-motivation) and extrinsic motivation (pressure from outside). Feedback can provide feedback about the individual performance and a comparison to other users. Both intrinsic and extrinsic motivational factors can increase the quality of data entry. A comparison of data quality before and after the implementation of feedback will detect whether data quality feedback works. This aspect is explored in a different disease research registry and explained in the following chapter.

5 REALISATION OF WEB-BASED AUTOMATED FEEDBACK

This chapter applies the framework of the previous chapter in the context of data quality demanding disease registries. With this case study we aim to assess if automated (web-based) feedback during data entry actually improves the overall quality of the data that is entered into clinical data registries or not. We also consider the use of the systems more generally and whether the adoption and use of the systems encourages an uptake of data quality. Specifically, we track patterns of user behaviour through web-based registries and discuss the utilisation patterns and whether there are any indicators that might increase the motivation of end users to enter more data and/or data of higher quality.

5.1 Case Study on the International Niemann-Pick Disease Registry

5.1.1 Niemann-Pick Context

Niemann-Pick Disease (NPD) is a rare metabolic disease that is caused by an accumulation of fat cells in the liver, spleen, bone marrow and brain. NPD is categorised into three types (A, B, C). There are at present no specific cures for this disease, just symptoms that can be partially treated (Schuchman, 2007). Researchers from six centres (Prague, Barcelona, Udine, Birmingham Children's Centre, University of Birmingham and Dublin) collect disease specific information as part of the EU funded International Niemann-Pick Disease Registry (INPDR[9]) to gain a better insight into the basic biology and clinical management of the disease. The registry captures extensive data about clinical symptoms and results of diagnostic tests.

The data that is captured can have different processes used for data capture. The INPDR supports manual data entry by clinicians and associated researchers – along the lines of ENSAT, however INPDR also captures data from patients directly through web-based systems and mobile applications. Mobile apps for INPDR patients used for capturing data for diagnostic purposes was presented in Sinnott et al. (2015). The registry captures a range of items clustered into several categories: demographic data, biochemical results, family history, clinical history, treatments, disability scale, further investigations and completed visits. While "further investigations" and "completed visits" are non-mandatory, such information supports the clinicians and research community by providing insight into the progression of the disease. Visits support longitudinal data covering laboratory results, treatment plans and disability scales. At the beginning of the evaluation and feedback period, INDPR included 142 NPD Type C (NPC) patient records and 18 NPD Type B (NPB) patient records. This work thus focused primarily on NPC patient data only.

To commence the work a baseline analysis of data was conducted. It was identified that the centres with the highest baseline completeness data were ITUD2 and CZPR1. However, the 21% completion rate of ESBA needs to be viewed in relation to their contribution of more than 100 patients into INPDR. Taking a closer look, the lowest scores are shown in treatment along with the visit information. Excluding ESBA, all scores would be 7-10% better (see Table 5-1). ESBA has far more patients entered however it has a lower overall data completeness level.

[9] www.inpdr.org

Table 5-1 Baseline data completeness NPC (1st January 2016)

Centre	patients	Lab Results	Family History	Clinical History	Treatments	Disability Scale	Add. Invest.	Visit	total
ITUD2	2	0.90	1.00	0.96	0.67	1.00	1.00	0.00	**0.82**
CZPR1	8	1.00	1.00	0.96	0.54	1.00	1.00	0.00	**0.81**
IEDU1	3	0.67	0.89	0.71	0.11	0.90	1.00	0.00	**0.66**
UKBI1	3	0.67	0.67	0.64	0.00	0.67	0.67	0.58	**0.61**
UKBI2	23	0.93	0.91	0.89	0.13	0.17	0.15	0.06	**0.53**
Total	**142**	**0.73**	**0.79**	**0.73**	**0.26**	**0.66**	**0.68**	**0.13**	**0.57**

5.1.2 INPDR Study Design

To explore the benefits of real-time data quality feedback and community data quality feedback a staged observational study was developed to assess data completeness trends and the impact of data quality feedback categories (see Table 5-2) of the INPDR.

Table 5-2 Study Design - 2 Stages of DC Feedback

- Stage I - No feedback (data collection)
- Stage II - Data Completeness feedback to the user
- Stage III - Stage II feedback with users average Data Completeness
- Stage IV - Stage III with community average Data Completeness (see Figure 5-1)

Figure 5-1 Stage IV Data Completeness Feedback

During data entry, every user receives warnings and alerts if the data entered is implausible (see Figure 5-2) or compulsory data is missing (see Figure 5-3). Within INPDR, the basic demographic patient information is captured when a patient record is initially created and this is all compulsory data. This includes a data element confirming that patient consent has been obtained. The actual compulsory (demographic) data is a very small subset of the total data that is to be captured within the INPDR however.

Figure 5-2 Warnings in INPDR

Figure 5-3 Missing compulsory data

This staged approach is in accordance to the motivation scale described in chapter 4.2. The human is driven ('motivated') by intrinsic and extrinsic factors. While intrinsic factors rely on self-motivation, extrinsic factors are based on pressure from outside with expectations that have to be fulfilled. Stage II & III assess whether the user is primarily driven by intrinsic factors. Stage IV assesses if the user is driven by comparison with colleagues. As noted, this could also be construed as 'name and shame'.

Every stage needs a separate assessment. Ideally, each stage should go live after a fixed time window, e.g. 2 months, or after a given number of patients were entered, e.g. at least 50 patients were entered into the INPDR. Unfortunately, between January and August 2016 only 35 records from 7 different centres were entered into the registry. This amount of patient information is not useful to draw significant conclusions. A power calculation indicated a minimal recruitment of 358 patients within 6 months. However, an analysis of data completeness scores can identify trends of quality and whether real-time feedback can reduce data inconsistencies or motivate the data entry user to enter better data.

Due to the low recruitment rate, Stage III went live after Stage II was live for 2 weeks. After both stages were combined in the final analysis. Data completeness of Stage II & III was analysed after 3 months and data completeness of Stage IV analysed after 3 months of Stage II & III (see Figure 5-4).

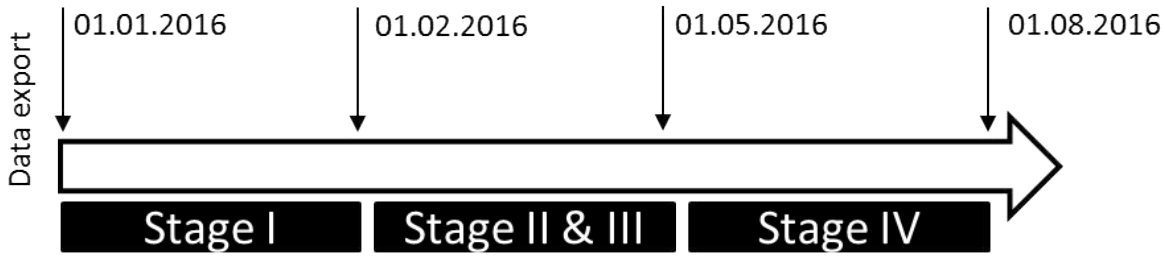


Figure 5-4 Time of data export and analysis from INPDR

The results showed a general increase of data completeness in five of six centres. CZPR1 entered just one record within these seven months (before stage II), which was more complete than the average records. ESBA1 entered five records and slightly increased the data completeness, but this was caused by the entered record information, not by the data quality feedback. IEUD1 showed a significant increase of data completeness before Stage II by improving existing records. ITUD2 had a dramatic fall of data completeness, caused by entering more incomplete records into INPDR. UKBI1 reduced data completeness by around 6% between stage I and II&III, by entering one incomplete record. UKBI3 entered no records and updated one patient record in stage IV (see

Table 5-3 where the number in brackets is the total number of patients entered into the registry at this time).

Table 5-3 Data Completeness Trend - Stage I - IV of the INPDR

	1-Jan-16		1-Feb-16 Stage I		1-May-16 Stage II & III		1-Aug-16 Stage IV	
CZPR1	81.23	(8)	81.40	(9)	81.40	(9)	81.40	(9)
ESBA1	30.93	(103)	31.37	(106)	31.35	(107)	31.89	(108)
IEDU1	66.00	(3)	72.67	(3)	72.67	(3)	73.17	(3)
ITUD2	81.56	(2)	38.51	(17)	38.51	(17)	38.51	(17)
UKBI1	61.08	(3)	61.08	(3)	55.31	(4)	68.10	(6)
UKBI3	53.17	(23)	53.17	(23)	53.52	(23)	53.61	(23)

INPDR Conclusions

The results of analysis of INPDR showed an increase of data completeness of existing records, however this needs a much larger sample size. To gain further insights into the DQF and a user's intrinsic/extrinsic motivational factors, the study needs to be extended with more patients and patient related information.

5.1.3 INPDR Discussion

This INPDR study design is a classical quasi-experimental study with a one-group pre-post test design. There is no actual control group because the data quality was measured from the same group before feedback was implemented into the data management system, i.e. there was a large collection of data entered into the INPDR before the feedback mechanisms were implemented. However, this study was performed to exclude a possible confounders related to data manager/data entry personnel that want to increase the data quality without using data quality feedback. Such confounders have the characteristics that they are directly associated with the outcome (higher data quality) and the exposition (data quality feedback). Monitoring increases the feedback rates that are reported to the data entry users and monitoring directly encourages the users to enter more data (see Figure 5-5).

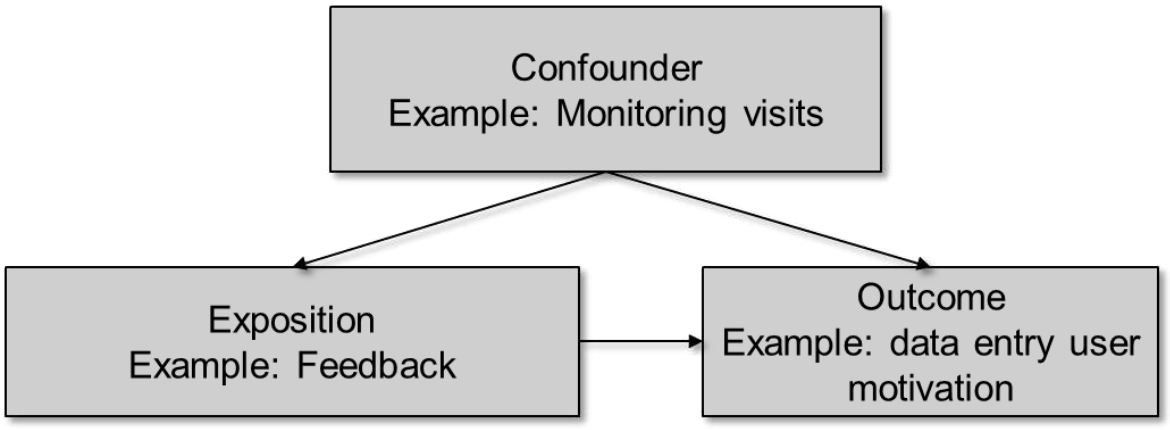


Figure 5-5 Confounder: Monitoring visits

Confounders can be detected through various mechanisms. One example of a confounder in medical cohort studies is the age (high age influences an exposition and the outcome) of the subject or the gender (Rothman, Greenland, & Lash, 2008). Research in medical informatics can be biased by, ambiguous temporal precedence, selection of wrong conditions, events that happen concurrently

during the observation, maturation, regression of scores due to multiple interventions (features), attrition, testing, instrumentation and interactive effects (Shadish, Cook, & Campbell, 2002).

In such case studies, there are two ways to exclude confounders in the final analysis. Cohorts like the INPDR research community need to be stratified in two groups: one group that had monitoring visits, and one group that had no monitoring visits. A comparison of both groups could subsequently detect and exclude the confounder. A second possibility is a restriction, e.g. considering just users without monitoring in the final analysis. Other possibilities that deal with confounders that could not be applied in this particular case study are randomisation (one group with feedback and one group without) or matching of cases with controls (case-control study) (Rothman et al., 2008; Rothman & Monson, 1973).

This case study could also be influenced by events that impact on data completeness. One expected event is the entry of imaging, pathology or laboratory results. For example, such capabilities and hence data are not available due to local clinical processes, or where the results can only be collected and entered after a certain amount of time. Such localised processes can have a major impact on the evaluation and data quality and completeness, especially where potentially minor improvements through web-based feedback are *lost in the noise* of major local organisational data access and usage issues.

5.2 Tracking and Monitoring User Online Data Entry Behaviour

To understand the data entry habits and impact of data quality feedback, it is also important to understand how often users are accessing the system; what records they are editing and how long they stay online for. This can be achieved through web-based logging tools. It is obvious that feedback only works if the systems are actually being used.

Logging is not a new feature. Companies with medium and huge traffic on their webpages typically try to improve interfaces and design their application with support of logging and/or metadata information. One popular public user logging example is Google ‘My Activity’. In case the same Google account is registered on multiple devices (phone, PC, laptop, car navigation system), this approach offers a very detailed log of the use of various systems by a given user (see Figure 5-6).

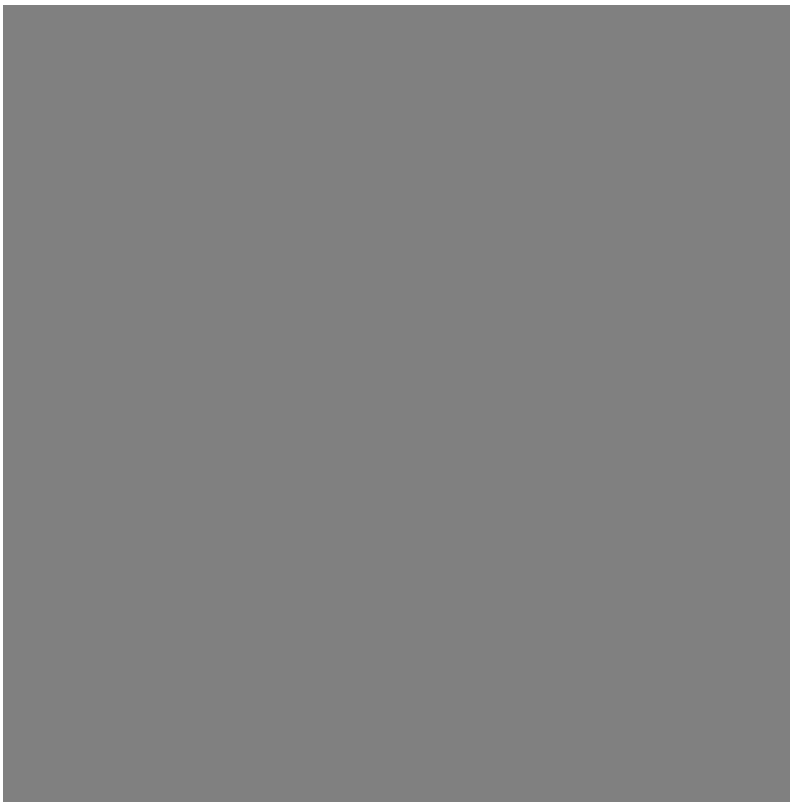


Figure 5-6 Example Google Logging

The literature describes the use of logging information in regards to the quality dimension of timeliness, which is also often referred as currency. In the clinical context, logging tools can track whether patient information is entered into the registry within a reasonable time period after the original data collection (Weiskopf & Weng, 2013). As a biomedical example, the translational research platform REDCap, a metadata-driven research registry, uses a 'Data Logging Module' to show every user data entry transaction and to provide a platform for data auditing during the lifecycle of the usage (Harris et al., 2009). ECRIN (described in previous chapters) provides a research network offering standards and frameworks for biomedical research in Europe, has addressed logging as minimum quality criteria for importing and uploading data into a research registry. It is also used "to maintain accuracy in identifying outstanding data" (ECRIN, 2010, 2013).

The majority of logging systems in clinical applications that are not used for security purposes are developed for event tracking. Patient events are tracked and clinical pathways suggested, often utilising other data (Z. Huang et al., 2014). Agosti, Crivellari, & Di Nunzio (2012) defined two research sources of general user logging: they focused on what the user actually entered and with regards to the preferences of the user. While the first point in this research is the evaluation and report of data quality, the second point is the analysis of user interaction with logging tools. One widely accepted tool is the Java-based library log4j. With this logging utility several levels of logs can be created. These level can be turned on/off or used to report debugs, fatal errors, report warnings, user interactions, or capture and report all information (Gülcü, 2012). This information is typically stored in a text file on the server where the web application is hosted (although this need not be the case and as a security issue it is generally preferred to store log files on other servers). Such log files can be analysed with other tools. Due to the volume of logging information that can be captured in large scale web applications, services like CouchDB and ElasticSearch can be used to clean and process the log files and subsequently identify events, e.g. logged in; edited record #23; logged out. The successor project

to log4j, LOGBack was designed by the developer of log4j (Ceki Gülcü). LOGBack is a more dynamic solution than other logging tools like log4j or SLF4J (Simple Logging Facade for Java)[10].

Studying user interactions based upon log information provides the possibility to track the efficacy of implemented data quality tools and the possibility to draw conclusions about the user knowledge and motivation of data entry.

5.2.1 Logging in ENSAT

One of the tools developed during this research for the ENSAT registry has been a log analyser. Key actions within the registry that indicate interactions with the data need to be recorded (record creation, modifications, deletion, reads, etc.) A typical log file is shown in Figure 5-7. Understanding who has accessed and used the registry is an important part of the auditing demands of the VRE but also part of data security standards. The log analyser uses a combination of a custom log-parser software system with cloud storage and a NoSQL database supporting Map-Reduce based visualisation. The scale of the logging information can be very large (many Gb per year) especially for ENSAT with over 170 registered user accounts and an average of 10 extended visits per working day (a typical research user can spend around one-hour interacting with the application), and many significant actions within a given web session can occur. Efficient cloud storage and rendering solutions were therefore required so a NoSQL database was chosen as a high volume, scalable and easy-access data repository solution. This system was deployed on the National eResearch Collaboration Tools and Resources (NeCTAR[11]) Research Cloud – the national academic research cloud across Australia.

```
31 May 2016 00:09:31 rootLogger  - username (page_nav.jsp) - i.piscaer@aig.umcn.nl
31 May 2016 00:09:31 rootLogger  - a5user - false
31 May 2016 00:09:31 rootLogger  - ('i.piscaer@aig.umcn.nl') - detail.jsp (Ensat ID = 'NLNI-0108')
31 May 2016 00:09:31 security.PolicyHandler  - ROLE 0: 1
31 May 2016 00:09:31 security.PolicyHandler  - sqlAction: SELECT DISTINCT SP_Role.sp_role_id,sp_action_name FROM SP_Role,SP_Role_Action_Assignment,SP_Action WHERE
SP_Role_Action_Assignment.sp_action_id=SP_Action.sp_action_id AND  SP_Role_Action_Assignment.sp_role_id=SP_Role.sp_role_id AND ( (SP_Role.sp_role_id=?) );
31 May 2016 00:09:36 rootLogger  - username (page_nav.jsp) - i.piscaer@aig.umcn.nl
31 May 2016 00:09:36 rootLogger  - a5user - false
31 May 2016 00:09:36 security.PolicyHandler  - ROLE 0: 1
31 May 2016 00:09:36 security.PolicyHandler  - sqlAction: SELECT DISTINCT SP_Role.sp_role_id,sp_action_name FROM SP_Role,SP_Role_Action_Assignment,SP_Action WHERE
SP_Role_Action_Assignment.sp_action_id=SP_Action.sp_action_id AND  SP_Role_Action_Assignment.sp_role_id=SP_Role.sp_role_id AND ( (SP_Role.sp_role_id=?) );
31 May 2016 00:09:36 rootLogger  - ('i.piscaer@aig.umcn.nl') - modality/home.jsp (modality = 'tumordetails')
31 May 2016 00:09:39 rootLogger  - username (page_nav.jsp) - i.piscaer@aig.umcn.nl
31 May 2016 00:09:39 rootLogger  - a5user - false
31 May 2016 00:09:39 rootLogger  - ('i.piscaer@aig.umcn.nl') - modality/detail.jsp (modality = 'tumordetails')
31 May 2016 00:09:39 security.PolicyHandler  - ROLE 0: 1
31 May 2016 00:09:39 security.PolicyHandler  - sqlAction: SELECT DISTINCT SP_Role.sp_role_id,sp_action_name FROM SP_Role,SP_Role_Action_Assignment,SP_Action WHERE
SP_Role_Action_Assignment.sp_action_id=SP_Action.sp_action_id AND  SP_Role_Action_Assignment.sp_role_id=SP_Role.sp_role_id AND ( (SP_Role.sp_role_id=?) );
31 May 2016 00:09:39 update_main.UpdateSub  - paramName (getParameterValues - updateSub): tumor_date
31 May 2016 00:09:39 update_main.UpdateSub  - paramName (getParameterValues - updateSub): tumor_form_date
31 May 2016 00:09:39 update_main.UpdateSub  - paramName (getParameterValues - updateSub): tumor_resected
31 May 2016 00:09:39 update_main.UpdateSub  - paramName (getParameterValues - updateSub): tumor_resected
31 May 2016 00:09:39 update_main.UpdateSub  - paramName (getParameterValues - updateSub): identification_date
```

Figure 5-7 ENSAT Log file output

The MapReduce visualisation application provides a variety of views to interpret the record histories – through patient IDs, centre codes, individual users (see Figure 5-8). It then renders the information through different graph representations (timelines, pie charts, bar graphs, etc.), again with efficient processing times in mind.

[10] http://java-source.net/open-source/logging

[11] www.nectar.org.au

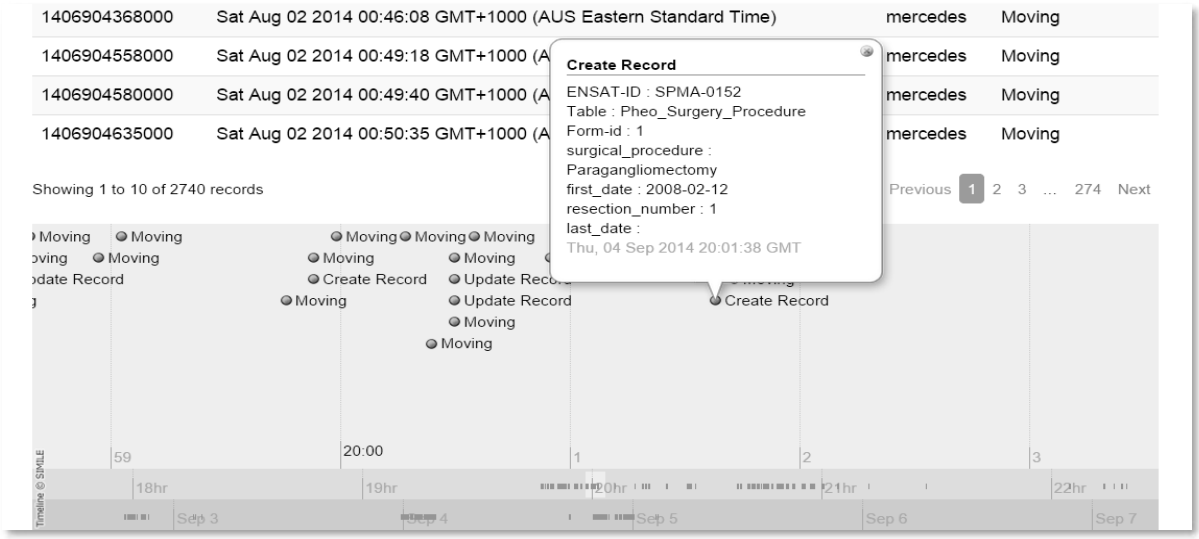


Figure 5-8 Tracking ENSAT user activities through log analysis

The architecture of the ENSAT log analyser provides information for users and internal/external auditors (see Figure 5-9). All internet-based access attempts are logged (successful and unsuccessful) and all activity inside of the ENSAT registry is captured. The logging data itself is periodically sent through a CronJob to the log analyser, which parses and processes the raw log file data. The log analyser application utilises CouchDB, ElasticSearch and MapReduce on the NeCTAR cloud to support converting the log data into understandable (human readable) content. The visualisation itself uses a timeline library (timeline.js), HTML, Jquery, AJA, CSS, JS and the CouchDB API (all software languages and libraries are deployed on the NeCTAR servers to ensure scalable and robust processing of logging information. Auditors have access to all of this information. The logging itself encompasses a range of key user tracking features:

- user movement or activities within the registry, e.g. navigation;
- interactions of users with the registry, e.g. searching;
- creating records;
- updating records;
- user logins, and
- access to and use of (specific) patient records.

The original log analyser was implemented by Effendy and its architecture and performance are described in Effendy (2014).

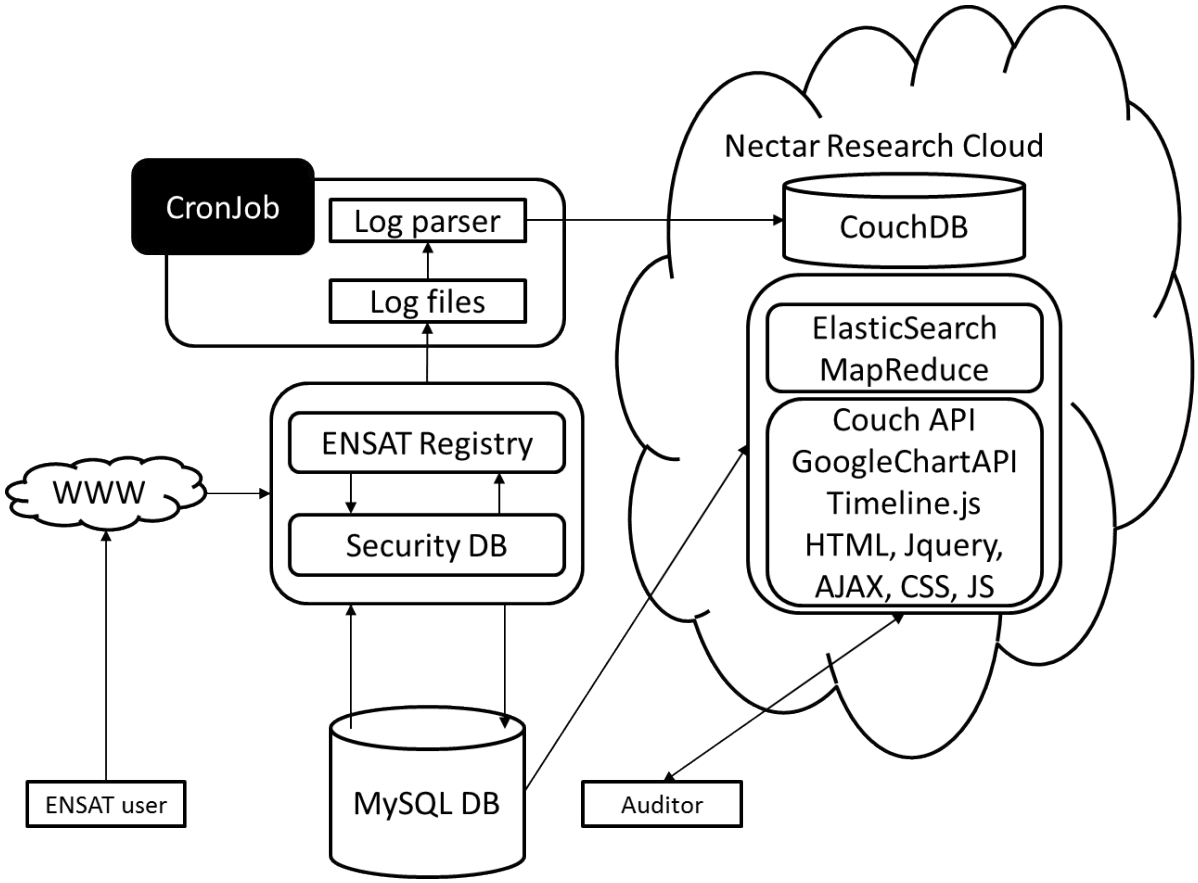


Figure 5-9 ENSAT registry logging architecture adapted from Effendy (2014)

The visualisations of the logging features are accessible to all members of the ENSAT research consortium. Dashboard, timelines and bar charts provide insights into user activities, their interactions and a record of the registry use as a whole.

It is important to note that the logging information provides two key services: to capture the history of the use of the ENSAT registry and to show the longitudinal history of the data that is collected on particular patients. Such a history is shown in Figure 5-10. This shows the history of editing of a particular patient (from Munich, Germany). The timeline shows all logged information from the original surgery and diagnosis in 2008 through to multiple treatments and follow up that have occurred throughout the patient's clinical care. This information flow captures the editing information captured in a log file. Similar capabilities exist to show the patients that are accessed and used the most in the registry, e.g. in search terms, although it is noted that this information is often limited by the security settings related to the individual patient records.

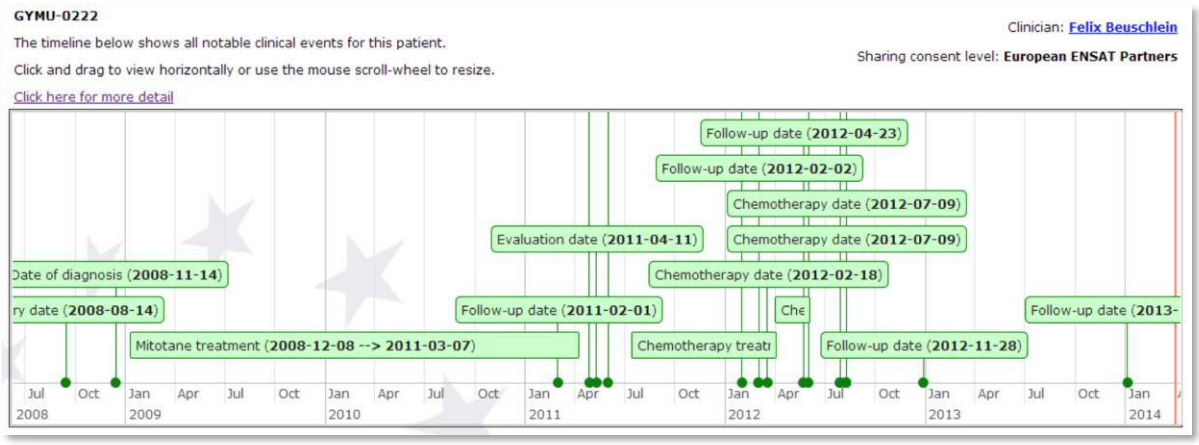


Figure 5-10 Visualisation of patient longitudinal information captured in log files

5.2.2 Google Analytics and ENSAT

Google analytics is a service to track user interactions with websites. The typical data that is collected includes country of origin, city of origin, session times and redirects. Such information can be used for features like budgeting and cost planning. It is traditionally used for benchmarking of web pages. For clinical research, Google analytics provides direct possibilities to monitor user interactions, session times and calculate a count of pages visited during a particular user session. This information is similar to logging tools, but Google provides this service for free for all registered users (see Figure 5-11).

Figure 5-11 Example of Google Analytics Interface for ENSAT

It is important that data from services and platforms such as Google analytics is interpreted to contextualise clinical user patterns (Crutzen, Roosjen, & Poelman, 2013). For instance, how long does a user stay on a site, how long does the user need to complete a particular record? Whilst a useful model to understand the global access and use of the ENSAT website, such information does not allow finer-grained access to information on use of the registry and user/session related information, e.g. how long particular users log in for and whether are frequent accessing the system. Nevertheless, tracking of information on the web site gives another dimension of the uptake of the ENSAT systems more generally.

5.3 Log Analysis Results of User Activity

5.3.1 ENSAT Log Analysis

The first assessment of the user logs provides a range of insights into the practical handling and usage of the ENSAT registry. In the observed time period, a genetic researcher within the ENSAT consortium recorded 32.7% of all activities when updating patient records (see Figure 5-12 and Figure 5-13).

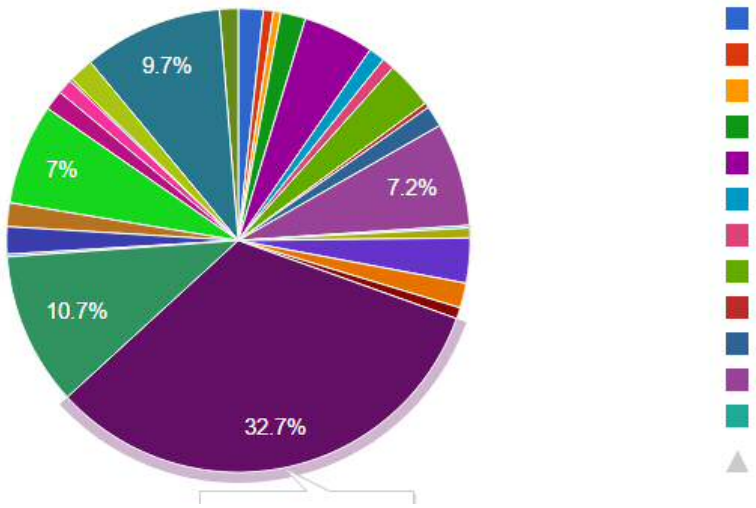


Figure 5-12 ENSAT updated records (user)

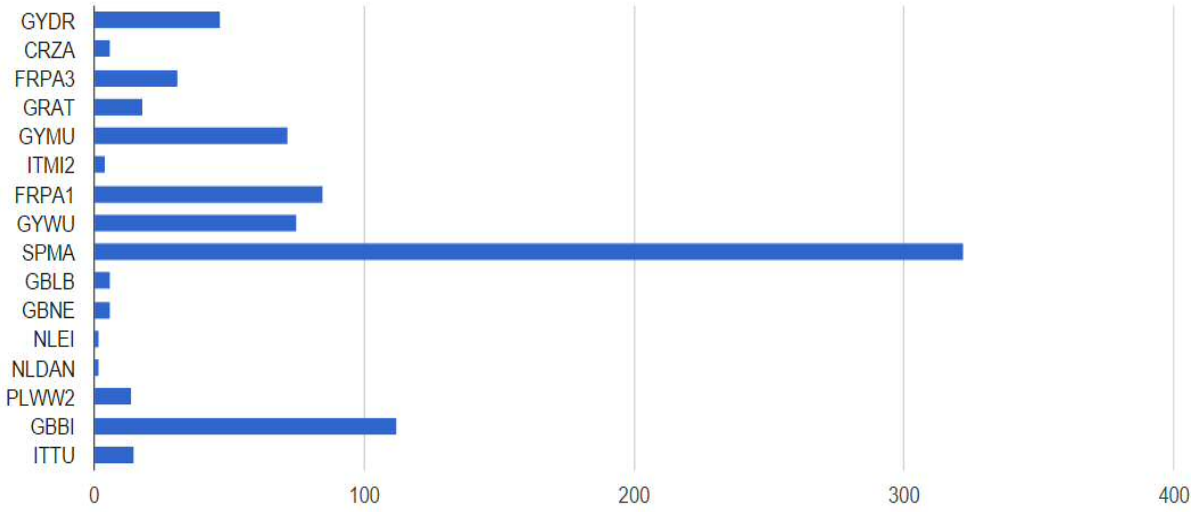


Figure 5-13 ENSAT Updated records (centre)

The research centre in Spain was by far the most active centre over the observed time. The centre in Spain receives biosamples and tumour tissues of adrenal tumours from all ENSAT centres to analyse genetic information. This information is entered into the genetic record of every subject that is captured in the ENSAT registry. Other centres did not show any significant activity in the observed time period. This may be because no patient information was entered or the centres only participated in biomedical basic science. For example, other centres like NLEI resample tumour tissue only for local research purposes.

Such data can track user activities that can inform data quality feedback and interaction. Logging provides the possibility to contextualise and inform the assessment of the efficiency of data quality feedback. Google analytics also provides key features that can be used to assess the overall utilisation and external (international) visibility of the system.

5.3.2 ENSAT Google Analytics Analysis

For the evaluation of user interaction within ENSAT, the year 2014 was chosen. In 2014, the EURINE-ACT case study with investigator and community feedback was performed. Centres were visited from January to May; reports were mailed to the community between March to September and the consortium was updated about the data quality assessments in April (Investigator) and November (all ENSAT members).

In 2014, an average of 435.67 sessions occurred per month. After the initiation of the ENSAT data quality project at the end of January, the user activity increased by 200 sessions (see Figure 5-14). Three German centres were visited at the end of January/beginning of February. By mid-February, the largest ENSAT centres that provided ACC information in Paris were visited, followed by three visits in Italy at the beginning of April. Unfortunately, the statistics showed no direct correlation between the data quality and the user activities. There were also no other events in the research consortium at this time (increased patient entry, new record forms, new clinical trials), so it was only possible to detect a correlation between the monitoring visits in Germany and France with regards to increased user activity. Specifically, we consider the motivation and data entry associated with the EURINE-ACT study. There was no increase in user activity detected after the visits to Italy, which might be due to the fact that the Italian centres where mainly focused on genotype data and the work/research balance for young medical scientists is unfavourable as explained in chapter 3.2.2.

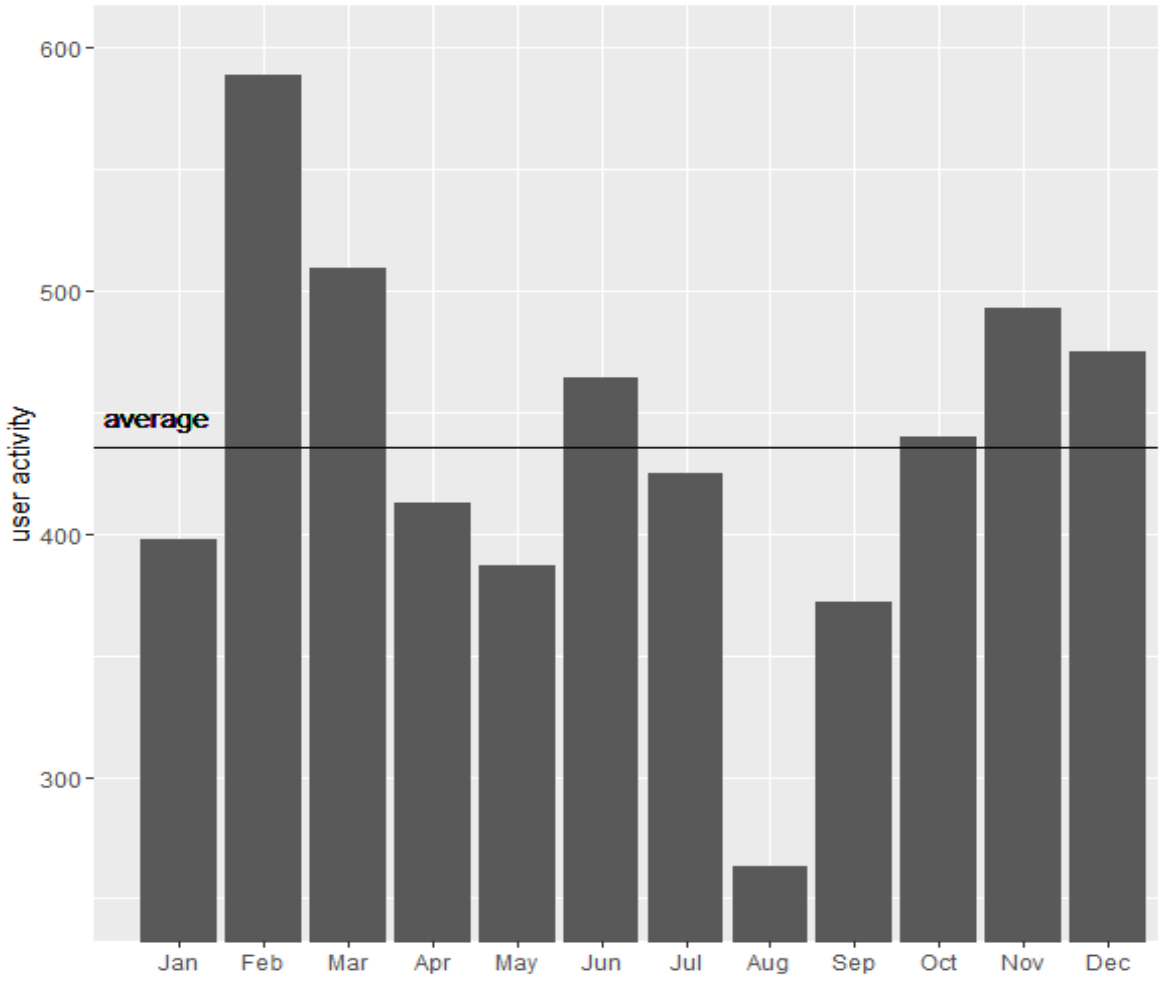


Figure 5-14 ENSAT Registry User activity 2014

Google analytics of the ENSAT registry commenced in January 2014, and no comparison data prior to this year is available. Nevertheless, we observed that in 2015 there was an average user activity of 694.08 sessions per month performed up to January 2016, and a further 1299.78 sessions per month up to September 2016. This is most certainly caused by the increased number of users and study activities. It is noted that the average activity was increased in the month of February and March 2014. This leads to the conclusion that monitoring visits and regular feedback forms a factor in the increase of user activity.

5.3.3 INPDR Log Analysis

Log analysis in INPDR was performed with Apache text analysis tools and use of a MySQL database (Rui, 2016). Every new session was logged and the number of actions performed on each record was recorded and subsequently analysed. In this analysis the time was measured and the pages through which the user performed particular actions. A typical example showing the creation, edit/modification and deletion of information is given in Figure 5-15. It is worth noting that the actual identification of records in the log files was made possible through the REST-based architecture, where http commands such as PUT, POST specifically included the record end-point (the uniquely identified record identifier such as NPC-UKBI-12).

Records	Start	End	Pages
3	30/Aug/2016 22:19:06 +1000	30/Aug/2016 22:20:41 +1000	patient
20	16/Sep/2016 09:20:05 +1000	16/Sep/2016 09:22:30 +1000	centre, physician,lab
99	23/Sep/2016 01:19:54 +1000	23/Sep/2016 03:24:53 +1000	various, mainly patient
13	23/Sep/2016 10:28:45 +1000	23/Sep/2016 10:36:31 +1000	physicians,subject
23	23/Sep/2016 12:04:33 +1000	23/Sep/2016 12:10:10 +1000	centre, physician, lab
29	26/Sep/2016 18:51:19 +1000	26/Sep/2016 19:52:39 +1000	npc_search, npc_visit
1	28/Sep/2016 14:33:09 +1000		nothing

Figure 5-15 INPDR log analysis

The figure shows that some centres only performed search queries, i.e. discovering patients that meet particular criteria. Such queries take little time to run and the user sessions are often far shorter, e.g. 1 minute or so. Other centres (Lyon, France) entered data for NPA and NPB patients that took of over 3 hours of online data entry time (Rui & Sinnott, 2016).

Similar to the data completeness analysis and primarily due to the lack of adoption of the platform (compared to ENSAT), no significant conclusions can be drawn from the INPDR log analysis. The users have a range of usage patterns showing that they have only sporadically used the system for data entry. This makes any detailed results difficult to establish. One other dimension to data entry is through the patients themselves. INPDR has developed a range of game application on mobile devices to help improve the diagnostics of children (Sinnott et al., 2015). Other mobile apps targeted to specific quality of life questionnaires are also being completed. It is hoped/expected that these will improve the quality of the data and the usage of the systems more generally.

Finally, it is worth discussing the importance of non-technical considerations on data quality. What makes a community such as ENSAT flourish whilst others such as INPDR struggle to truly galvanise the international community. A large part of this is getting to a critical mass of data, but it is also important that the clinicians and researchers themselves have extrinsic motivation. In ENSAT this is often through the very many clinical trials and studies that are taking place that demand and drive use of the registry to improve the data quality. For INPDR, at present no clinical trials have yet been established. It would be expected that when this arises, the usage and quality of the core registry data would increase considerably.

5.4 Chapter Conclusions

Unfortunately, for INPDR no significant conclusions can be drawn in regards to the impact of feedback on data quality. The primary cause of this is insufficient data and record entries within the observation period. Still, INPDR is a fully working platform that went through all of the proposed feedback stages. It was intended that Stage 2 should be as long as all other stages, however due to the observed low data entry, this stage was combined with stage 3 which itself went live after 2 weeks of stage 2. The justification for this was that both stages assess intrinsic motivational factors. Therefore, no other outcome could be detected in stage 2 than in stage 3.

Furthermore, it was shown that defining mandatory items works for improvement of data completeness in ENSAT and to a lesser extent in INPDR (where only a small number of data items have been defined as compulsory). It is expected, that all optional data items depend upon improvements to user motivation. Therefore, data completeness of mandatory items can be more readily increased and actually enforced than data completeness rates of optional items.

ENSAT logging and INPDR case study explored whether the patterns of use and impact on motivation of the data entry impacted on the overall data quality. The results here were mixed due to the amount that INPDR was used over the time period. Furthermore, it should be noted that to generalise the results, research hospitals and research units need better access to novel technology and solutions should be equally applicable to those centres in rural areas or in third world countries. Researchers needs to be more motivated to participate in research by showing them their impact on the research and clinical study objectives.

User logging information and general uptake is a key for the assessment of the efficiency of data quality feedback. Logging information is a rich source of information both for auditing and for visualisation purposes. During the logging analysis of INPDR and ENSAT technical solutions were developed to explore many factors of the usage. A more detailed analysis of the visualisations and the impact they have on data quality combined with resultant data quality scores is important to understand.

One major challenge with the INPDR analysis was the fact that once centre (Barcelona) has entered the vast majority of data into the registry with a low overall data completeness rate. Specifically, Barcelona entered 108 patients with an average completion rate of 31%. As a comparative example, Prague has entered 9 patients and an average completion rate of 81%. The interpretation of any meaningful results from the feedback systems that were put into place needs to be carefully understood.

6 USAGE AND ROLE OF PATIENTS ON DATA QUALITY

This chapter considers other usage patterns and their impact on data quality of disease registries. We also consider other dimensions of motivation and the impact that they can have on data quality in clinical registries. Specifically, we consider the impact of patients entering data themselves. We hypothesise that patients that have a disorder/disease will be more motivated to enter better quality data to help researchers and clinicians find cures for particular diseases. Given the ubiquity of mobile devices, we explore the role of such technology for data capture and investigate whether patients achieve higher data quality than researchers themselves. This chapter considers the example of the Environmental Determinants of Islet Autoimmunity (ENDIA[12]) project and the mobile applications that have been made available for patients to enter their own information.

6.1 Mobile Health

In 2007, Steve Jobs, former COE of Apple Inc., introduced three new products to the public: a new "widescreen iPod with touch controls", a "revolutionary mobile phone" and a "breakthrough internet communicator". His team put all of these devices in one mobile device, the iPhone. The following year changed the way in which *mobile phones* evolved into *mobile communication devices* that supported all aspects of communication: speech, video and text. Mobile devices are now an integral part of everyone's life. With the constant access to news, information, games and apps that support daily activities, new dimensions of data collection are now possible. This includes apps that can be used to collect daily information that can support the clinical and biological data typically collected in disease registries.

Nearly 75% of the world now has mobile access and developing countries have more mobile coverage than the developed world (ICD, 2012). The approach to support health related circumstances with mobile applications – so called mobile health (mHealth) offers many potential opportunities to capture and indeed deliver health-related data. mHealth has no general definition nor a primary public health related context that it can be applied to (World Health Organization, 2011). According to Kay et al. (2011) "mHealth is being applied in maternal and child health, and programmes reducing the burden of the diseases linked with poverty, including HIV/AIDS, malaria, and tuberculosis (TB)" (World Health Organization, 2011). Mobile health applications (apps) have been developed and applied to multiple disorders including developmental disorders, cognitive disorders, substance-related disorders, psychotic disorders, mood disorders, anxiety, eating and sleep disorders. Many apps have been developed for collecting data automatically from device sensors, e.g. tracking users mobility (or lack of mobility) through the location-based services that many devices now support whilst others have been developed for targeted data entry to collect information specific to the condition of interest (Luxton, McCann, Bush, Mishkind, & Reger, 2011).

Several systematic reviews showed that mHealth has a high potential to improve treatments or diagnostics of diseases, but more research must be performed (Catalani et al., 2013). In HIV research, a systematic review investigated mHealth applications that support the prevention, care and treatment of HIV. In 62 analysed articles between 2001 and 2011 they discovered that the majority of all studies were evaluation studies (56%), while they categorised all studies into monitoring, operational and evaluation research. Catalani et al. (2013) concluded that mHealth applications have to be implemented into the HIV treating cascade and that "the overall rigor of mHealth research […] needs to be improved …" (Catalani et al., 2013).

Similar findings can be confirmed through other systematic reviews, which investigated the impact of mHealth applications in chronic disease management. The authors selected 107 articles between 1980

[12] www.endia.org.au

and 2014 that met the inclusion criteria and described mobile applications that support diabetes mellitus, cardiovascular diseases, chronic lung diseases and hypertension. The research team identified several used tools: SMS, video messaging, applications or devices that added extra capabilities to the phone such as blood pressure monitors. They defined "mAdherence" as the core motivation whereby the patient is actively involved in the care of their disease through such mobile apps. It has been shown that, in 15 out of 27 randomised controlled trials mAdherence has directly impacted on patient's adherence behaviour. On the other hand, the authors conclude, that there is no strong evidence that mHealth improves patient's adherence. Recognition that more research has to be performed regarding low patient motivation through mobile devices was discussed in (Hamine et al., 2015).

A third systematic review by Aranda-Jan et al (Aranda-Jan et al., 2014) investigated the advantages and disadvantages of mHealth projects in Africa. Of the 44 research articles reviewed the authors discovered eight mHealth research categories: adherence, motivation/training of staff, evaluation research, stock management, patient education, surveillance, data collection and mHealth overviews. The review of articles classified as staff training, support and motivation (n=2) described two pilot studies that supported rural care scenarios and giving easier access to clinical information and facilitating data sharing between physicians. Research articles also described the impact of mHealth on data collection, transfer and reporting (n=10) where it showed that the use of SMS for data collection improves the overall data quality. However, several issues need to be addressed: low technology skills of staff members, high setup costs and risk of theft. The authors conclude that the "evidence remains poor" regarding the positive impact of mHealth applications in Africa (Aranda-Jan et al., 2014).

All three systematic reviews performed a literature analysis review of mHealth technologies applied to developing countries settings, to chronic diseases and to important infectious diseases. Whilst all three research groups established a positive impact of mHealth on the treatment, patient adherence and goals of data collection, they recognised that further research needs to be performed that can evaluate the findings and show a more general and more significant impact of mHealth technologies. Tomlinson et al. (2013) summarise this as, there is "insufficient programmatic evidence" of mHealth applications. The authors also recommend improvements of current standards and development of new standards and guidance for mHealth solutions. This is needed to scale-up mHealth in low and middle income countries, to implement factorial designs, to focus on user behaviour and to establish an open mHealth architecture (Tomlinson, Rotheram-Borus, Swartz, & Tsai, 2013).

The research group of Deborah Estrin (University of California) proposed an open mHealth architecture[13], which supports the interconnectivity of multiple mHealth apps (C. Chen et al., 2012; Estrin & Sim, 2010). The idea is that every user generates their own data with various applications running in the Cloud, which is managed by a centralised system (Estrin, 2014). Several applications could focus on special medical fields such as rheumatic diseases (Say et al., 2015) or HIV (Swendeman et al., 2015). Many standards for the evaluation of mHealth technologies have been published (Kumar et al., 2013), but the concrete evaluation and uptake of these standards remains under-achieved.

Nevertheless, mobile technology is advancing and almost every human will have access to a mobile device. As such mHealth needs to support a wide range of stakeholders including health organisations, professionals and patients to improve the prevention, diagnosis and treatment of a spectrum of diseases. In the previous chapters the intrinsic and extrinsic motivation factors of researchers and clinicians was assessed through collaborative research registries. In this chapter we consider the motivation of patients in clinical trials and if patients are driven to offer better data quality given they

[13] www.openmhealth.org

are the ones with the actual diseases or the risk of such diseases developing to themselves or their offspring.

6.2 Patient Incentives

The first approach to check whether patients can drive data quality is to assess the motivation of participation in clinical trials themselves. In a survey of 153 healthy participants of a volunteer drug injecting clinical trial, seven primary motivational categories were identified (see Table 6-1) (Fry & Dwyer, 2001).

Table 6-1 Motivation to participate in trials (healthy volunteers) (Fry & Dwyer, 2001)

Motivation	% (n)
Economic gain (money)	46 (71)
Citizenship (for the community)	38 (58)
Altruism (help)	19 (29)
Personal satisfaction	17 (26)
Drug user activism	16 (24)
Seeking information/assistance	5 (8)
Non-specific	5 (8)

Other research showed that people participating (n=13) in qualitative research are driven by: subjective interest, enjoyment, curiosity, introspective interest, social comparison, as well as therapeutic, material and economic interest (Clark, 2010).

Quilici et al. (2013) investigated the impact of using SMS technology after coronary stenting. The research compared the adherence of patients in a mHealth group with a control group and discovered that those in the mHealth group had a better completion rate than the control group (Quilici et al., 2013). Another article examined the design of a mobile application for diabetes mellitus type 1 patients (n=20, between 12-16 years). They implemented gamification features and discovered that frequency of blood glucose sampling improved, however they identified that more research utilising more data was needed (Cafazzo, Casselman, Hamming, Katzman, & Palmert, 2012). This research shows the significance of introducing mobile devices in daily care of patients and that they can indeed increase data collection, especially with some form of rewards. Other studies (Tatara, Arsand, Skrøvseth, & Hartvigsen, 2013) have confirmed these results with focus on type-2 diabetes patients. Their focus was on visualizing automated feedback from patients about blood glucose and their activity levels (steps walked) initially without data entry. Feedback was used to show the patient the journey to their goal (relation between blood glucose, physical activity and nutrition). During interviews, two of the 12 patients talked about a lack of motivation for self-management. The authors concluded that their qualitative study design only provided limited generalizable results, because the majority of all participants were highly motivated (Tatara et al., 2013).

In summary, technological advancement through mobile devices and apps are increasingly implemented into personalised health tracking, tracking movements like ‘quantified self’ and open mHealth platforms are increasingly required. They can be beneficial for the community and researchers, especially due to their always-accessible nature. Often they can be used for collecting data automatically, i.e. without explicit data entry through use of the accelerometer and the location-based services now existing in mobile devices. Systematic reviews discovered improvements in data collection and motivation, but the evidence is not yet established about their impact on data quality due to the typically small population sizes. Furthermore, more research has to be performed regarding the behavioural change of users and the impact on data quality itself (Sinnott et al., 2015).

To explore this, a case study was performed where the user interaction was logged and conclusions about the efficiency and quality of patient data entry explored. This was conducted in the context of the Environmental Determinants of Islet Autoimmunity (ENDIA) project.

6.3 ENDIA Context

The Environmental Determinants of Islet Autoimmunity (ENDIA) project was funded by the Australian National Health and Medical Research Council (NHMRC), the Juvenile Diabetes Research Foundation (JDRF) and the Helmsley Charitable Trust in 2013. The aims of ENDIA are to investigate the environmental factors that might be involved in the onset of type-1 diabetes (T1D) and specifically those related to pancreatic islet autoimmunity. ENDIA is a prospective cohort study across Australia (n=11 centres) with a recruitment aim of 1400 subjects. The inclusion criteria for participants include: age (less than 6 months) and where an infant has a first-degree relative with T1D. Mothers and infants are monitored before birth and subsequently followed until their early childhood. The trial aims are focused on the development of T1D during pregnancy, the microbiome during the first 3 years, weight changes and infections (Penno et al., 2013). The clinical databases include an extensive range of phenotypic data on both the parents and the pre-born and infant child. This includes extensive dietary information on the mother and more generally on the environment in which they live. This data is captured over an extended time period as shown in Figure 6-1 (for a small subset of the patients). Each column corresponds to a 3-month period (first trimester, second trimester, third trimester, birth, 3 months, 6 months etc.). At each time period, an extensive amount of information needs to be collected.

				Pregnancy			Postnatal											
Id	M	F	S	T1	T2	T3	B1	B2	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
NSW-SYD-CHW-001	✓	✓	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
NSW-SYD-CHW-002	✓	✓	✗	✗	✗	✗	✓	✗	✗	✓	✓	✗	✗	✓	✗	✗	✓	✗
NSW-SYD-CHW-003	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
NSW-SYD-CHW-004	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
NSW-SYD-CHW-005	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✓	✗	✗	✗
NSW-SYD-CHW-006	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗
NSW-SYD-CHW-007	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
NSW-SYD-CHW-008	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
NSW-SYD-CHW-009	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓	✓	✗
NSW-SYD-CHW-010	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗
NSW-SYD-CHW-011	✓	✓	✗	✗	✗	✗	✓	✓	✓	✓	✗	✓	✗	✓	✓	✗	✗	✗
NSW-SYD-CHW-012	✓	✓	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
NSW-SYD-CHW-013	✓	✓	✗	✗	✗	✗	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
NSW-SYD-CHW-014	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗
NSW-SYD-CHW-015	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗
NSW-SYD-CHW-016	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
NSW-SYD-CHW-017	✓	✓	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
NSW-SYD-CHW-018	✓	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✓	✗	✓	✓	✗	✗	✗
NSW-SYD-CHW-019	✓	✓	✗	✗	✗	✗	✓	✗	✗	✗	✓	✓	✓	✗	✗	✗	✗	✗
NSW-SYD-CHW-020	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✓	✗	✗	✗	✗
NSW-SYD-CHW-021	✓	✓	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
NSW-SYD-CHW-022	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
NSW-SYD-CHW-023	✓	✓	✗	✗	✗	✗	✓	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗
NSW-SYD-CHW-024	✓	✓	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✓	✗	✗	✗	✗	✗
NSW-SYD-CHW-025	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗

Figure 6-1 ENDIA Phased Data Collection

Many of these data forms are especially data demanding and place considerable onus on the data entry personnel and indeed the study participants (mothers/fathers). For example, the dietary information guidelines alone are based on a 167-page manual that documents all of the dietary – related data items that need to be collected and the processes that must be taken to collect the data. The actual number of patients entered into the system is shown in Figure 6-2 and the number of patients (involved) per site shown in Figure 6-3.

Figure 6-2 ENDIA Recruitment Actual (October 2016) and Target (September 2017)

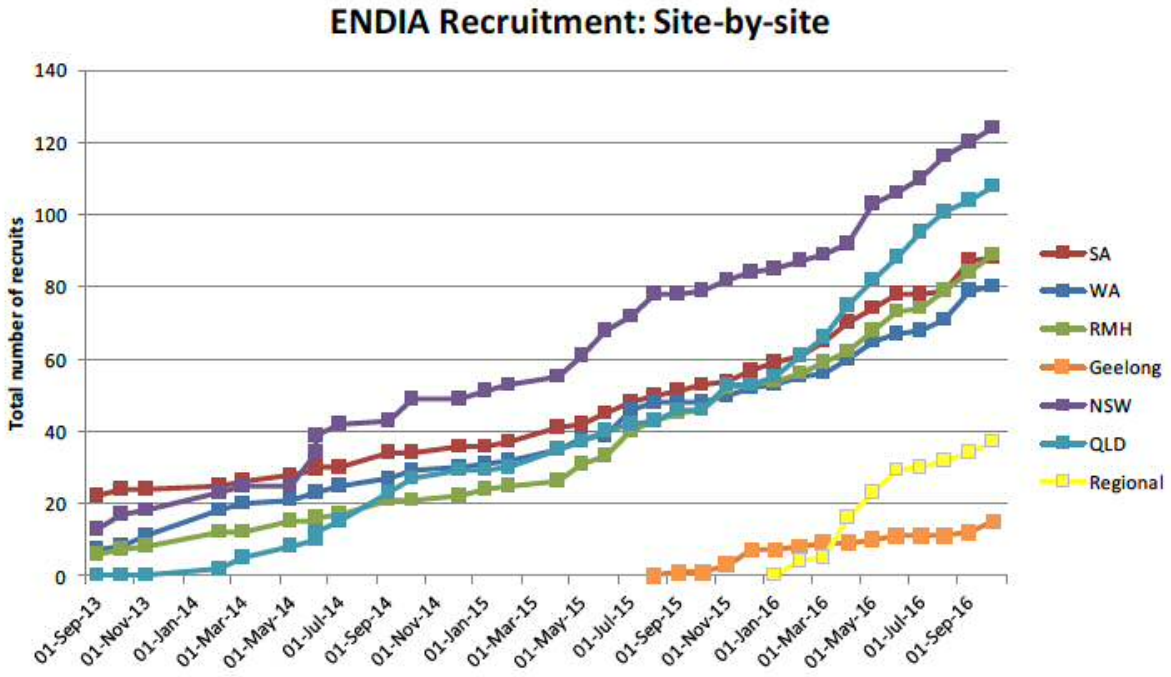


Figure 6-3 ENDIA Recruitment by Site

Audit information on access and use of the ENDIA database has been collected since the start of the project in 2013. The evaluation of data completeness was implemented in the registry, similar to INPDR. Analysis of data was conducted to detect potential data quality issues with specific focus on incompleteness and any identified data inconsistencies. A log analysis of the use of the registry was conducted through CouchDB using MapReduce and the Python API (see Figure 6-4).

Field	Value
_id	"1ca88650a63792e97bf8f243b7f6850f"
_rev	"1-5fb7800b0831cb69440aa4a8d6ece6a9"
Centre_id	"5"
activity	"Update"
average_completeness	0.4
completeness	0.4
date	"2014-02-20"
document_finished	1
local_id	"022"
newcontent	"utf8=>✓, authenticity_token=>kx/VPhXIroyI4cMwOi0u7zbvKak561yY5Z5AK6mimaU=, participant=>local_id=>22, user_id=>4, clinical_centre_id=>5, con..."
old_content	"utf8=>✓, authenticity_token=>e4iW0jEH8RkrYGojrqoe9ZcakxDQ5YDdILJoEgFFQeg=, participant=>local_id=>022, user_id=>4, clinical_centre_id=>5, co..."
participant_id	26
time	"13:19:38"
total_edits	2
type	"Participant"
version	2

Figure 6-4 Example of ENDIA logging information

All user actions (login, deletion, withdrawal, update or creation of information) were analysed. For visualisation the software d3[14] (data-driven documents) was used. D3 is a JavaScript based library suitable for a range of data visualisation scenarios. The advantage of D3 is that it does not require a "toolkit-specific lexicon of graphical marks" (Bostock, Ogievetsky, & Heer, 2011). As such, it can be used on a large variety of platforms because it selects the object not based on the mouse curser, but based on the data element, class attribute or tag names (Bostock et al., 2011). The library has a large amount of different layout schemes including calendar views, box plots, scatterplots, bar charts, geographical maps, hierarchical edge bundling and force-directed graphs.

A key aspect for understanding the quality of the ENDIA data is the patterns of data entry by data entry personnel – these are typically research nurses employed at the sites that are involved in ENDIA. Figure 6-5 shows the times of day that data entry personnel typically enter data into the registry. These results are obtained through running MapReduce algorithms focused on the specific time of day for any data entered into the registry. Importantly, through such algorithms it is directly possible to detect the statistical patterns of individual users and centres. For example, it is possible to understand how often do personnel edit the same records and at what point does this become the optimal – in terms of the quality of the data that is entered. Whilst one would assume that the quality and completeness of the data increases indefinitely as more edits are made, this is not typically the case. Since the amount of data that is to be entered follows a longitudinal timeline, different forms are activated at different time periods, e.g. when the infant has reached 6 months the 6-month data entry form is activated and available for editing.

[14] https://d3js.org

Figure 6-5 ENDIA Timeline of Data Edits

It was detected, that users with at least 35 edits of the same record have the highest data completeness scores (see Figure 6-6). This is based on the completeness of the records that have been entered up to that point and not all possible data that might be completed up to the child reaching 30 months (V10).

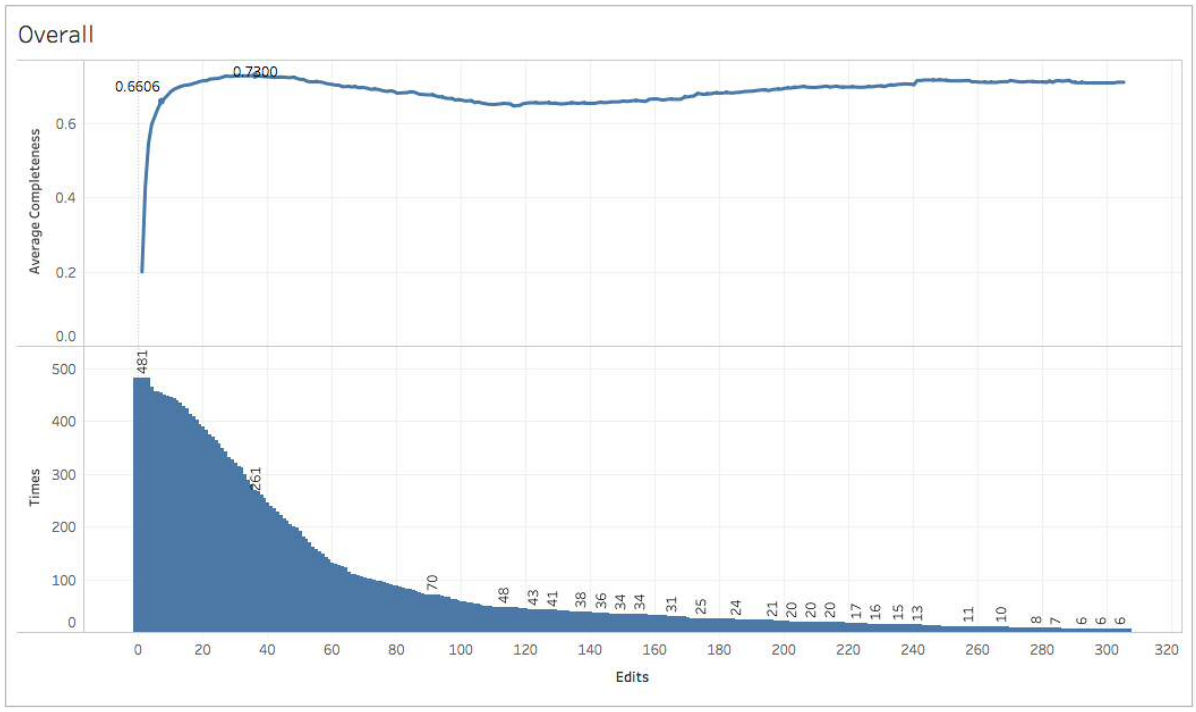


Figure 6-6 ENDIA edits vs. data completeness

The centre with the highest data completeness rates (up to October 2016) is the Women's & Children Hospital in Adelaide. It is also noted that this centre has an average amount of edits compared to the centres with lower data completeness rates. This might be construed as data completeness is not influenced by average and total record edits of the centres. However, Barwon Health has only 10

subjects entered in the protocol and has achieved a data completeness of 73%. The Women's and Children's Hospital and Barwon Health can't be compared because of the large discrepancy in the number of records entered into the registry (see Figure 6-7).

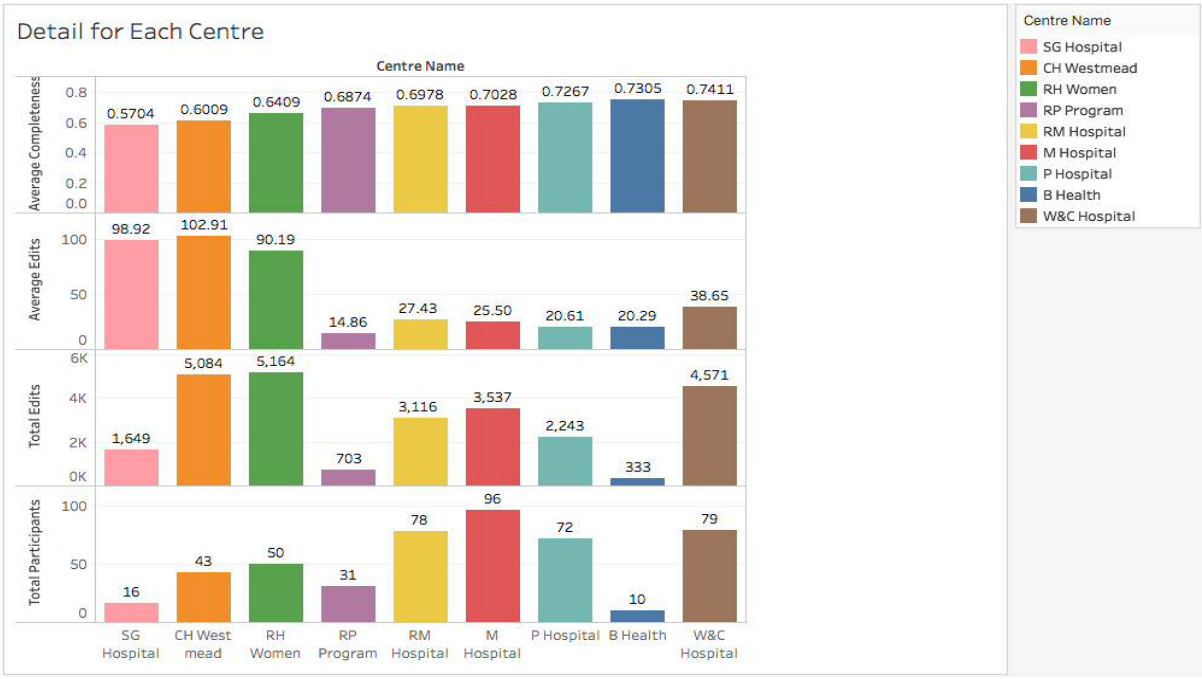


Figure 6-7 ENDIA average completeness rates (centre)

This indicates that average numbers of edits do not automatically lead to higher data completeness rates – at least for ENDIA. Westmead Hospital in Sydney has the highest average but the second lowest overall completeness score. They have also entered less than half the number of patients (43) compared to the other centres, e.g. Mater Hospital in Brisbane (96). From this, one can conclude that data entry personnel in Westmead may require further training or guidance in using the system. As identified in previous chapters, this may also be an artefact of the local processes and environment in which study personnel are collecting data. A discussion with the principal investigator and the report of these findings may thus be advisable.

6.4 Can Patients drive Data Quality?

As shown in Figure 6-1, there are numerous (longitudinal) points where patients are expected to provide information. One of the key aspects of ENDIA is considering how environmental factors that can impact on the onset of type-1 diabetes for children. One of the major environmental considerations under investigation within ENDIA is dietary information. As noted, this information is extensive with 167-page manual covering all of the potential information that should be collected in the system. In the first two years of the project it was identified that the research nurses were not systematically collecting this data in their visits and discussions with the patients. To help tackle this, it was decided to develop mobile applications that could allow the patients themselves to enter their own dietary data directly.

This mobile app was introduced in April 2016 and has a small cohort of users (n=20), however this is gradually increasing, as more patients are involved in the study. The app was delivered through both Apple iOS and the Android platforms and made available to all patients involved in the study. It is noted that whilst these apps are available within the iPhone AppStore and Google Play environments and hence downloadable by anyone (including those outside of the study). The app itself requires an activation code that is created by the ENDIA study personnel for the patients that wished to be

involved in using the mobile app. The app itself focused on a subset of the key dietary intake questions that were essential with 50 different data items to be entered each week. The app was developed through close interactions with the ENDIA study teams and evaluated initially with a few patients, before being made available more widely to the ENDIA patient cohort. Figure 6-8 to Figure 6-11 show some of the core functionality of the mobile application that was developed for dietary information collection. Figure 6-8 shows the initial opening screen of the application (after it has been activated). Figure 6-9 shows the weekly dietary forms that have to be completed by the patients. As noted there are no identifying information that tie the app to the specific patients themselves, e.g. their names, dates of birth etc. All patients are assigned a unique code (e.g. VIC-MEL-MEG-001 for the first (001) dummy patient from Victoria (VIC), Melbourne (MEL), Melbourne eResearch Group (MEG)).

A small subset of the data that is collected by the dietary form is shown in Figure 6-10. This can be sent directly to the ENDIA database (Figure 6-11) or saved and edited. Once the dietary information has been submitted for a particular week, the data is sent to the database and a local (non-editable) copy of the data stored on the mobile device. That particular week is then no longer possible to complete by the end users, i.e. the week is greyed out.

The mobile apps also include feedback mechanisms (notifications) that can be sent to the patients to remind them to complete their dietary information. The apps also include a range of data quality enforcement capabilities directly. Thus the apps require that all of the data for a particular week is collected and they raise warnings if patients attempt to submit incomplete data. The apps also do not allow data to be entered in the future, however they do allow for previous (historic) non-completed or partially completed dietary information to be entered and submitted to the ENDIA database.

Figure 6-8 Patient Select

Figure 6-9 Weekly overview

Figure 6-10 Survey Example Question 1

Figure 6-11 Saving actions

Before the ENDIA mobile app was developed for the patients, nurses and participants met directly (face-to-face) and patient related dietary information collected. In October 2015 it was identified that the data dropped by more than 20% - due it was presumed to the verbosity of the data to be collected and the fact that the data had to be continually collected each week. With the implementation of the mobile app in April 2016, it can be observed that the overall completion rate increased slightly (see Figure 6-12) although this is far from 100% completion rates as originally hoped/expected by having patients enter their own data at any time. This may be an artefact of new technologies needing more training time and support for patients to enter more data.

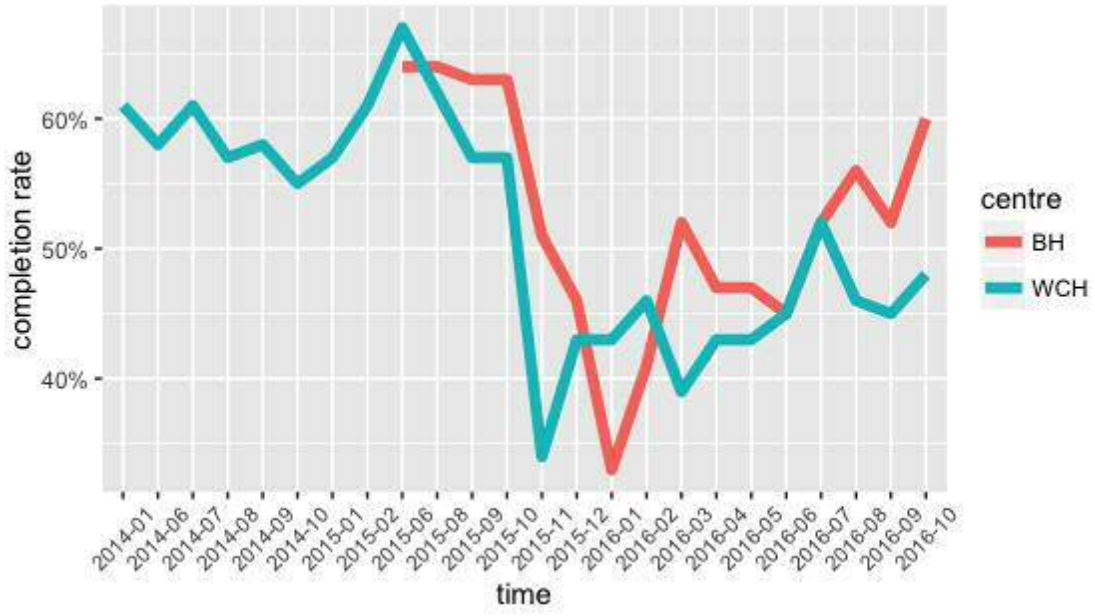


Figure 6-12 ENDIA data completion rate

The overall completion rate for the *dietary information* has improved since the mobile app went live. However, the overall completion rate for ENDIA as a whole remains a challenge. Indeed, as with any major study such as ENDIA, the attrition rate of patients being initially enthusiastic and/or convinced to be involved can gradually wane as shown in Figure 6-13.

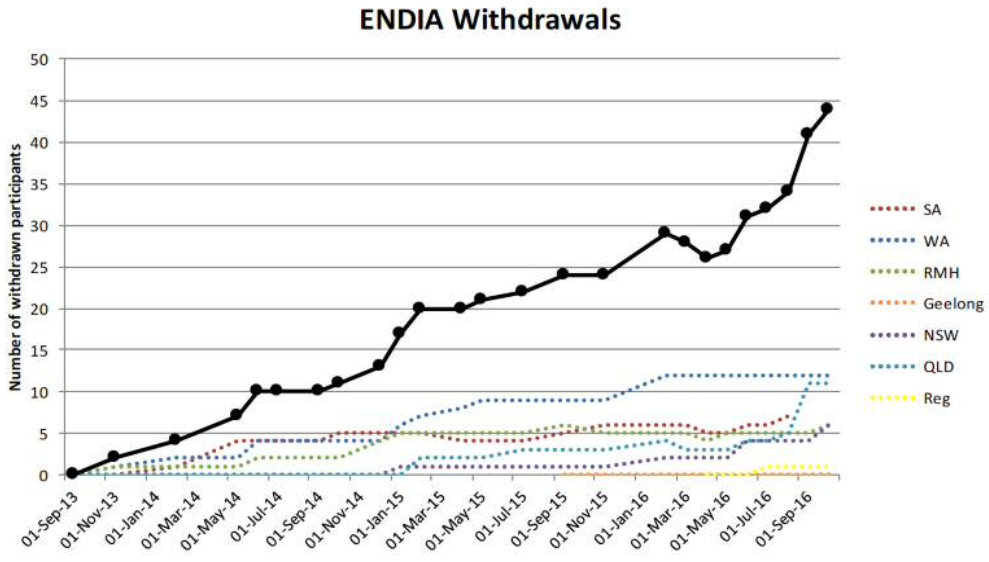


Figure 6-13 ENDIA dropout trend (September 2016)

The evaluation of the ENDIA study and patient engagement through the use of mobile technology remains somewhat inconclusive. At present more patients and more time is needed to systematically evaluate the improvement in data quality achieved through patients being self-empowered to enter their own data. A further factor in this is the feedback that patients themselves receive. Simply asking patients to enter data with no reward may not be continually sustained. Feedback mechanisms currently under consideration include targeted reports to the patients that benchmark their baby with other babies involved in the study. ENDIA is also a great adopter of social media with Facebook groups established. This has shown direct correlations and engagement with the study with targeted recruitment activities resulting directly from the periodic updates to the Facebook groups (see Figure 6-14).

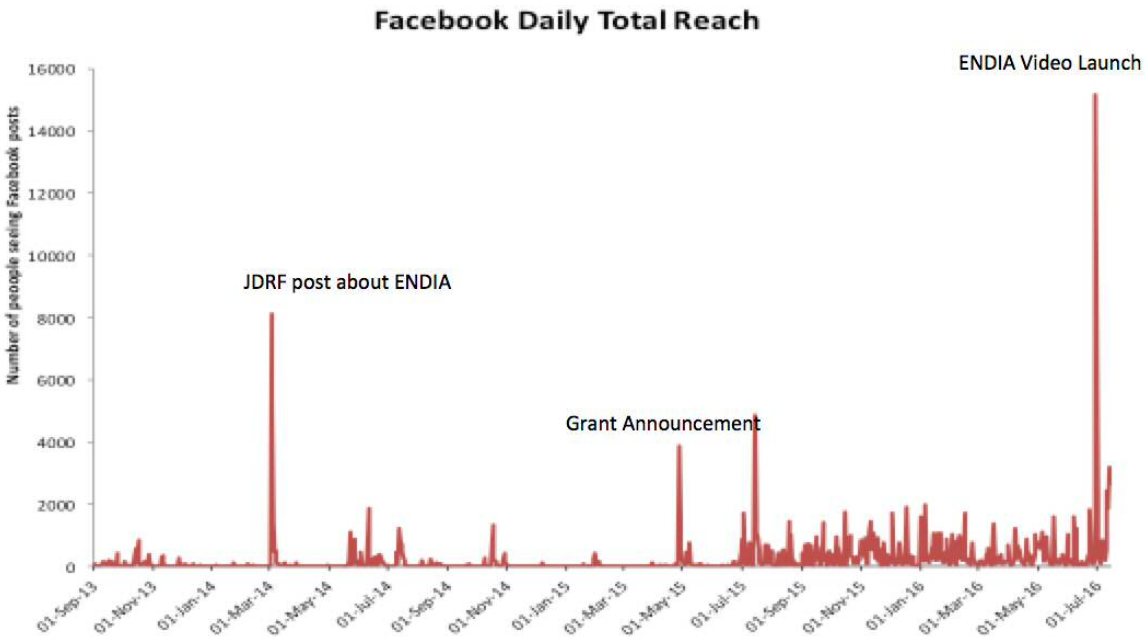


Figure 6-14 ENDIA Social Media Engagement

This model of adoption of social media is another mechanism to keep the participants involved and motivated. Such models of using the web and targeted technologies offers another dimension to web-based feedback proposed here that targets the specific needs of the patients and families (Moorhead et al., 2013; Nurse et al., 2014).

6.5 Chapter Conclusions

Through the technological advancement and the availability of mobile devices, nearly every human now has access to the Internet from everywhere. mHealth technologies are increasingly important and will be a key part of new research initiatives. There is still no significant evidence that the application of mobile devices in clinical care has quantitatively improved data quality and data capture processes. This chapter has presented an approach based on an experimental pre- and post-test regarding the change of data completion and data quality in the ENDIA project focused specifically on dietary information considering the research nurse (pre-) and the patient (post-).

The results suggest that patients do capture slightly more data through the mobile apps, however they do not yet achieve the desired data completion rates. Furthermore, the study design needs to be improved regarding post-tests and longer observation times in order to draw significant conclusions. Thus ideally having patients enter the data themselves through the web application and then through mobile applications could be assessed and whether the mobile technology itself has systematically

increased the quality of the data in the registry. The data shows that there is an increased trend of data completion, which can in part be based on the learning curve of trial subjects. As time progresses and more data is collected, we would be able to assess more rigorously whether patients offer better quality data as they are directly (intrinsically) motivated.

7 IMPACT OF INVOLVEMENT IN CLINICAL STUDIES ON DISEASE REGISTRY DATA QUALITY

This chapter explores how involvement in clinical trials and studies can influence the quality of data in disease registries. We focus specifically on two clinical trials associated with the ENSAT registry: the Prospective Monoamine Tumour (PMT) study and the prognostic value of urine steroidobolomics in patients with adrenocortical tumours (EURINE-ACT) study. We assess whether centres that have a direct involvement in these clinical studies have improved the quality of the data in the data registry overall, and/or whether the data quality has only improved with regards related to the data required for the clinical trials themselves. We consider participating centres in the studies and the lead study sites and the role of vested interest. Thus in this chapter we seek to identify the extrinsic motivational factors that can improve data completeness and data quality. Implicit in this assessment is the relationship with the central disease registry and the associated clinical studies. This motivational factor can be considered directly when one considers that the data can flow from the disease registry to the clinical study and vice versa. We describe some of the capabilities and challenges in facilitating such data flows through ENSAT.

7.1 Data Interoperability

Key to the establishment of disease registries is the knowledge exchange between collaborators, and allowing others to take advantage of centralised and shared data for new trial intentions and allowing secure access to any associated biomaterials. Literature in this area of research is often labelled as 'data exchange', 'data warehousing' and/or 'data pooling'. Implicit in this is ensuring data interoperability and tackling data heterogeneity.

For data interoperability several clinical data standards such as CDISC[15] (Clinical Data Interchange Standards Consortium) or HL7[16] (Health Level 7) are now well established, both in industry and academia. These standards have been developed to enable interoperability of medical data in clinical care and medical research. CDISC has different data models. CDISC-ODM[17] (Operational Data Model) is used for data interchange between regulatory submissions and (meta-) data collected in clinical trials. EDC systems can write ODM compatible files and thereby strengthen the network between clinical health records and CRFs (Ohmann & Kuchinke, 2009; Prokscha, 2007). Other CDSIC models are used for laboratory purposes (CDISC-LAB[18]), study data tabulation (SDTM[19]), analysis datasets (ADaM[20]), protocol representation (PRG[21]), standards for exchange of non-clinical data (SEND[22]) and case report tabulation data definition specifications (CRT-DDS). On the other hand, HL7 is widely used in healthcare data exchange (Ohmann & Kuchinke, 2009) with rich and extensive information models and associated messaging capabilities.

Every standard has its own purposes and the implementation of any given standard is driven primarily by concerns about costs, quality, patient safety and knowledge transfer speed from bench to bedside (Richesson & Krischer, 2007). The Biomedical Research Integrated Domain Ground (BRIDG[23]) project developed a data standard for biomedical research that provides a semantic foundation for applications of CDISC. HL7 and the National Cancer Institute (NCI) through the project Cancer

[15] https://www.cdisc.org
[16] http://www.hl7.org/
[17] https://www.cdisc.org/standards/transport/odm
[18] https://www.cdisc.org/standards/foundational/lab
[19] https://www.cdisc.org/standards/foundational/sdtm
[20] https://www.cdisc.org/standards/foundational/adam
[21] https://www.cdisc.org/standards/foundational/protocol
[22] https://www.cdisc.org/standards/foundational/send
[23] https://www.cdisc.org/standards/domain-information-module/bridg

Biomedical Informatics Grid (caBIG) project focused on data exchange and processing in the cancer domain also supported similar functions (Buetow & Niederhuber, 2009). The stakeholders of these domains included CDISC, HL7, NCI and the US Food and Drug Administration (Ohmann & Kuchinke, 2009). It is worthwhile noting that in 2011 a report on caBIG raised significant questions about the effectiveness and oversight, and its budget and scope were significantly trimmed. In May 2012, the National Cancer Informatics Program (NCIP) was created as caBIG's successor program (Mongkolwat, Kleper, Talbot, & Rubin, 2014).

In these efforts it was shown that CDSIC-SDTM was a helpful tool for data warehousing and CDISC-ODM was suited for clinical trial data and trial audit information (Kuchinke, Aerts, Semler, & Ohmann, 2009). ECRIN requirements recommended the use of CDISC-ODM to target data especially in CSV or XML files. However, it was identified that, despite the advantages of single centre clinical trials, it was "impractical… to insist on the use of specific data standards" (Ohmann et al., 2011). In short, they identified the major challenges of data heterogeneity and IT heterogeneity across the clinical landscape: from primary care, to secondary care through to clinical research systems and databases. A multitude of hospital systems exist and mandating any given standard or solution was impractical and unworkable especially given the demands for sustaining existing (legacy) systems. However, many reports have stressed the importance of standards and harmonisation in large multicentre trials (ECRIN, 2013; Ohmann et al., 2011).

In this context, it has also been recognised that there is a lack of defined research data standards for descriptive information including the study design, the type of randomization, data blinding and standardised (or commonly accepted) coding of record forms (Richesson & Krischer, 2007). It was also shown that for clinical care - case report forms in areas of physical exams, medical history and eligibility criteria have typically no established named standards and ad hoc approaches are typically manifest (Richesson & Krischer, 2007).

The GCP Inspection Group recommended an implementation of described data standards into clinical research databases to simplify data exchange with hospital systems (GCP Inspections Working Group, 2007). While the majority of volunteers in CDISC have a background in biotechnology (51%), only 21% have an academic or government-related background (Souza, Kush, & Evans, 2007). The GCP Guidelines suggest that long-term storage of essential trial-related documents and data has to be captured in study databases and subsequently address longer term storage and preservation demands (International Conference on Harmonization, 1996). Often heterogeneous data management solutions, even in the same institution, make new trials difficult to undertake or in certain cases, difficult to understand and interpret the results (Kuchinke et al., 2009). Ideally data standards such as CDISC-ODM for data exchange have to be applied on data sets in a more rigorous and widely adopted (ubiquitous) manner (Richesson & Krischer, 2007). It was further argued that this should be mandatory to achieve interoperability of clinical and biomedical datasets (Brandt et al., 2006).

Data standards are recommended for virtual (collaborative) research environments and especially those in multicentre research settings. In ENSAT, data has largely been defined by its purpose for the clinicians that are involved in the individual centres. Consensus and agreement has been a core achievement of the broader ENSAT network and whilst all hospitals involved have their own independent IT systems in place, agreement on the core data to be captured in the centralised databases has been achieved. In particular, every working group (NAPACA, ACC, PHEO, APA) defined their own relevant data needs (data variables) including the data field to be captured and the range and purpose of the data. However, not every data item is needed in every clinical trial. Therefore, many more variables can be found in the ENSAT registry that are related to the type of tumour and it is typically the case that a subset of these variables may be relevant to particular trials. The trials

themselves will define their own data requirements as part of the study protocol and these will typically be far more demanding than the core data captured in the registry.

The following evaluation of "vested interests" and extrinsic motivational factors by involvement in clinical studies focuses on the ENSAT PMT and EURINE-ACT clinical studies. These studies are focused on the PHEO and ACC parts of the ENSAT registry respectively. In particular, we consider the extrinsic motivating factors that might influence the quality of the relevant PHEO and ACC data sets on the registry related to these clinical studies. Specifically, do sites with vested interest in clinical trials improve the quality of the data (in the registry) that is related to those clinical studies?

7.2 PMT Study

7.2.1 Background

As noted, the PMT study (Prospective Monoamine Tumour study) is a four-phase clinical trial that tracks patients who exhibit clinical indications of suspected pheochromocytoma through any of the following criteria:

- Signs and symptoms;
- Therapy-resistant hypertension;
- Incidental finding on imaging for related condition;
- Routine screening due to known mutation or hereditary syndrome, and
- Routine screening due to previous history of pheochromocytoma

Patients are admitted to the study from the specialist clinics in Germany (Dresden, Munich, Wurzburg), Netherlands (Nijmegen) and Poland (Warsaw). These patients are then tracked through the four phases of the clinical trial, which includes screening, clonidine tests, characterisation, follow-up. These data are collected throughout the trial and up to five years after the study. The clinical path of patient progress through the study is outlined in Figure 3-3. The full set of information captured in this study covers the sections outlined in Table 7-1.

Table 7-1 Overview of the PMT study database

Phase 1: Screening
Identification
Demographics
Medications
Screening biochemical tests
Signs and symptoms
Tumour details
Other cardiovascular diseases and malignancies
Hereditary PPGL syndromes (genetics)
Phase 2: Clonidine Testing
Medications
Overnight urinary metanephrines
24-hour urinary catecholamines
Plasma chromogranin A
Clonidine tests
Complications
Post-clonidine metanephrines/catecholamines

Phase 3: Tumour Characterization
Medications
Biochemical tests (carried over from phase 2)
Ambulatory blood pressure monitoring
Cardiovascular tests
Echocardiography
Electrocardiogram
Metabolic tests
Imaging tests
Phase 4A: Excluded Follow-up
Follow-Up
Medications
Phase 4B: Pheo Follow-up
Genetics
Unresectable tumor: imaging
Resectable tumour: surgery/pathology
Resectable tumour: post-operative verification
One year follow-up: medications
One year follow-up: biochemical tests
One year follow-up: cardiovascular
One year follow-up: blood pressure
One year follow-up: echocardiography
One year follow-up: electrocardiogram
One year follow-up: metabolic

The tables that translate directly to the ENSAT registry exist primarily in phase 1 (Identification, Demographics, Tumour Details and Genetics). The PMT eCRFs were designed explicitly with similar parameter fields so that transfers between both the ENSAT registry (the PHEO/NAPACA databases) and the PMT clinical study could occur. As a result, patients that are recruited prospectively to the PMT study can subsequently be entered into the ENSAT (PHEO/NAPACA) registry and patients from the ENSAT (PHEO/NAPACA) registry can be recruited to the PMT study. It is noted that all clinical studies associated with the ENSAT registry have to apply for their own individual ethics approval and obtain consent from the patients that are involved. Thus the data flows are not completely automatically enabled but require strict adherence to the standard operating procedures and protocols that have been established throughout the course of the ENSAT project more generally.

In the case that a PMT patient has or had a metanephrine secreting adrenal tumour (PPGL) or hormonal inactive tumour (NAPACA), the record can be/has to be transferred from PMT to the ENSAT registry. With the transfer, data and biomaterial can in principle be available for all ENSAT researchers (provided the patient has signed the ENSAT informed consent form for biobanking and subsequent (physical) data exchange).

To understand the information flow within the PMT and ENSAT systems, we describe a typical use case scenario that illustrates how patients with an adrenal tumour are recruited to a clinical study such as PMT.

During a routine diagnostic, an adrenal mass is detected at the top of a patient's kidney. Following this identification, the patient is referred to a local endocrine research unit. At this point, the patient is

interviewed and receives consent forms related to the PMT study including the trial aims and the treatment plan (PMT flow chart). Once the patient agrees to be involved in the PMT study, i.e. after they have signed the consent forms, they are introduced to a clinical researcher, who collects all PMT Phase 1 related information. This information will be entered directly in the PMT eCRFs or captured on paper based case report forms (pCRFs) so that the data entry can be performed retrospectively. When registering the patient data, the patient is assigned a pseudonym. This is a unique identifier generated the PMT study system. Importantly this identifier is linked through the ENSAT registry through an incremental clock/counter-based system (Stell & Sinnott, 2012), e.g. if this patient is the 6th patient from Dresden then they are assigned with a unique identifier by the ENSAT registry (ending with GYDR-006) where the prefix reflects Germany, Dresden. This central ID generation greatly simplifies the subsequent data management and data relationships between the registry and the eCRFs related to the clinical study. After the ID is generated, the information identified in Phase 1 of Table 7-1 can be entered (captured) into the PMT eCRFs.

Data entry users have the possibility to transfer the patient information into the ENSAT-registry. The user can choose the level of access to the subsequent data (local, national, PMT study members or international collaborators) along with the ENSAT database (NAPACA or PHEO) depending on whether the patient has been confirmed with a PHEO or a NAPACA tumour. This data transfer web interface is shown in Figure 7-1.

Figure 7-1 Data transfer from the PMT study to the ENSAT Registry

The same transfer possibility exists within the ENSAT registry and data flow to the PMT study database. Often this data flow correlates with patients that meet the criteria for recruitment to the PMT study and once consent is approved, they can be recruited into Phase 1 of the study with the pre-existing ENSAT data automatically transferred to the PMT database. Both ENSAT Pheo and NAPACA patients can be transferred into the PMT study. It is noted that the ENSAT registry data entry user has to select the specific inclusion criteria for transfer of the patient record to the PMT study (see Figure 7-2).

Figure 7-2 Data transfer from the ENSAT Registry to the PMT study

The structure of the vast majority of data items in the ENSAT registry and those required specifically for PMT is different, since PMT is a single study related to the ENSAT database. It is noted that at present over 25 major clinical trials were ongoing and associated with ENSAT (November 2016). Furthermore, much PMT data is not required for ENSAT (for example very detailed medication information; history of cardiovascular diseases information or TMN classification). Registries and clinical studies differ in their goals and needs. Typically, a study will have much finer grained data requirements than a general disease registry. During data transfer, the data may need to be restructured so that it is compatible with the other database. Data validation on data transfer is essential in supporting such data flows.

7.2.2 Assessment of ENSAT-PMT Data Quality Exchange

PMT utilises a subset of the ENSAT data and requires different data items in each of the clinical trial phases (as outlined in Table 7-2). Importantly the data items entered into the registry can be used as the basis for determining whether a patient meets the criteria for the study and whether the minimum data required has actually been entered. The study has far greater demands on data quality and data volume. The core data items of patients in the PMT Study covering phase 1-phase 4b comprises around 400 data items. However, it is important to note that these data items often repeat, e.g. multiple measurements are taken repeatedly at specific time points. Any given patient can have over 10,000 individual data items in the PMT Study (where the items in Table 7-1 are typically repeated at each follow up visit)

Table 7-2 PMT data items per phase

Phase	Data items	Data items in ENSAT-Pheo
1	112	82
2	56	12
3	68	50
4a	10	0
4b	154	92

The ENSAT Pheo database itself has 339 core data items. However most of the PMT data items are not transferred or requested by ENSAT PHEO. For example, data items such as follow up examinations (Phase 2 Clonidin test, Phase 4a follow up information and Phase 4b) are not required in ENSAT Pheo.

By calculating data completeness of all data entered for ENSAT PHEO patients, calculation of an overall average 291.8 errors per record, where here error is given by data that is not completed. ENSAT Pheo has 339 data items, which leads to an average data completeness of 14%. This data will subsequently be compared with PMT phase 1 patients – noting that all patients (tumour or no tumour) can be enrolled in PMT entered Phase 1. For the PMT Phase 1 patients, an overall average error rate of 60.1 errors per record (based on missing data) was calculated. Based on a total of 112 data items, this leads to a data completeness rate of 46% (see Table 7-3). It is noted that such errors are calculated and not entered information.

Table 7-3 Data completeness of Pheo Datasets

Dataset	Avg. errors per record	Data items per record	Data completeness
ENSAT Pheo	291.8	339	0.14
PMT Phase 1	60.1	112	0.46

An evaluation of just PMT records in ENSAT Pheo has no impact on the calculation, however the impact of the involvement of a centre in the study needs to be considered. It is also important to consider those patients that meet the inclusion criteria for PMT (or not). The data collection and data entry (data processes) will be the same for both scenarios, i.e. as discussed the ENSAT system is largely based on a best effort data entry effort. It is presumed that a comparison of centres that are participating versus centres that are currently not participating in PMT will have an impact on the evaluation and quality of the related data in the registry (see Table 7-4).

Table 7-4 Data completeness impact of PMT

ENSAT Pheo Dataset	Avg. errors per record	Records	Data completeness
Centres participating in PMT	292.6	1026	0.137
Centres not participating in PMT	291.6	1765	0.140
Principal investigator centre (GYDR)	299.94	50	0.115
Not participating (GBLK)	223.33	30	0.341
Not participating (SPMA)	234.41	165	0.309
Participating (NLNI)	260.38	245	0.232

Centres with the highest completeness rates and more than 10 records are reported for comparison purposes. The results show that the PMT study is not improving the overall data completeness of the common (overlapping) ENSAT Pheo data subset. However, the 12 PMT participating centres out of 60 overall contribute an average of 85.5 records each compared to 36.7 records for other non-participating centres. Results of both the ENSAT and INPDR registries showed that the more records

that are entered, the lower the overall data completeness. Furthermore, the difference between 13.7% data completeness and 14% data completeness is not significant (p=0.90). There is thus no direct evidence that PMT is improving or not improving ENSAT Pheo.

In a new assessment of PMT inclusion criteria, similar problems of data transfer appear. Specifically, 154 records out of 2149 (=7.1%) were not transferred due to inclusion criteria from ENSAT Pheo to PMT (marked red in Table 7-5). Also four inclusion criteria due to language issues (in Dutch) were discovered in the assessment that should be harmonized with the trial language English (see Table 7-5).

Table 7-5 Distribution of PMT inclusion criteria

Inclusion criteria	Total
doorverwijzing met feochromocytoom mgl maligne	1
doorwijzing feochromocytoom	1
Family History positive	1
gemetastaseerd paraganglioom	1
known malignant pheo	1
paraganglioom verwijderd in abdomen, onbekende mutatie, genetisch onderzoek loopt	1
Right adrenal tumor found in the USG and CT, signs and symptoms present	1
Routine	1
several glomus tumors	1
Excluded	6
Routine screening due to known mutation or hereditary syndrome	109
--	154
Routine screening due to previous history of pheochromocytoma	181
Incidental finding on imaging for unrelated condition	422
Therapy resistant hypertension	473
Suspicion based primarily on signs and symptoms	795
Total	2149

This leads to the conclusion that transferrals from ENSAT Pheo to the PMT study have a 7% error rate with regards to inclusion criteria. The inclusion criteria is a mandatory criteria when transferring records from ENSAT Pheo to the PMT study. GYWU has 91 inclusion criteria from 317 entered patients (=29%). All other trial critical defined data items (e.g. biochemical assessments) could be correctly transferred. Nevertheless, the inclusion criteria are important variables for the final stratification.

7.3 EURINE-ACT

7.3.1 Background

EURINE-ACT is an international prospective study that examines the diagnostic and prognostic value of urine steroidobolomics in patients with adrenocortical tumours (ACC). Urine steroidobolomics is the combination of steroid profiling by gas chromatography/mass spectrometry (GC/MS) followed by data analysis by machine learning analysis. The focus of the work is to differentiate between benign and malignant adrenocortical tumours (ACC) by a simple urine test should have a higher specificity and comparable sensitivity than existing imaging techniques.

The EURINE-ACT data flow is shown in Figure 7-3. The study aims to recruit patients with any adrenal mass >1cm. It requires entry of the clinical information into the appropriate ENSAT database (ENSAT ACC for adrenocortical carcinoma and ENSAT NAPACA for any other adrenal mass/adrenal

incidentaloma). The required biomaterial consists of a 24-h urine, a spot urine, a serum and a heparin plasma sample.

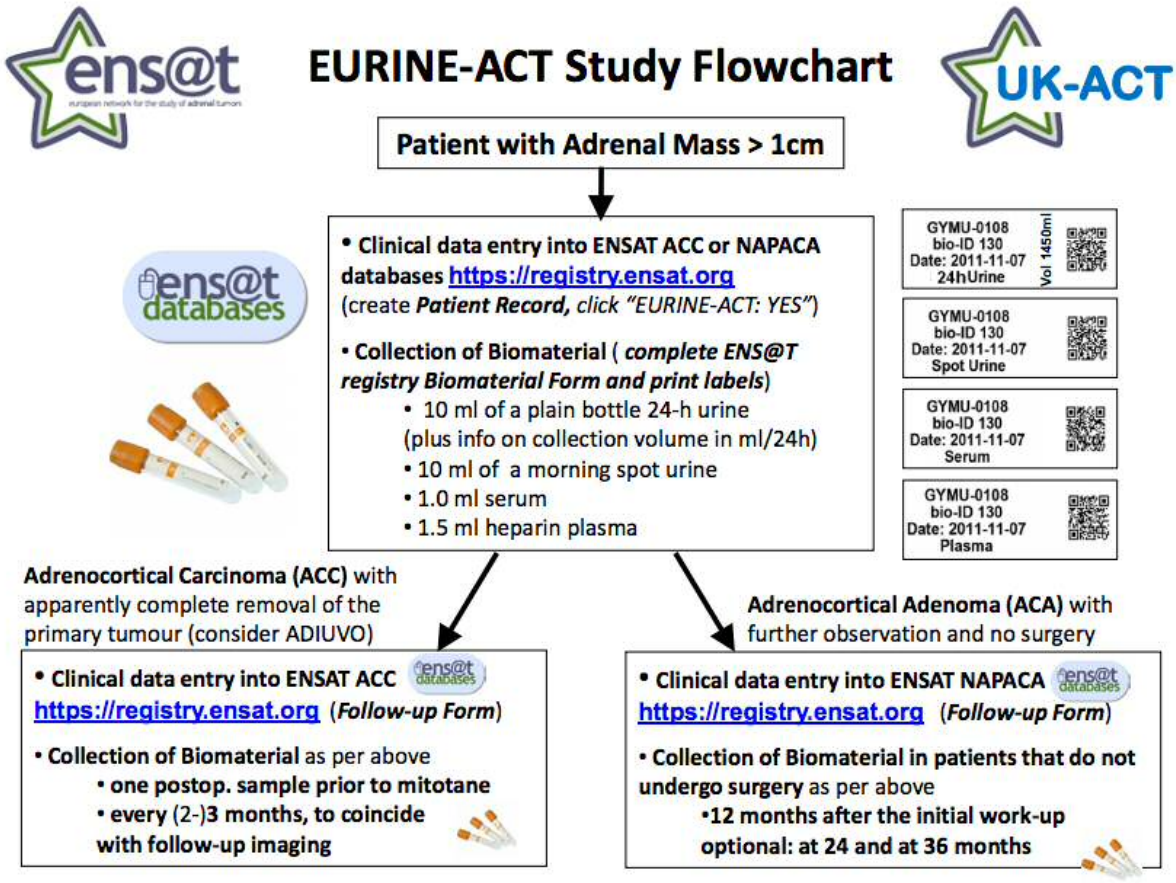


Figure 7-3 EURINE-ACT Flowchart

Patients with confirmed ACC can participate in the EURINE-ACT ACC follow-up arm that offers urine steroidmetabolomics every 3 months following apparent complete surgical removal of the primary tumour. This arm aims to assess the sensitivity of the approach for detection of recurrence in comparison to imaging and also allows analysis of steroid changes induced by mitotane treatment.

Patients with an adrenal incidentaloma that are thought to be an endocrine inactive adrenocortical adenoma can participate in the EURINE-ACT ACA follow-up arm, with biomaterial required 6 and 12 months after initial diagnosis and then annually. This arm aims to detect whether the adrenal incidentaloma shows evidence of potentially deleterious excess hormone production during the first three years after initial diagnosis. All patients with active ENSAT ethics or active local biobank ethics can participate.

Patients enrolled in EURINE-ACT are directly entered into the ENSAT registry and not into separate eCRFs like PMT patients. Compared to PMT, patients with a confirmed adrenocortical cancer are enrolled into the study protocol. In PMT, patients with therapy resistant hypertension or suspected adrenal tumour (the majority of cases are not confirmed during the study phases) can be enrolled into the trial. EURINE-ACT patients with a confirmed adrenocortical tumour are seen in an outpatient setting where they are consented and enrolled into the trial. Data is typically collected through a study assistant, study nurse or clinician. There is no treatment advantage (or disadvantage) for any patient participating in EURINE-ACT (or indeed in PMT).

7.3.2 Assessment of ENSAT-EURINE-ACT Data Quality Exchange

Similar to PMT, EURINE-ACT uses a subset of the ENSAT registry data. Specifically, the study requires 64 ACC related data items from the 142 ACC data items that are available. It is noted that the link between EURINE-ACT and ACC has direct advantages over PMT, because EURINE-ACT is not using external eCRFs.

The results show that EURINE-ACT centres have a 3% higher data completeness level over centres that are not involved, which is not a significant difference (p=0.5). Despite the statistical insignificance, EURINE-ACT centres improve the ACC data sets with on average 3 errors less per record. In comparison to the best centres with the highest data completeness (for centres entering more than 10 records), the centres participating in EURINE-ACT have around 5 errors less per record entered into the ENSAT ACC data subset (see Table 7-6).

Table 7-6 Data completeness impact of EURINE-ACT ACC

ENSAT ACC Dataset	Avg. errors per record	Records	Data completeness
Centres participating in EURINE-ACT	84.06	1940	0.41
Centres not participating in EURINE-ACT	87.49	664	0.38
Principal investigator centre (GBBI)	96.33	63	0.32
Participating (FRBO)	59.17	42	0.58
Not Participating (FRNA)	64.13	16	0.54
France			
Participating	64.92	207	0.54
Not participating	74.03	480	0.48
Germany			
Participating	86.49	1155	0.39
Not participating	119	2	0.16
Italy			
Participating	82.97	328	0.42
Not participating	100.90	75	0.29
Great Britain			
Participating	91.67	65	0.35
Not participating	101.93	71	0.28

It was observed that the majority of centres with higher data completeness were situated in France. Therefore, a stratified data completeness analysis of all ENSAT founding countries (with the greatest involvement and experience) and the country of the principal investigator (GBBI – Birmingham, Great Britain) was conducted. The results show that in every country, the centre participating in EURINE-ACT achieved higher data completeness levels than non-participating centres. Furthermore, in France EURINE-ACT centres have 10 errors less per record, Italy – 18 errors less per records and Great Britain, 10 errors less per record. The results of Germany can't be used to lead to any significant conclusion, because the non-participating centres only entered two ENSAT ACC records into the ENSAT registry. All German centres are highly motivated to participate in research and have been involved in ENSAT from the outset. In summary, we can conclude that EURINE-ACT is improving the ENSAT ACC data subset.

7.4 Data Trustworthiness

7.4.1 Background

Health informatics is the connection between biomedical science and information science. As explained in chapter 2, major subspecialties are clinical research informatics and translational informatics. Another important area of research is the focus on eHealth including spatial research. We consider this here since it provides a relevant example of data trustworthiness. Like every other field of science, spatial research relies on the quality of source data. One widely available form of spatial data is social media, e.g. Twitter. Web developers can use the Twitter APIs to receive data from (tweeted by) users and this can often include the geo-location of the tweet. Users of Facebook and Facebook messenger can also activate the option to share their location with others. Other open source web application like wunderground[24] provide a weather forecasting app that provides users with the possibility to share their personal weather station data with others to improve the weather forecasting around their geo-region. However, in such scenarios, data can be wrong or misleading either deliberately or accidently and this phenomenon gives rise to the question: how much data can be trusted?

As an example of this, Nurse, Agrafiotis, Goldsmith, Creese, & Lamberts (2014) developed a trustworthiness score related to social media information. In a literature review they put forward that trustworthiness should consist of four categories that have to be weighted appropriately (see Table 7-7).

Table 7-7 Categories of Data Trustworthiness (adapted from Nurse et al., 2014)

Category	Weight
Competence of a source (i.e., they have been found not to be very competent)	x5
Information recency (i.e. the information is up to date)	x10
Proximity (i.e. the source is physically close to the event of interest)	x20
Integrity (e.g. there may be concerns about the integrity of the information such as the origins of the data and the infrastructure where it is made available)	x10

The overall calculated scored is called data trustworthiness. The authors suggest to communicate this score in traffic lights (red, yellow and green) and in radar graphs as illustrated in Figure 7-4 (Nurse et al., 2014).

[24] www.wunderground.com

Figure 7-4 Example of data trustworthiness of social media data (Nurse et al., 2014)

7.4.2 ENSAT Data Trustworthiness

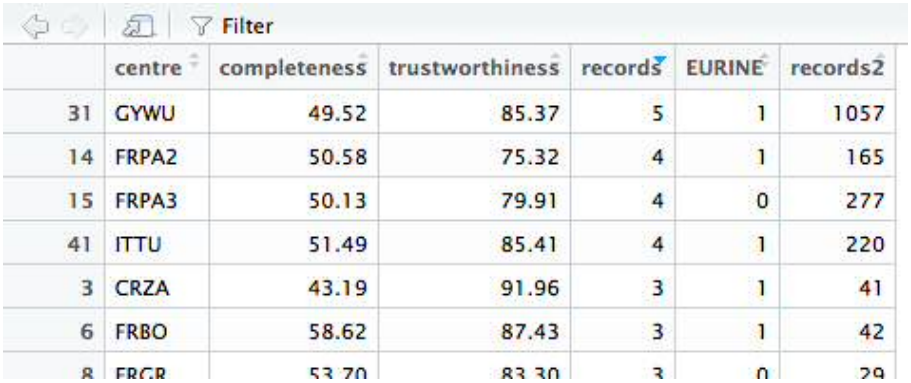
Data trustworthiness is a data quality dimension that has great impact especially in the clinical domain and for translational science. After the exploration of data eligibility scores presented previously, data trustworthiness is a second novel dimension that should be explored. Trusting data entered by potentially non-trusted personnel, e.g. collaborators involved in ENSAT places requirements on the data analysis. One example of such potentially untrustworthy data is from patients (or individuals more generally) that download mobile applications for their own data entry. One such mobile application - NAPACA app has been developed for ENSAT patients with focus on collecting data based on their quality of life. The app is available on the Android and iOS platforms. However, unlike many open mobile applications that can be used by anyone, the NAPACA app requires an activation code that is generated through the ENSAT server by the clinicians responsible for the patients. Hence only those patients that are being seen by ENSAT doctors can activate and subsequently use the app. This is unlike many other crowd sourcing efforts and social media activities, where there is no policing of the users or the data/tweets that they send.

Within ENSAT it is assumed that through the process of a trial and the ongoing data quality assurance methods that are applied (investigator communication, data quality feedback, staff training and protocol amendments), the quality of data and the data completeness will improve over time. It might be considered that data entered at the beginning of a trial has more inconsistencies than data that is entered at the end of trial.

The ENSAT registry captures the date when a patient was registered for research purposes by a contributing clinical centre. However, it is noted that this registration might well have been before the development of the ENSAT registry in 2010. Given this it might be suspected that older data are assumed as having lower trustworthiness than more recently registered patients. We consider this hypothesis in more detail through exploring the trustworthiness and completeness of the ENSAT data within the PMT and EURIN-ACT studies. Specifically, and for simplicity we assume a linear scale for data trustworthiness based on the date of entry of the data. To support the analysis and the visualisation, the source data and R script are given for both EURINE-ACT and ENSAT ACC respectively

in Appendix III. The trustworthiness of PMT is measured in a similar to EURINE-ACT but Pheo-related source datasets are used.

The final graph plots the data completeness of centres with their associated data trustworthiness. To show the different record contributions in ENSAT ACC, each centre will have a plot in relation to the entire set of patient records that have been contributed to the registry. Therefore, the source data is split into several dimensions: the centre; the completeness score; the trustworthiness score; the record rank for the visualization; whether the centre participates in the related trial, and the overall records for weighting the final calculation (see Figure 7-5).


Filter

	centre	completeness	trustworthiness	records	EURINE	records2
31	GYWU	49.52	85.37	5	1	1057
14	FRPA2	50.58	75.32	4	1	165
15	FRPA3	50.13	79.91	4	0	277
41	ITTU	51.49	85.41	4	1	220
3	CRZA	43.19	91.96	3	1	41
6	FRBO	58.62	87.43	3	1	42
8	FRGR	53.70	83.30	3	0	29

Figure 7-5 Calculation of ACC data trustworthiness (based on time of entry)

The column 'completeness' is a percentage value that represents the average entered values of the entire ENSAT ACC dataset. The 'trustworthiness' column is also a percentage value and is calculated as an average of all centres records based on the age of the records. This calculation was performed by assessing the difference between the record registration date and the data trustworthiness calculation. All records of the centre (displayed in 'records2') are assigned in 5 different categories. Centres with more than 300 records are assigned in the category 5 (n=1), centres with 300-100 records are assigned into category 4 (n=3), centres with 100-40 records assigned into category 3 (n=14), centres with 40-10 records into category 2 (n=11) and centres with less than 10 records into category 1 (n=25). This category is displayed in column 'records' and is only used for visualization purposes. The column 'EURINE' is for those centres participating in the EURINE-ACT study (value = 1) or not (value = 0). Not every record from a centre participating in EURINE-ACT or PMT is assigned to EURINE-ACT or PMT, however it is assumed that the data processes of non-trial records are similar to those of trial records.

For the visualization of the centres, the package RColorBrewer was used to allow larger centres to be visualised with larger circles and hence to better understand the data analysis. The code to achieve this analysis is shown here.

```
# color setup
library(RColorBrewer)
display.brewer.all()
# colorplate
somecols <- brewer.pal(5, "Set1")
red <- col2rgb(somecols[1])
blue <- col2rgb(somecols[2]) display.brewer.all()
# transparency for dot plots
Transred <- rgb(red = red["red",],
                green = red["green",],
                blue = red["blue",],
```

```
                    maxColorValue = 255, alpha = 60)
transblue <- rgb(red = blue["red",],
                    green = blue["green",],
                    blue = blue["blue",],
                    maxColorValue = 255, alpha = 60)
dat$colors <- ifelse(dat$EURINE == 1, transred, transblue)
# relative size of the plotting symbol, color in new column
dat$cex <- dat$records/max(dat$records)
oldpar <- par()
```

Using this code, we plot an ellipse for the overall visualisation. This ellipse displays the 95% confidence interval of centres that participate in EURINE-ACT and centres that do not participate. This ellipse is weighted by the overall records (see dataEllipse – weights command below).

```
layout(matrix(c(2,2,2,2,1,1,1,1,1,1,1, 1,1,1,1, 1,1,1,1,1, 1,1,1,1), nrow = 6, byrow = T))
library(car)
par(mar = c(5,5,1,0))
plot(0,0,type = "n", ylim = c(-10,40), xlim = c(60,105),
     ylab = "completeness in %",
     xlab = "trustworthiness in %",
     main = "",cex.axis=2, las = 1, cex.lab = 2, cex.main = 2,
     bty = "n", yaxs = "i", xaxs = "i")
# the maximum size of a plotting symbol is 8
points( dat$trustworthiness, dat$completeness, cex = 10*dat$cex,
     col = dat$colors, pch = 16)
with(subset(dat, EURINE == 1),
     dataEllipse(trustworthiness,completeness, weights = records2, levels = 0.95, add = T,
     center.pch = NULL, plot.points = F,
     col = red))
with(subset(dat, EURINE == 0),
    dataEllipse(trustworthiness,completeness, weights = records2,
    levels = 0.95, add = T, center.pch = NULL, plot.points = F,
    col = 'blue'))
```

The last part of this code segment is the labelling of the graph, which is not further explained, since is it redundant for the data quality aspects explored here. The final graph for ENSAT ACC and EURINE-ACT is shown in Figure 7-6 and Table 7-8.

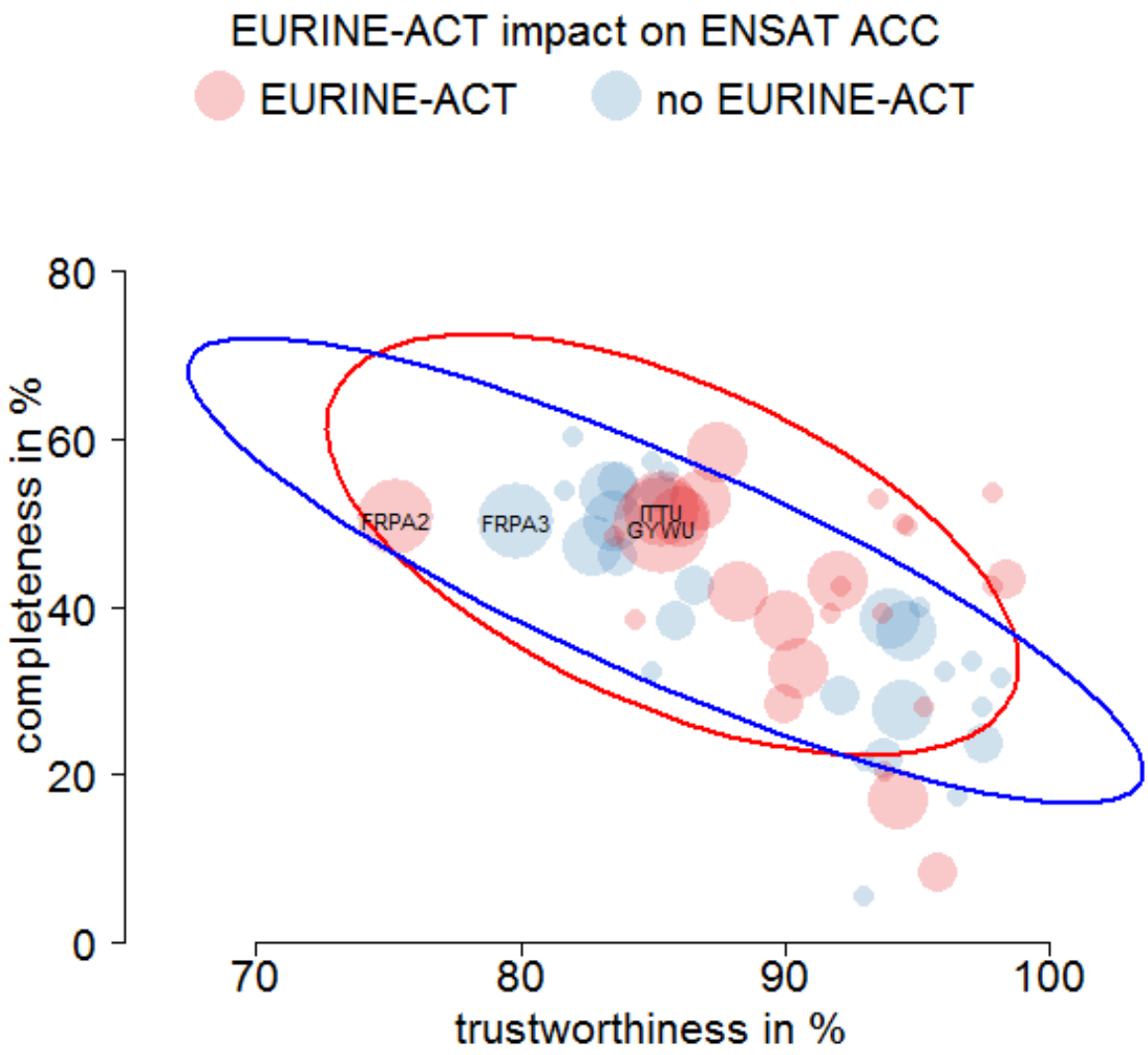


Figure 7-6 Visualisation of ENSAT ACC Data Trustworthiness (based on time of entry)

Table 7-8 EURINE ACT Trustworthiness Regression estimates in a linear regression model (weighted)

	(Intercept)	(slope)	p (EURINE-ACT vs no-EURINE-ACT)
EURINE ACT	157.40	-1.32	
no-EURINE ACT	140.16	-1.08	0.52

The size of the ellipse illustrates a non-visible difference between the centres involved in EURINE-ACT and those centres not involved in EURINE-ACT. The graph is expected the show that both ellipses are not covering themselves (at least by the majority of their area) and the data completeness has a positive linear correlation with data trustworthiness. The results in Table 7-8 show that there is no significant difference in data completeness and data trustworthiness for centres participating in EURINE-ACT compared to those centres that are not involved in EURINE-ACT. This is caused by large centres (ITTU, FRPA2 or GYWU) who provide a large proportion of 'older' data with an average overall data completeness level (≈ 55%) versus newer and typically smaller participating centres that provide lower data completeness levels.

The same analysis was performed for both PMT and ENSAT Pheo and is shown in Figure 7-7 and Table 7-9.

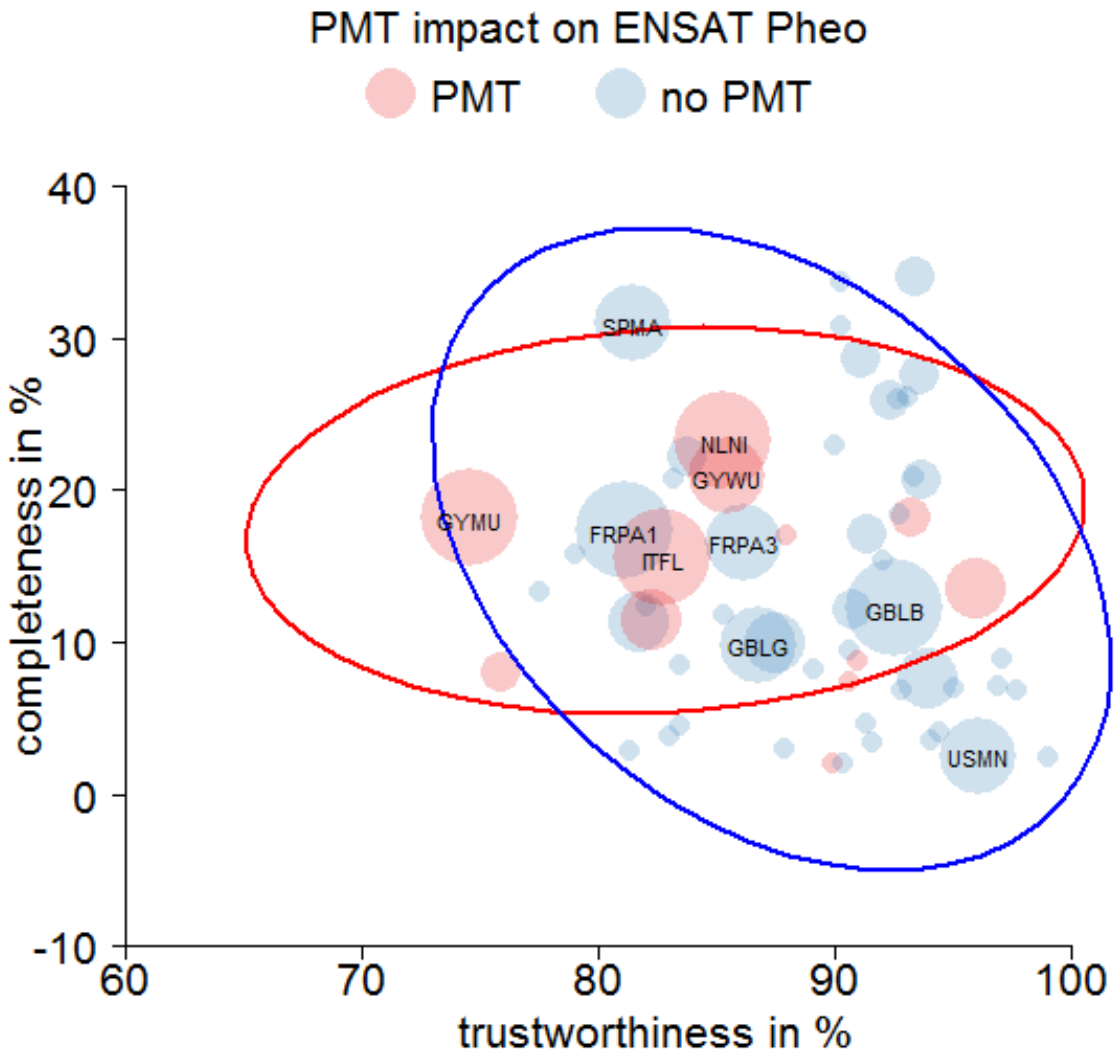


Figure 7-7 ENSAT Pheo Data Trustworthiness (based on time of entry)

Table 7-9 PMT Trustworthiness Regression estimates in a linear regression model (weighted)

	(Intercept)	(slope)	p (PMT vs no-PMT)
PMT	11.89	0.07	
no-PMT	61.65	-0.52	0.067

In contrast to EURINE-ACT and ACC, the weighted graph shows a positive correlation for those centres participating in PMT compared to those centres not participating PMT with regards to data trustworthiness and data completeness. This confirms the hypothesis that new centres with newer data provide more complete information (of Pheo datasets) in the ENSAT-registry. It is also noted that the ellipse is narrower for non-participating centres, which implies that this assumption is more valid. In general, the overall data completeness is low, which indicates that the included data items in the registry should be discussed in the ENSAT working group meetings, and that very specific data is needed for research purposes, which is potentially not needed for other trials.

To explore this further, a more detailed investigation of the assumption that there is a positive correlation between data completeness and data trustworthiness can be achieved through a linear regression analysis per record. The results of the linear regressions for the Pheo and ACC data are shown in Figure 7-8 and Figure 7-9, linear regressing values displayed in these figures are given in Table 7-10.

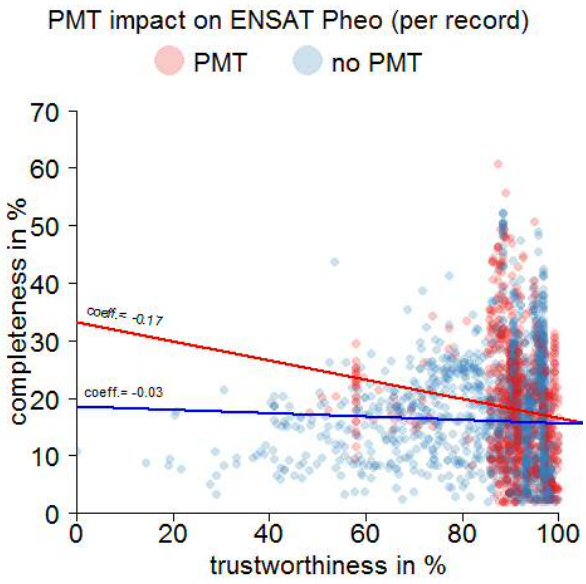


Figure 7-8 Linear regression ENSAT Pheo

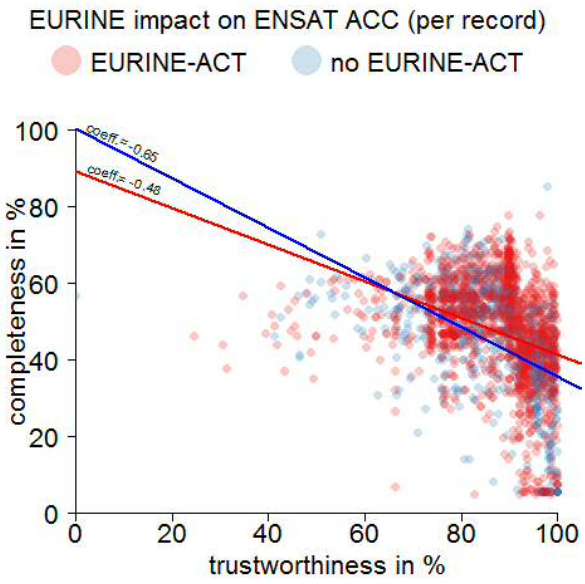


Figure 7-9 Linear regression ENSAT ACC

Table 7-10 Linear regression values for single centre data trustworthiness comparison

	(Intercept)	(slope)
PMT	0.33	-0.17
no-PMT	0.19	-0.03
EURINE-ACT	0.89	-0.48
No-EURINE-ACT	1.00	-0.65

The linear regression shows a negative correlation coefficient in all cases. The slope of PMT, no PMT, EURINE-ACT and no-EURINE-ACT records shows a negative slope, which indicates that there is decreasing completeness with higher levels of data trustworthiness. This indicates that the data of the records do indeed have an impact on the ENSAT-registry, however new records (higher trustworthiness) have lower data completeness. This may be caused by too many redundant data items that are used for research purposes.

Furthermore, in ENSAT Pheo there are a large proportion of records in the region of 20% data completeness and 95% data trustworthiness. Inspecting only this area (trustworthiness 90-100%) a slope is seen for data completeness. The reason for this might be caused by an 'information delay'. In other words, the records are created with baseline information of the patient. After a certain time, laboratory information, pathology information, follow up information is entered and the records are gradually made more complete. The data shows that this has indeed been the case since 2011. The same can be confirmed for ENSAT ACC, where a large proportion of records show a data completeness drop between 90% and 100% with regards to data trustworthiness.

7.4.3 Conclusions on Data Trustworthiness

The assumption that more recently entered can be more trusted needs to be reconsidered. Factors such as the discovered information delay arise in ENSAT ACC and ENSAT Pheo and need to be considered. The assumption that data trustworthiness has a positive correlation on data completeness also needs further evaluation in other situations and with other registries. It is a logical assumption but the analysed data could not confirm this to actually be the case.

In conclusion, data trustworthiness as a data quality dimension needs more attention, especially in clinical research settings. Data quality procedures improve data and the data can be more trusted and subsequently used in other research purposes, e.g. in other clinical trials. The performed analysis is a novel approach to investigate the relationship and correlations between two quality dimensions. There is a need to investigate dynamics between defined quality dimensions for clinical research. Data completeness is a well-researched data dimension for record eligibility for example, whilst data trustworthiness dimensions will have a growing impact in the era of mobile health and other citizen science data-oriented activities.

In summary with regards to ENSAT, data older than 10 years (lower trustworthiness) is less complete, whilst data from 2011/2012 is more complete. Conversely we find that more recent data (higher trustworthiness) is also less complete. Specifically, we observe that data has most 'usefulness' for research purposes, when it is around 3-4 years old, i.e. it was entered 1-2 years after the establishment of the ENSAT registry.

7.5 Chapter Conclusions

Data standards are an important tool for data exchange, for research interests and to achieve system interoperability more generally. So-called data warehouses must be built leveraging established standards. CDISC and Hl7 provide several modules for different study intentions. International platforms and research networks suggest to use data standards to improve data exchange between electronic health records or hospital health records and electronic case report forms (eCRFs). Due to policy and security barriers, ENSAT is not using electronic data captured from hospital records into the ENSAT registry. Manual data entry is the used method to complete eCRFs with standardised information. Here, no general data standard was used, but a standard that matches the intensions and purposes of the international network of adrenal tumour researchers.

In this chapter we considered the research question as to whether PMT or EURINE-ACT improved the ENSAT registry with regards to data completeness of the overlapping data due to extrinsic motivational factors, i.e. by being involved in the trials. This was found to be the case for EURINE-ACT but not so for the PMT study. One reason for this result is, that EURINE-ACT is using the same data elements as the ENSAT registry and does not have its own eCRFs for data entry – thus its success depends on the quality of the data in the registry. It should be noted that these calculations were performed with just one type of tumour: pheochromocytoma for PMT and adrenocortical carcinoma for EURINE-ACT. Other ENSAT tumour types (NAPACA, APA) were not explored. The results suggest that data transfers to study-specific eCRFs do not in themselves lead to improvements in data quality of records in centralised registries. On the other hand, PMT is enrolling many patients (as controls) that do not have an adrenal tumour and these kinds of patients cannot be captured within the ENSAT registry.

8 CONCLUSIONS AND FUTURE WORK

High data quality is essential for translational research, especially in the clinical domain. The evaluation of data quality in clinical research is needed to improve scientific conclusions. It is also essential that data quality assessment is undertaken throughout the data collection and analysis phases of clinical research, and where required improve the behaviour of data collection personnel and their data entry practices. The intention of this thesis was to provide a framework realised through an automated web-based application that provided data quality feedback for data entry users throughout the lifecycle of data collection and future use. As part of this aim an exploration of data quality dimensions and data entry motivation was conducted. This was undertaken in a range of national and international disease registries and included the patients themselves performing data entry through targeted mobile applications.

The following chapter summarizes the primary and secondary aims of this thesis and provides an overview of possible future research that may be followed to improve data quality in clinical research settings.

8.1 Conclusions

This research has outlined the necessity for data quality in clinical research and especially for its translation into a clinical setting. We explored the state of the art approaches that have been adopted in the area of data quality with specific focus on clinical research demands. We articulated why clinical research especially relies on data quality based on the specific demands for accuracy and challenges in collecting data from patients, e.g. during patient visits or from medical devices. We discussed how data might be entered manually by data entry personnel or electronically through software systems leveraging pre-existing (digitised) clinical data. Both methods have their own advantages and disadvantages: manual data entry is more realistic at present given the heterogeneity of IT systems and standards for data that currently exists across the healthcare sector both nationally and internationally. It is envisaged that this situation will eventually be overcome through continued adoption of IT within the clinical sector. One example of this is the growing research field of mobile health applications. As US President Thomas Jefferson once said: “Information is the currency of democracy” (The Jefferson Monticello, 2016). In the internet age, a refinement of this quote is applicable: ”Information is the currency of the digital age” (Nurse, Agrafiotis, Goldsmith, Creese, & Lamberts, 2014). However, more information in itself does not necessarily indicate that more insight (knowledge) can be garnered. Whilst big data promises much for the future through data mining and advanced data analytics, it is still the case that clinical and biomedical research demands high quality data that is difficult to access and use due to the sensitivities that surround it, e.g. need for confidentiality and it’s heterogeneous ‘at source’ nature. In a clinical setting, it should be the case that “Information that matches the need of the user is the currency of the digital age”. This would imply that information is standardised and organised such that knowledge can be derived from it directly. As discussed in this thesis, this is still a fraught issue for breakthroughs in biomedical research due to the challenges in establishing sufficiently robust and accurate data repositories and/or providing live access to high quality data sets at source. Biomedical data registries represent one approach to overcome this. Such registries offer ways to establish targeted sets of data relevant to specific disease areas that support networks and communities of biomedical researchers. Such data resources can underpin clinical trials and biomedical studies more generally, but this demands that the data is of high quality.

During the exploration of the research hypothesis: “web-based feedback should be an integral part of data collection processes in disease registries that demonstrably improves the overall quality of data that is entered into such registries”, three key aspects were identified and explored in the thesis:

- how to measure data quality (data quality dimensions),
- how to improve data quality (user motivation), and
- how feedback can "be an integral part of data collection processes" (categories of feedback).

These aspects are concluded in the following chapters.

8.1.1 Exploration of Data Quality Dimensions

For the evaluation of data quality in clinical and biomedical research, the definition of relevant data quality dimensions is crucial. There are standard quality dimensions that are often measured in the majority of clinical trials. As discussed, these typically include data validity checks related to data completeness, data accuracy and data plausibility.

This research covered in this thesis explored the five most used data quality dimension of clinical research (Weiskopf & Weng, 2013):

- Data completeness;
- Data correctness (here accuracy);
- Data concordance (here comparability);
- Data plausibility (here trustworthiness), and
- Data currency (here timeliness).

We discussed the recommendations and findings of each of these data quality dimensions in this research through case studies involving a range of international clinical disease registries and clinical trials. We summarise these findings here.

Data completeness

Data completeness is the most commonly used dimension for data quality evaluation. It is calculated based on the ratio of entered/given items of a (requested) dataset with all items of a dataset. Often these requested and given are based on web-based forms for data collection. NIH collaborators distinguish data completeness with overall completeness through 'column', 'row' and 'ascertainment' (record eligibility) completeness (Zozus et al., 2014).

Within this thesis it was highlighted, that data completeness needs to be adjusted depending on whether data is compulsory or optional for particular clinical studies. Making all data compulsory is typically unworkable in most settings: data items may not exist at a given time, and data entry personnel may be unwilling or unable to provide all such data. A more typical approach is to mix compulsory data with optional data. In this situation the predominant approach for data collection is based on best effort. The calculation of data completeness provides a general procedure that can be used to provide a generic form of data quality assurance in health based data registries. Literature suggests that percentage values of >90% for data completeness are recommended (Mathers et al., 2005; Nonnemacher et al., 2014). Data completeness is generally reported as percentage values in the literature (Larsen et al., 2009). For the visualisation of data completeness, tables with percentage values and given (Hills et al., 2010; Hirdes et al., 2013; Porgo, Moore, & Tardif, 2016) or tables with arithmetic means (Köpcke, Trinczek, et al., 2013). Data completeness can also be visualized as bar charts (Herzberg et al., 2011) or as process diagrams with percentage values (Arts et al., 2002). Timmermans et al. (2016) also suggested using p-value dot plots to visualize data inconsistencies (see chapter 4.3). This thesis also suggests to weight every data dimension and to set a threshold, but these demand could not be explored within this research.

For this work, data completeness assessment provided the baseline information for the research hypothesis. The first data completeness evaluation was performed for the PMT-Study where all records for Phase 1 patients were analysed. Here data completeness was used to detect data items

that were not entered into the ENSAT-CANCER registry, but needed for the research hypothesis. In PMT data completeness was assessed as part of data validity, where the completeness was combined with data accuracy and plausibility checks. In the EURINE-ACT study, a separate data completeness evaluation was performed, since the dataset was incomplete and could not be used directly to assess accuracy and plausibility. In this case, data completeness was used as feedback to researchers to help to increase their motivation to enter more data. The resultant bar charts visualising the data completeness scores before feedback are presented in Figure 8-1 & Figure 8-2.

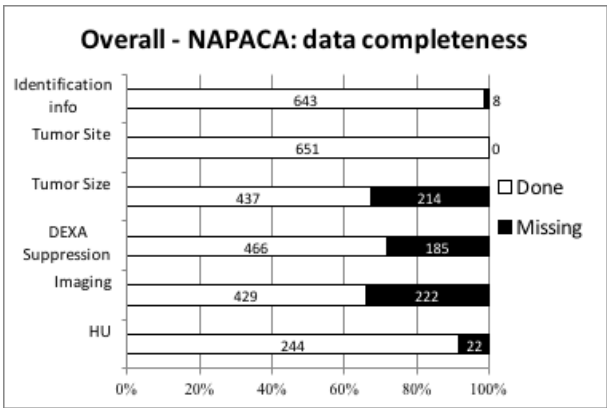


Figure 8-1 EURINE-ACT NAPACA Data completeness visualisation example I

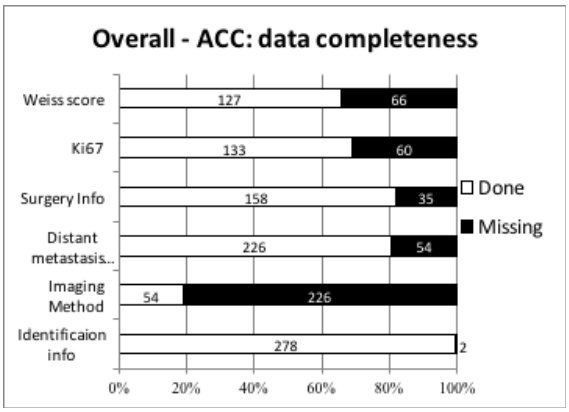


Figure 8-2 EURINE-ACT ACC Data completeness visualisation example II

Data completeness was also reported in tables as percentage values with errors per record also provided, where an error represents a non-entered value. In the literature data completeness also referred as data accessibility, availability, missingness, omission, presence, rate of recording or sensitivity (Weiskopf & Weng, 2013).

Data accuracy

As discussed in this thesis, data accuracy is often combined with data completeness to assess the overall degree of data validity. The completeness level of data in clinical registries is more readily reported than the degree of data accuracy. Data accuracy also has a much more diverse definition in the literature. Accuracy is also defined as the "extent to which the results of a method agree with the independent external citation" (Stone, 1986). Accuracy is often given in terms of possible transfer or transcription errors that arise when moving (transferring) data into a data registry (Datta, Findlay, Kortbeek, & Hameed, 2007). An example of such data accuracy measurements occurs in source data verification based clinical studies.

In this research, data accuracy was defined as the extent to which a data item agrees with the defined range for the item. Data accuracy was assessed in the PMT study, where every item of phase 1 patients was assigned with a range that was required to be fulfilled study. This excluded items that were not filled out or where default information was provided, e.g. an item with the possible responses "—", "not done", "no" or "yes" in case the item "—" should be selected since otherwise it is assumed that no data entry was performed and hence the item is inaccurate (incorrect). A similar procedure was applied to EURINE-ACT for the data completeness calculation. The reason is that "—" reflects non-entered information and therefore where such a moot response is given, this corresponds to an incomplete data item in the dataset.

As an extension (sub-category) of data accuracy, record eligibility was developed for the EURINE-ACT study. This quality dimension has the potential to be applied to other clinical trials. Records entered by the data entry personnel need to meet defined minimum criteria to be eligible for the final analysis and ultimately to be useful for research. The report of eligibility during a trial (e.g. a cohort study)

displays important information about possible issues in the data process. With regards to the EURINE-ACT study, these issues included items such as pathology and surgery reporting problems (see Figure 8-3).

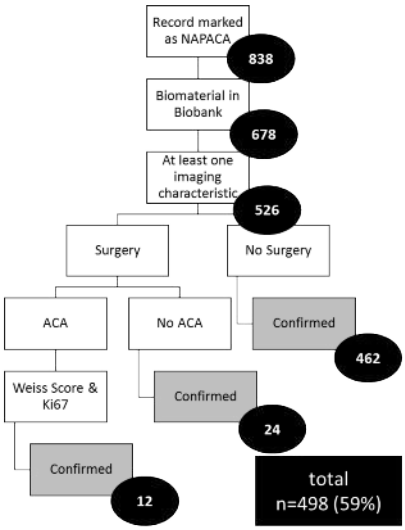


Figure 8-3 Example of Record Eligibility Feedback

For EURINE-ACT, record eligibility and data completeness scores were combined to provide an overall data quality score. In a pre-post test study design (without control group) within the EURINE-ACT study, feedback of the record eligibility score increased the overall data quality within 6 months (see chapter 3.4). This was an essential fact that provided the basis for much of the experiments and the research hypothesis that has driven this thesis. Specifically, it was demonstrated that feedback works. In this case the feedback was based on face-face meetings with the data entry personnel. The core thesis we wished to explore here was whether the human face-face feedback loop could be replaced with automated web-based feedback systems.

Data comparability

In disease registries, data comparability is often regarded as the extent of classification procedures of a registry that agrees with international guidelines. One example of this is ICD-10 (Bray & Parkin, 2009). Such coding of data in a registry can allow it to be directly compared with other registries based on the common coding scheme and ability to compare directly. However, clinical registries often cannot be compared with other registers since the research purposes are diverse, and importantly the data itself is coding using many different coding systems. Comparison with other registries can be explained as challenges of 'external' comparability, whilst comparison with datasets derived from the same registry, e.g. data used to support particular clinical trials, can be considered as 'internal' comparability.

The optimal epidemiological practise for clinical trials is based on comparison with control groups. One example of such an approach is the PMT trial, which is used to improve pheochromocytoma diagnostics based on the measurement of blood-based metanephrines. Measuring blood-based metanephrines in a healthy (and similar) cohort offers a baseline assessment of what constitutes high and low levels of metanephrines.

In the evaluation of the PMT data quality, data monitoring of blood sampling approaches detected statistical outliers (see Figure 8-4). A 'health' cohort had higher blood metanephrines levels when they did not fast, if they drank coffee, if they cycled or did sport more generally. The importance of standard operating procedures (SOPs) for blood sampling was thus essential, because SOPs forbid these actions

(Därr et al., 2014). Thus data quality and comparability is far reaching and goes beyond the data entered into a database. Rather it can involve the systematic approaches and procedures by which data is collected and analysed in a clinical setting. The graphical visualisation of such performance issues and deviance from the norm is one way that such issues can be addressed. This subsequently implies that clinical staff are subsequently trained to adopt best practices in blood taking. This is one example, but many others are essential for data to be truly comparable.

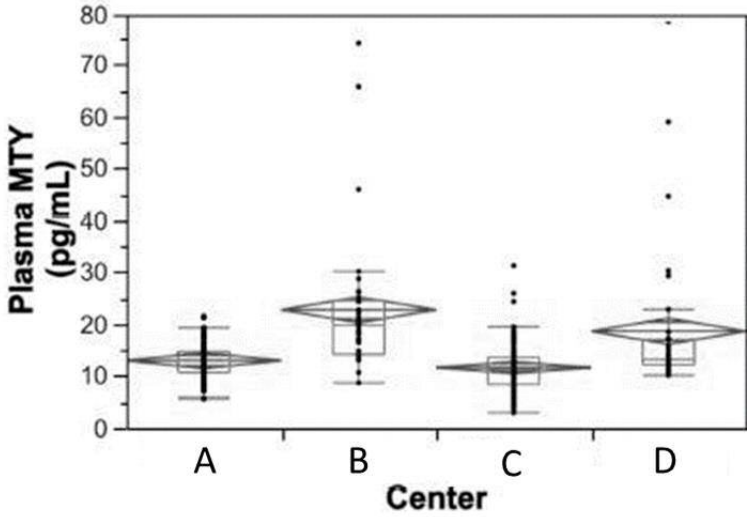


Figure 8-4 Box-Plot chart displaying outliers used to detect data issues in PMT

Another statistical approach for data comparability is the calculation of interclass correlation coefficients (ICC) and their visualisation, e.g. through boxplots. There are several propositions in the literature for such visualisation. Guthrie et al. (2012) display the value in a table where values are used to indicate a performance differences between centres. The display of ICC in boxplots or RICC vs R^2 (Wynants et al., 2013) is another possibility. With regards to investigator feedback, the usage of ICC or RICC vs. R^2 was suggested (see chapter 4.1.1).

Data trustworthiness

Assessing data trustworthiness is often used to quantify the quality of the original source of the data. There is no general definition of data trustworthiness. Questions that data trustworthiness needs to address are: “Where did the data come from? How trustworthy is the original data source? Who handled the data? Are the data managers trustworthy?” (Dai et al., 2008). Alternative terms for data trustworthiness can be data provenance or believability (Weiskopf & Weng, 2013).

For this research the first approach taken was to assess data trustworthiness with regards to data completeness. Here the hypothesis was that data has high trustworthiness when it is recently entered into a research registry or into an electronic case report form (eCRF). It was considered whether older data would have lower trustworthiness on the premise that the data processes were not as established compared to more recent data or where the older data was based on evolving understanding of the data that was to be entered into the registry.

Based on correlation with data completeness for the ENSAT-CANCER registry, it was discovered that data has higher completeness when it is 3-4 years old or when it is entered 1-2 years after the establishment of the registry (see chapter 7.4.2). As an example for the visualization, dot plots of every centre (see Figure 8-5) and for every record can be used (see Figure 8-6).

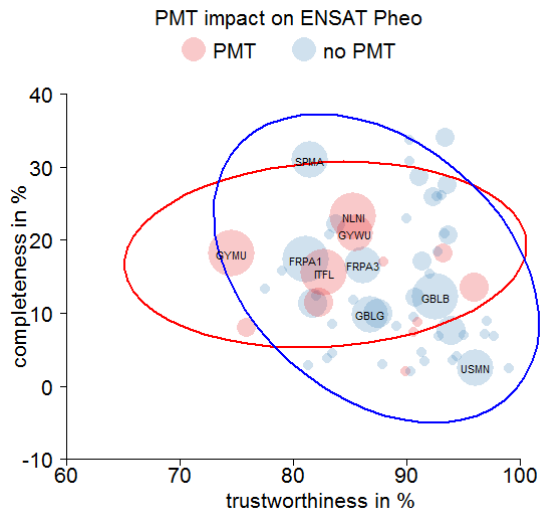


Figure 8-5 Data completeness visualization by centre (example ENSAT Pheo & PMT)

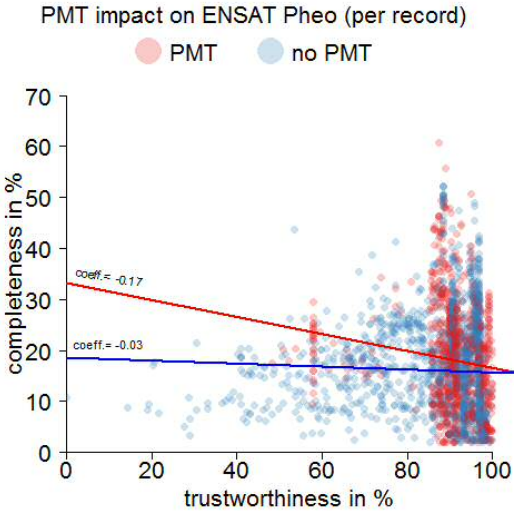


Figure 8-6 Data completeness visualization by record (example ENSAT Pheo & PMT)

Data timeliness

In clinical research, data timeliness is a "benefit to health providers and researchers, and early provision of data usually enhances the reputation of the registry" (Bray & Parkin, 2009). This research defined data timeliness as a core indication of good data entry practice.

To assess data timeliness, timestamps in the data sets that were entered was used. These timestamps were collected through logging tools associated with the data registries such as INPDR and ENSAT-CANCER. Other important dates include the date of the original collection of the data, which is in the majority of cases different to the date of registration and actual data entry. Logging data can also be used to assess the need for staff training purposes and use of registries more generally. This can be used to better understand local processes, and where necessary to improve local data collection and data entry processes.

Another key capability that is often required for clinical research is the visualization of patient timelines. Longitudinal visualisations offer the ability to readily assess patient diagnosis and treatment to improve follow-up and generate feedback that can subsequently be used for advanced patient care. An example of this from the ENSAT-CANCER registry is shown in Figure 8-7.

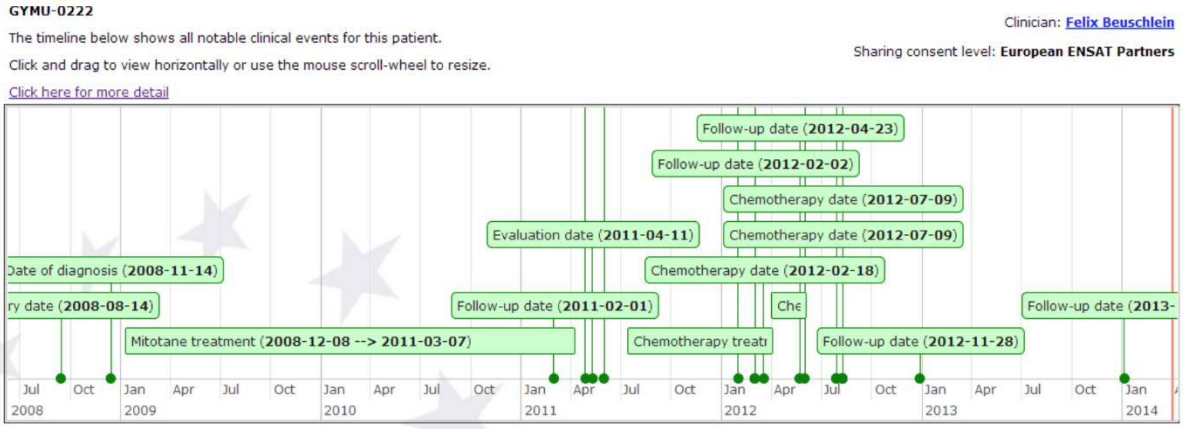


Figure 8-7 Visualisation example of longitudinal data timelines of a patient

Similar analysis can be performed for researcher interactions with a clinical data registry. This can be used for data protection purposes but also for monitoring when data entry personnel have problems entering data. This assessment should be part of an overall data quality feedback process (see Figure 8-8.

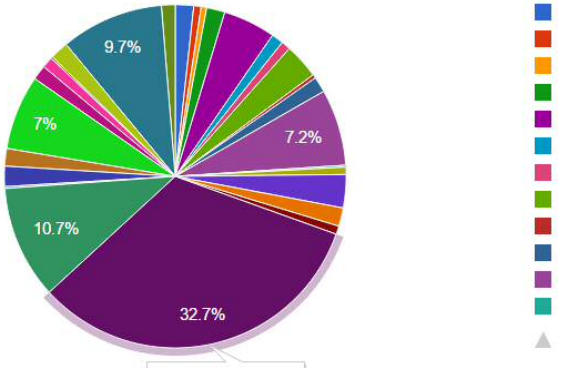


Figure 8-8 Example of ENSAT visualisation of user logging information

Data Quality Conclusions

The data quality dimensions of completeness (all trials), accuracy (EURINE-ACT & PMT), eligibility (EURINE-ACT), comparability and timeliness (PMT) were all explored in this research. Some dimensions need targeted correlations with data completeness, which is the most widely adopted data quality indicator currently used in clinical research. Other quality dimensions could not be statistically tested with other registries. For the INPDR data registry, there were insufficient data entries within the observed period to make any overall conclusions on data quality practices. Other registries where originally considered to be part of the final analyses of this work included registries focused on Polycystic Ovarian Syndrome (PCOS), however the research collaborators failed to utilise these resources (this work was tied to a research grant that was submitted, and whilst the PCOS registry and associated mobile applications were developed, the grant was unsuccessful and the subsequent use of the system did not materialise).

A methodological disadvantage of the overall efficiency of assessing data feedback is that this is typically not based on use of a control group. Thus there are no known data quality measures that can be systematically applied. Instead each disease registry has its own communities and data collection practices that differ between one another. Thus comparing the data quality of ENSAT-CANCER does not immediately transfer to INPDR or PCOS for example. Nevertheless, the process of applying web-based feedback is generic and can be applied in all cases.

As outlined in the introduction (section 1.2.3), the impact of user motivation with an overall data quality score could not be fully explored in this thesis. The reason for this was that only PMT and EURINE-ACT explored data completeness along with a second data quality dimension (for PMT – data accuracy and for EURINE-ACT – data eligibility). For ENDIA and INPDR only data completeness was assessed. This indicates that the impact of an overall data quality score (an arithmetic mean of all calculated data quality dimensions) still has to be investigated with regards to user motivation.

Data quality has an essential overall impact on the final conclusions of any given clinical study or trial. As a suggestion, all data quality dimensions can be given as feedback where single values can be reported in tables or visualised in bar charts, box plots, dot plots, timelines or pie charts. The overall

data quality itself can be reported in radar charts and targeted to specific users or specific centres involved in a given clinical collaboration as shown in Figure 8-9.

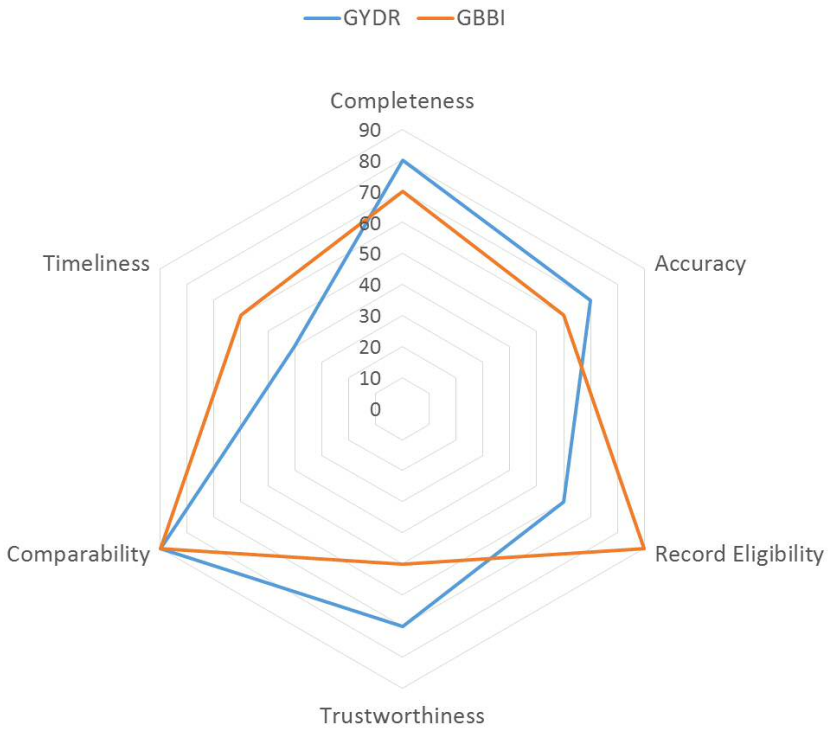


Figure 8-9 Visualisation example of all ENSAT data quality dimensions

High data quality should be the goal of all clinical research collaborators: from data entry personnel through to data users including those beneficiaries involved in associated clinical trials. Data quality feedback can motivate and encourage collaborators to collect higher quality data. We have demonstrated that feedback works. Ultimately higher data quality is an essential cornerstone that provides the bridge between clinical research and clinical care.

8.1.2 Exploration of User Motivation

The second element that has driven the work in this thesis was evaluating user motivation. People have by far the greatest impact on data quality especially based on manual data entry processes (which reflect the vast majority of clinical research registries existing today). Following the literature review, a survey was performed and the responses to this survey were correlated with data completeness for the ENSAT-CANCER registry. It was found that data completeness was directly influenced by the number of other research registries that the (contributing) centres were involved with, and especially where the data entry personnel had to enter data from the same patient multiple times. Furthermore, the time of data entry showed an impact on data completeness rates. The more time spent by the user on data entry, the lower the data completeness – this is somewhat against intuition. It may have a relation with the usability of the system more generally, e.g. it requires a longer time based on unfamiliar user interface to the system for example. Non-significant results of the survey included a negative correlation with the age of the data entry personnel, or more specifically the trend showed that the older the user, the better the data. The survey results showed no relation regarding the time delay to actual data entry (i.e. the time between original data collection and data entry into the registry), the gender, how many trials the centre is participating in, nor the original source data type, e.g. paper or electronic based data forms.

For the motivational question of the survey, users expressed that they have higher motivation when they see the impact of their data entry efforts on the treatment of patients or in supporting

international collaborations. Various tools have been implemented in the ENSAT-CANCER registry to motivate the users and centres to enter more and better quality data. These tools include leader boards and feedback features that dynamically inform the user about their own data quality and how it is tracking over time as well as comparing their data entry efforts with other user/centre data entry efforts. Other possibilities include badges and wards to help motivate better data quality efforts, however these have only been considered indirectly in the work conducted here and by and large, all of the registries are based on best (unrewarded) effort.

For the PMT study, pre-post tests on data quality could not be performed (since the work had already commenced), but in structured interviews during monitoring visits, researchers had positive feedback on feedback features. One feature that was tested was the reporting of data completeness scores at varying levels of feedback. This was explored for INPDR in chapter 5). The study design was developed to distinguish between intrinsic (self-motivation) and extrinsic (external pressure) motivational factors. This included a variety of feedback mechanisms of increasing detail and comparison with the other data collected by collaborators. Unfortunately, no statistically relevant results could be seen, largely due to the limited amount of data entry during the time of observation to the INPDR registry.

It is worth noting that the technological underpinnings of ENSAT-CANCER and INPDR are largely similar and both deal with a rare disease, however the ENSAT-CANCER registry has galvanised global research into adrenal tumours whilst INPDR (at present) only includes the original collaborators involved in the Niemann-Pick grant. It is assumed that the ENSAT network has been key to this. A second major contributing factor is the number of clinical trials that are now supported through the ENSAT-CANCER registry. At present INPDR does not support any clinical trials, although several are being discussed with pharmaceutical companies.

8.1.3 Exploration of web-based Feedback to increase Data Quality in biomedical Registries

The Good Clinical Practise Guidelines (GCP) define two mandatory criteria for clinical trials: to protect human subjects and to collect high quality data (International Conference on Harmonization, 1996). Other important standards and guidelines describe the need for ongoing quality assurance during clinical trials leveraging centralized repositories and on-site monitoring visits (U.S. Department of Health and Human Services et al., 2013). It is widely recognised and reported that quality assurance procedures are cost intensive (R. M. Califf et al., 1997). Furthermore understanding the cost-effectiveness of clinical trials is challenged by the lack of reported information regarding the data quality, the data quality processes and their overall contribution to the trial performances (Macefield et al., 2013). There are efforts to improve centralized data monitoring with better bias detection (Baigent et al., 2008) and suggestions on how to perform cost efficient on-site visits with targeted focus on information collection processes (J. Bakobaki et al., 2012). The literature and the case studies undertaken in this thesis confirm that most problems stem from source-to-registry data transfer, which suggests that single source data pools with remote and quality-controlled data entry are the best way to improve clinical research (Glöckner et al., 2015).

Data quality implies appropriate data management processes are in place and actually adhered to. All trial protagonists need to have an overview of the need for the data; for the subsequent collected data and its quality. All protagonists should be made aware of the impact of lower data quality on the study. As described, the defined data quality dimensions must focus on:

- *Record Completeness feedback* including statistical information on those records required with their associated values compared to those that were actually entered;
- *Record Accuracy feedback* focused on records with their validated values;

- *Record Eligibility* which focuses on the data needs as they pertain to existing (ongoing) and upcoming clinical trials and studies;
- *Record Timeliness* related to the temporal information of records from when the data was originally recorded to when it was actually entered into the registry;
- *Record Comparability* allowing statistical consistency verification between centres, and
- *Overall Record Data Quality feedback* and how this can/should be calculated.

These dimensions should be assigned and reported to data entry personnel, to investigators and to the clinical research collaborators/community more generally (see Figure 8-10). Automated data quality feedback in combination with the improvements and suggestions conducted through this research should be an ongoing and key part of clinical trials to ensure that higher quality data is collected in a timely manner, and so that the data is subsequently useful for all parties involved in the clinical collaborations and the future researchers wishing to better understand the results of the trials that have taken place. Such repeatability of science is increasingly demanded and a focal area for many research activities, but it is especially critical in a clinical and translational setting.

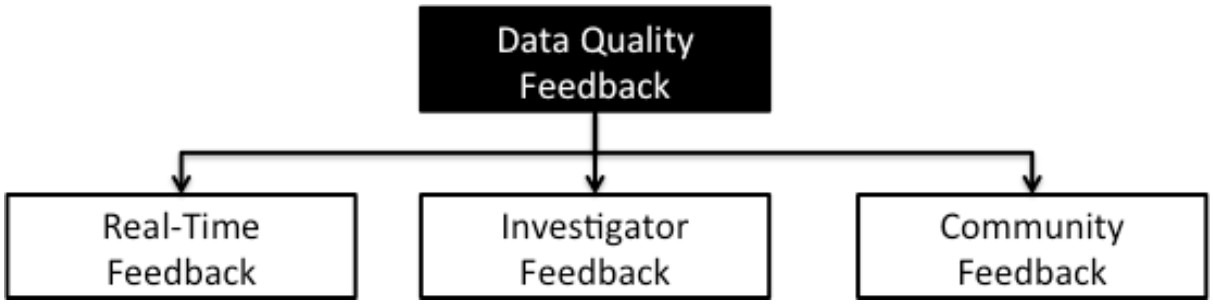


Figure 8-10 Suggested categorised targets for data quality feedback

This approach taken on data quality for the PMT and EURINE-ACT studies showed an overall improvement in data quality. However, this thesis argues that automated feedback for each of these dimensions should be supported with focus on community feedback, investigator feedback and ideally real-time feedback to the data entry personnel. A study framework was applied on INPDR to show the efficacy of the community feedback however the results cannot be generalised, primarily due to the limited data entry over the time of the observation. In general, data quality reports should be delivered through generic tools and processes that are not tightly coupled with any specific study. The basis for any clinical trial or study is the ethics and clinical protocol, which outlines the data that is to be collected and how it should be processed/analysed to achieve a particular outcome. Moving forward, a clear specification of the data quality requirements should be a key part of such protocols.

Data quality feedback mechanisms and tools need to be implemented, supported and considered at the beginning of any new trial, and indeed we argue that these need to be continuously and consistently applied with results shared between all study personnel. Analysis at the end of the clinical study is too late, but unfortunately currently the norm. This thesis has shown that tools and processes now exist to address this situation and should be more systematically adopted and applied.

8.2 Future Work

As already pointed out, some parts of this research were underpowered due to the limited amounts of data entry that occurred during the period under consideration (specifically with regards to INPDR and PCOS). The work on ENDIA and support for mobile applications for patients themselves was also in the early phases of adoption. To gain insights into user motivation and demonstrate clear data quality improvements through feedback, the proposed stage wise study design framework has to be applied to other clinical registries and trials. Data feedback showed an improvement of data quality,

but for good epidemiologic practice a comparison with a control group should ideally be performed. This can happen during randomization of patients in a particular clinical study for example where healthy individuals and used and compared (subject to appropriate randomization) with the actual disease group.

The thesis discovered a correlation of research success and research funding schemes with focus explicitly on adrenal research. Deeper research needs to be conducted to show any true causality however. This can be performed in other research fields (e.g. coronary or selected infectious diseases) with further detailed analysis on R&D spend by country for example. This would potentially show how important multicenter research and ultimately improved VREs could be establish to avoid research funding that does not deliver true value.

Another aspect that needs further research is the impact of national health services on data collection for research purposes. The comparison of centres in a multicentre trial must also focus on comparison with national health policies. As stated in the ENSAT monitoring visit reports, Italy does not financial supporting research on benign tumours, hence patients that are enrolled into a given study protocols have to travel their own costs. More importantly, physicians are not payed for research. In France, a general patient hospitalisation of a malignant tumour is covered by the national health service, which consequently improves data collection possibilities for research purposes. Germany covers only routine care and hence all research must be carried out during standard diagnostics or treatment. Each European health system is different to the NHS in the UK. Germany has a multi-contributing system, covered by the government and/or the individual (depending on insurance type). The UK provides health care services covered by the taxpayers through the NHS. Such differences in health service policies need to be explored and taken into account for the evaluation of data quality in multicentre international trials.

For further projects in the field of eHealth and/or mHealth technologies, this proposed stage wise framework could be applied to gain insights into the behavior of research collaborators, patients and health workers more generally. The work could also be used in other contexts and situations including for example the rising surveillance and case management research needs and demands of developing countries. It is obvious, that the behavior of end user researchers and data entry personnel in third world and developing countries are differently motivated and presumably under resourced compared to collaborators in more established settings, e.g. data entry users in North America, Europe or Australia/New Zealand.

Technology is continually evolving and mobile phones and applications are now a ubiquitous part of people's lives and society more generally. The ENDIA study and the use of mobile apps for dietary information has demonstrated that the technology exists for patient-specific data collection, however again the clear demonstration that these apps have provided a step change in data quality remains largely unanswered. One challenge here is that the ENDIA apps are used for collecting an extensive amount of information from patients on their dietary intake. The amount of data that is to be collected is an important factor that can directly impact upon data quality. Future clinical data collection opportunities with associated data quality reports through mobile applications are expected to occur.

Furthermore, with increased technological advancement, increased importance on the data quality dimension of data trustworthiness will arise. Ensuring that only the mothers are using the ENDIA mobile application can be difficult to enforce (if someone else uses their mobile phone for example). This is issue can of course arise in a hospital setting, but the risk is increased given that mobile phones can be easily mislaid or stolen. Quality dimensions need to be continually adjusted to new research purposes and clinical trials. It should be mandatory that every principal investigator plans the employment of data managers with expertise in data quality assurance and software engineering,

however this can be a cost that may not always be afforded. Nevertheless, detailed training should be a key part to all personnel involved in clinical research and data entry. As shown by this research, people are by far the greatest factors on data quality. Whilst automated data quality feedback can support improvements in data quality, this will have little impact for those with direct intrinsic or extrinsic motivational challenges.

In conclusion, high data quality is an achievement, which is reached by the collaboration and motivation of all researcher in a research community. Data quality feedback encourages collaborators to collect high quality data. Therefore, high data quality will be the tool to improve the efficiency and translation speed of findings in clinical research into clinical care.

REFERENCES

Abate, M. L., Diegert, K. V., & Allen, H. W. (1998). A Hierarchical Approach to Improving Data Quality. *Data Quality Journal*, *4*(1). Retrieved from http://dblp.uni-trier.de/db/journals/dq/dq4.html#AbateDA98

Agosti, M., Crivellari, F., & Di Nunzio, G. M. (2012). Web log analysis: a review of a decade of studies about information acquisition, inspection and interpretation of user interaction. *Data Mining and Knowledge Discovery*, *24*(3), 663–696. https://doi.org/10.1007/s10618-011-0228-8

Allison, M. (2012). Reinventing clinical trials. *Nature Biotechnology*, *30*(1), 41–9. https://doi.org/10.1038/nbt.2083

American College of Medical Quality. (2010). *Medical Quality Management. Theory and Practice.* (P. Varkey, Ed.). Boston, Toronto, London, Singapore: Jones and Bartlett Publishers.

Aranda-Jan, C. B., Mohutsiwa-Dibe, N., & Loukanova, S. (2014). Systematic review on what works, what does not work and why of implementation of mobile health (mHealth) projects in Africa. *BMC Public Health*, *14*(1), 188. https://doi.org/10.1186/1471-2458-14-188

Arlt, W. (2009). In the arena of adrenal disorders significant progress has been made over the last few years. Preface. *Best Practice & Research. Clinical Endocrinology & Metabolism*, *23*(2), vii. https://doi.org/10.1016/j.beem.2009.04.002

Arlt, W. (2011). Disorders of the Adrenal Cortex. In D. Longo, A. Fauci, D. Kasper, S. Hauser, J. L. Jameson, & J. Loscalzo (Eds.), *Harrison's Principles of Internal Medicine: Volume 1* (18th ed., pp. 2940–2961). New York: McGraw-Hill Professional.

Arlt, W., Biehl, M., Taylor, A. E., Hahner, S., Libé, R., Hughes, B. a, … Stewart, P. M. (2011). Urine steroid metabolomics as a biomarker tool for detecting malignancy in adrenal tumors. *The Journal of Clinical Endocrinology and Metabolism*, *96*(12), 3775–84. https://doi.org/10.1210/jc.2011-1565

Arts, D. G. T., De Keizer, N. F., & Scheffer, G.-J. (2002). Defining and improving data quality in medical registries: a literature review, case study, and generic framework. *Journal of the American Medical Informatics Association : JAMIA*, *9*(6), 600–11. https://doi.org/10.1197/jamia.M1087.

Bacchieri, A., & Della Cioppa, G. (2012). Methodological Foundations of Clinical Research. In R. L. Richesson & J. E. Andrews (Eds.), *Clinical Research Informatics* (pp. 49–79). London: Springer.

Baigent, C., Harrell, F. E., Buyse, M., Emberson, J. R., & Altman, D. G. (2008). Ensuring trial validity by data quality assurance and diversification of monitoring methods. *Clinical Trials (London, England)*, *5*(1), 49–55. https://doi.org/10.1177/1740774507087554

Bakobaki, J., Joffe, N., Burdett, S., Tierney, J., Meredith, S., & Stenning, S. (2012). A systematic search for reports of site monitoring technique comparisons in clinical trials. *Clinical Trials (London, England)*, *9*(6), 777–80. https://doi.org/10.1177/1740774512458993

Bakobaki, J. M., Rauchenberger, M., Joffe, N., McCormack, S., Stenning, S., & Meredith, S. (2012). The potential for central monitoring techniques to replace on-site monitoring: findings from an international multi-centre clinical trial. *Clinical Trials* , *9*(2), 257–264. https://doi.org/10.1177/1740774511427325

Balch, C. M., Gershenwald, J. E., Soong, S. -j., Thompson, J. F., Atkins, M. B., Byrd, D. R., … Sondak, V. K. (2009). Final Version of 2009 AJCC Melanoma Staging and Classification. *Journal of Clinical Oncology*, *27*(36), 6199–6206. https://doi.org/10.1200/JCO.2009.23.4799

Barker-Haliski, M., Friedman, D., White, H. S., & French, J. A. (2014). How clinical development can, and should, inform translational science. *Neuron*, *84*(3), 582–93. https://doi.org/10.1016/j.neuron.2014.10.029

Batini, C., & Scannapieco, M. (2006). *Data Quality: Concepts, Methodologies and Techniques*. Berlin, Heidelberg, New York: Springer. Retrieved from http://books.google.com/books?hl=en&lr=&id=SULMBFgtwQoC&pgis=1

Bennett, L. M., & Gadlin, H. (2012). Collaboration and team science: from theory to practice. *Journal of Investigative Medicine*, *60*(5), 768–75. https://doi.org/10.231/JIM.0b013e318250871d

Beuschlein, F. (2013). EJE Prize 2013: Regulation of aldosterone secretion: from physiology to disease. *European Journal of Endocrinology*, *168*(6), R85–R93. https://doi.org/10.1530/EJE-13-0263

Bhatt, A. (2011). Quality of clinical trials: A moving target. *Perspectives in Clinical Research*, *2*(4), 124–8. https://doi.org/10.4103/2229-3485.86880

Blake, M. a, Cronin, C. G., & Boland, G. W. (2010). Adrenal imaging. *AJR. American Journal of Roentgenology*, *194*(6), 1450–60. https://doi.org/10.2214/AJR.10.4547

Bollen, J., Van de Sompel, H., Hagberg, A., & Chute, R. (2009). A principal component analysis of 39 scientific impact measures. *PloS One*, *4*(6), e6022. https://doi.org/10.1371/journal.pone.0006022

Border, P. (2011). Genomics: Sequence sharing. *Nature*, *470*(7333), 169–170. https://doi.org/10.1038/470169a

Bornmann, L., Marx, W., Gasparyan, A. Y., & Kitas, G. D. (2012). Diversity, value and limitations of the journal impact factor and alternative metrics. *Rheumatology International*, *32*(7), 1861–7. https://doi.org/10.1007/s00296-011-2276-1

Bornstein, S. R., & Licinio, J. (2011). Improving the efficacy of translational medicine by optimally integrating health care, academia and industry. *Nature Medicine*, *17*(12), 1567–9. https://doi.org/10.1038/nm.2583

Bostock, M., Ogievetsky, V., & Heer, J. (2011). D3 data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, *17*(12), 2301–2309. https://doi.org/10.1109/TVCG.2011.185

Brady, L. S., & Insel, T. R. (2012). Translating Discoveries into Medicine: Psychiatric Drug Development in 2011. *Neuropsychopharmacology*, *37*(1), 281–283. https://doi.org/10.1038/npp.2011.106

Brandt, C. A., Argraves, S., Money, R., Ananth, G., Trocky, N. M., & Nadkarni, P. M. (2006). Informatics tools to improve clinical research study implementation. *Contemporary Clinical Trials*, *27*(2), 112–22. https://doi.org/10.1016/j.cct.2005.11.013

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(May 2015), 77–101. https://doi.org/10.1191/1478088706qp063oa

Bray, F., & Parkin, D. M. (2009). Evaluation of data quality in the cancer registry: principles and methods. Part I: comparability, validity and timeliness. *European Journal of Cancer (Oxford, England : 1990)*, *45*(5), 747–55. https://doi.org/10.1016/j.ejca.2008.11.032

Brouwer, H. J., Bindels, P. J. E., & Weert, H. C. (2006). Data quality improvement in general practice. *Family Practice*, *23*(5), 529–536. Retrieved from https://search.ebscohost.com/login.aspx?direct=true&db=mnh&AN=16868006&site=ehost-live

Buchan, I., & Bischop, J. W. C. (2009). A Unified Modeling Approach to Data-Intensive Healthcare. In T. Hey, S. Tansley, & K. Tolle (Eds.), *The Fourth Paradigm* (pp. 91–98). Redmond, Washington: Microsoft Research.

Buchan, N. S., Rajpal, D. K., Webster, Y., Alatorre, C., Gudivada, R. C., Zheng, C., … Koehler, J. (2011). The role of translational bioinformatics in drug discovery. *Drug Discovery Today*, *16*(9–10), 426–34. https://doi.org/10.1016/j.drudis.2011.03.002

Buetow, K. H., & Niederhuber, J. (2009). Infrastructure For A Learning Health Care System: CaBIG. *Health Affairs*, *28*(3), 923–924. https://doi.org/10.1377/hlthaff.28.3.923-a

Buneman, P., Khanna, S., & Wang-Chiew, T. (2001). Why and where: A characterization of data provenance. *Database Theory - ICDT 2001*, *1973*(ICDT), 316–330. https://doi.org/10.1007/3-540-44503-X_20

Burns, J. M. (2003). Transforming leadership: A new pursuit of happiness. *Grove Press.*, *213*.

Cafazzo, J. A., Casselman, M., Hamming, N., Katzman, D. K., & Palmert, M. R. (2012). Design of an mHealth app for the self-management of adolescent type 1 diabetes: A pilot study. *Journal of Medical Internet Research*, *14*(3), e70. https://doi.org/10.2196/jmir.2058

Califf, R. (2009). ACS and Acute Heart Failure Models. In *Transforming Clinical Research in the United States*. Washington DC.

Califf, R. M., Karnash, S. L., & Woodlief, L. H. (1997). Developing systems for cost-effective auditing of clinical trials. *Controlled Clinical Trials*, *18*(6), 651-60–6. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9408727

Campbell, H., & Sweatman, J. (2002). Quality Assurance and Clinical Data Management. In R. K. Rondel, S. A. Varley, & C. F. Webb (Eds.), *Clinical Data Management* (pp. 123–141). John Wiley & Sons, Ltd. https://doi.org/10.1002/0470846364.ch7

Cantor, M. N. (2012). Translational informatics: an industry perspective. *Journal of the American Medical Informatics Association*, *19*, 153–155. https://doi.org/10.1136/amiajnl-2011-000588

Catalani, C., Philbrick, W., Fraser, H., Mechael, P., & Israelski, D. M. (2013). mHealth for HIV Treatment & Prevention: A Systematic Review of the Literature. *The Open AIDS Journal*, *7*(1), 17–41. https://doi.org/10.2174/1874613620130812003

Cavusoglu, H. (2015). Can Gamification Motivate Voluntary Contributions ? The Case of StackOverflow Q & A Community. In *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work \& Social Computing* (pp. 171–174). ACM.

Chen, C., Haddad, D., Selsky, J., Hoffman, J. E., Kravitz, R. L., Estrin, D., & Sim, I. (2012). Making sense of mobile health data: an open architecture to improve individual- and population-level health. *Journal of Medical Internet Research*, *14*(4), e112. https://doi.org/10.2196/jmir.2152

Chen, J., Qian, F., Yan, W., & Shen, B. (2013). Translational biomedical informatics in the cloud: present and future. *BioMed Research International*, *2013*, 658925. https://doi.org/10.1155/2013/658925

Chen, K., Chen, H., Conway, N., Hellerstein, J. M., & Parikh, T. S. (2011). Usher: Improving Data Quality with Dynamic Forms. *IEEE Transactions on Knowledge and Data Engineering*, *23*(8), 1138–1153. https://doi.org/10.1109/TKDE.2011.31

Choi, D. W., Armitage, R., Brady, L. S., Coetzee, T., Fisher, W., Hyman, S., … Sherer, T. (2014). Medicines for the mind: policy-based “pull” incentives for creating breakthrough CNS drugs. *Neuron*, *84*(3), 554–63. https://doi.org/10.1016/j.neuron.2014.10.027

Clark, T. (2010). On “being researched”: why do people engage with qualitative research? *Qualitative Research*, *10*(4), 399–419. https://doi.org/10.1177/1468794110366796

Clinical Research Society. (2014). Translational Medicine Paradigms in Clinical Trials. Retrieved July 31, 2014, from http://www.clinicalresearchsociety.org/translationalmedicine/

Collins, F. S. (2011). Reengineering translational science: the time is right. *Science Translational Medicine*, *3*(90), 90cm17. https://doi.org/10.1126/scitranslmed.3002747

Colvis, C. M., & Austin, C. P. (2014). Innovation in Therapeutics Development at the NCATS. *Neuropsychopharmacology*, *39*(1), 230–232. https://doi.org/10.1038/npp.2013.247

Contopoulos-Ioannidis, D. G., Alexiou, G. A., Gouvias, T. C., & Ioannidis, J. P. A. (2008). Medicine. Life cycle of translational research for medical interventions. *Science (New York, N.Y.)*, *321*(5894), 1298–9. https://doi.org/10.1126/science.1160622

Contopoulos-Ioannidis, D. G., Ntzani, E. E., & Ioannidis, J. P. a. (2003). Translation of highly promising basic science research into clinical applications. *The American Journal of Medicine*, *114*(6), 477–484. https://doi.org/10.1016/S0002-9343(03)00013-5

Costa, F. F. (2014). Big data in biomedicine. *Drug Discovery Today*, *19*(4), 433–440. https://doi.org/10.1016/j.drudis.2013.10.012

Crutzen, R., Roosjen, J. L., & Poelman, J. (2013). Using Google Analytics as a process evaluation method for Internet-delivered interventions: an example on sexual health. *Health Promotion International*, *28*(1), 36–42. https://doi.org/10.1093/heapro/das008

Dai, C., Lin, D., Bertino, E., & Kantarcioglu, M. (2008). An approach to evaluate data trustworthiness based on data provenance. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *5159 LNCS*, 82–98. https://doi.org/10.1007/978-3-540-85259-9_6

Darr, R., Lenders, J. W. M., Hofbauer, L. C., Naumann, B., Bornstein, S. R., & Eisenhofer, G. (2012). Pheochromocytoma - update on disease management. *Therapeutic Advances in Endocrinology and Metabolism*, *3*(1), 11–26. https://doi.org/10.1177/2042018812437356

Därr, R., Pamporaki, C., Peitzsch, M., Miehle, K., Prejbisz, A., Peczkowska, M., … Eisenhofer, G. (2014). Biochemical diagnosis of phaeochromocytoma using plasma-free normetanephrine, metanephrine and methoxytyramine: Importance of supine sampling under fasting conditions. *Clinical Endocrinology*, *80*(4), 478–486. https://doi.org/10.1111/cen.12327

Datta, I., Findlay, C., Kortbeek, J. B., & Hameed, S. M. (2007). Evaluation of a regional trauma registry. *Canadian Journal of Surgery. Journal Canadien de Chirurgie*, *50*(3), 210–3. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17568493

Davis, J. R., Nolan, V. P., Woodcock, J., & Estabrook, R. W. (1999). *Assuring Data Quality and Validity in Clinical Trials for Regulatory Decision Making*.

Day, S., Fayers, P., & Harvey, D. (1998). Double data entry: what value, what price? *Controlled Clinical Trials*, *19*(1), 15–24. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9492966

de Lusignan, S. (2002). Does Feedback Improve the Quality of Computerized Medical Records in Primary Care? *Journal of the American Medical Informatics Association*, *9*(4), 395–401. https://doi.org/10.1197/jamia.M1023

de Lusignan, S. (2005). Using feedback to raise the quality of primary care computer data: a literature review. *Studies in Health Technology and Informatics*, *116*(Box 1), 593–598.

Dechartres, A., Boutron, I., Trinquart, L., Charles, P., & Ravaud, P. (2011). Single-center trials show larger treatment effects than multicenter trials: evidence from a meta-epidemiologic study. *Annals of Internal Medicine*, *155*(1), 39–51. https://doi.org/10.7326/0003-4819-155-1-201107050-00006

Deming, W. E. (2000). *Out of the Crisis*. MIT Press. Retrieved from http://books.google.com.au/books/about/Out_of_the_Crisis.html?id=LA15eDlOPgoC&pgis=1

Deterding, S. (2012). Gamification: designing for motivation. *Interactions*, *19*, 14–17. https://doi.org/10.1145/2212877.2212883

Deterding, S., Sicart, M., Nacke, L., O'Hara, K., & Dixon, D. (2011). Gamification. using game-design elements in non-gaming contexts. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems* (pp. 2425–2428).

DFG. (2016a). DFG Profile - Finances. Retrieved April 14, 2016, from http://www.dfg.de/en/dfg_profile/facts_figures/statistics/finances/index.jsp

DFG. (2016b). DFG Profile - Processing Times and Success Rates. Retrieved August 5, 2016, from http://www.dfg.de/en/dfg_profile/facts_figures/statistics/processing_times_success_rates/index.html

Division, B., & Pharmaceuticals, G. (2001). Mining the bibliome. In P. R. O. Payne, P. J. Embi, I. N. Sarkar, N. Shah, J. D. Tenenbaum, & A. B. Wilcox (Eds.), *Translational Informatics* (Vol. 2, pp. 88–89). London: Springer. https://doi.org/10.1038/35076512

Donabedian, A. (1988). The quality of care. How can it be assessed? *JAMA : The Journal of the American Medical Association*, *260*(12), 1743–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3045356

ECRIN. (2007). *ECRIN Report - Survey on data management, tools and procedures within ECRIN*.

ECRIN. (2010). Standard requirements for GCP- compliant data management in multinational clinical trials, (May), 1–24.

ECRIN. (2013). Requirements for Certification of ECRIN Data Centres. *Trials*, *14*.

Effendy, J. (2014). *Cloud-Based Logging and Auditing System for ENS@T*. University of Melbourne.

Eisenhofer, G. (2004). Malignant pheochromocytoma: current status and initiatives for future progress. *Endocrine Related Cancer*, *11*(3), 423–436. https://doi.org/10.1677/erc.1.00829

English, R. a., Lebovitz, Y., & Griffin, R. B. (2010). *Transforming Clinical Research in the United States. Challanges and Opportunities. National Academies Press*.

Estrin, D. (2014). Small data, where n= me. In *Communications of the ACM* (pp. 32–34).

Estrin, D., & Sim, I. (2010). Health care delivery. Open mHealth architecture: an engine for health care innovation. *PLoS Medicine*, *10*(2), e10011395. https://doi.org/10.1126/science.1196187

Fähnrich, C., Denecke, K., Adeoye, O. O., Benzler, J., Claus, H., Kirchner, G., … Krause, G. (2015). Surveillance and Outbreak Response Management System (SORMAS) to support the control of the Ebola virus disease outbreak in West Africa. *Euro Surveillance : Bulletin Europeen Sur Les Maladies Transmissibles = European Communicable Disease Bulletin*, *20*(12). https://doi.org/10.3201/eid1911.121843

Falagas, M. E., Kouranos, V. D., Arencibia-Jorge, R., & Karageorgopoulos, D. E. (2008). Comparison of SCImago journal rank indicator with journal impact factor. *The FASEB Journal*, *22*(8), 2623–2628.

https://doi.org/10.1096/fj.08-107938

Fassnacht, M., & Allolio, B. (2009). Clinical management of adrenocortical carcinoma. *Best Practice & Research. Clinical Endocrinology & Metabolism*, *23*(2), 273–89. https://doi.org/10.1016/j.beem.2008.10.008

Fernandes, J., Duarte, D., Ribeiro, C., Farinha, C., Pereira, J. M., & Da Silva, M. M. (2012). IThink : A game-based approach towards improving collaboration and participation in requirement elicitation. In *Procedia Computer Science* (Vol. 15, pp. 66–77). https://doi.org/10.1016/j.procs.2012.10.059

Franco, R., Hronec, M., & Slizgi, B. (2013). Higher Quality at Lower Cost: The Benefits of Portfolio-Wide Predictive Analytics in Clinical Trial Monitoring. Retrieved from http://www.appliedclinicaltrialsonline.com/higher-quality-lower-cost-benefits-portfolio-wide-predictive-analytics-clinical-trial-monitoring?rel=canonical

Fry, C., & Dwyer, R. (2001). For love or money? An exploratory study of why injecting drug users participate in research. *Addiction (Abingdon, England)*, *96*(9), 1319–1325. https://doi.org/10.1080/09652140120070373

Fulgoni, G. (2015). The Rise of the Digital Omnivore: What It Means for Advertisers, Publishers, and App Developers. *Journal of Advertising Research*, *55*(2). Retrieved from http://zenithinfostation.my/wp-content/uploads/2015/07/The_Rise_of_the_Digital_Omnivore_What_It_Means_for_Advertisers_Publishers_and.pdf

Gatta, G., Capocaccia, R., Trama, A., & Martinez, C. (2010). The Burden of Rare Cancers in Europe. In M. Posada de la Paz & S. C. Groft (Eds.), *Rare Diseases Epidemiology* (Vol. 686, pp. 285–303). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-90-481-9485-8

GCP Inspections Working Group. (2007). *EMEA Reflection paper on expectations for electronic source documents used in clinical*.

Glöckner, S., Arlt, W., Bancos, I., Stell, A. J., & Sinnott, R. (2015). Improving Data Quality in Disease Registries and Clinical Trials: A Case Study from the ENSAT-CANCER Project. In A. Maeder & J. Warren (Eds.), *8th Australasian Workshop on Health Informatics and Knowledge Management (HIKM 2015)* (Vol. 164, pp. 25–32). Sydney, Australia: ACS. https://doi.org/10.13140/RG.2.1.5177.3202

Glöckner, S., & Schoffer, O. (2012). *Praktikumsbericht. Praktikum an der Tumorepidemiologie des Universitätsklinikums Carl Gustav Carus Dresden*. Martin Luther University Halle/Wittenberg.

Golden, S. H., Robinson, K. a, Saldanha, I., Anton, B., & Ladenson, P. W. (2009). Clinical review: Prevalence and incidence of endocrine and metabolic disorders in the United States: a comprehensive review. *The Journal of Clinical Endocrinology and Metabolism*, *94*(6), 1853–78. https://doi.org/10.1210/jc.2008-2291

Green, E. D., Guyer, M. S., & National Human Genome Research Institute. (2011). Charting a course for genomic medicine from base pairs to bedside. *Nature*, *470*(7333), 204–13. https://doi.org/10.1038/nature09764

Grove, J. (2011, July 28). Triple miracle sees huge rise in EU funds for frontier research. Retrieved April 14, 2016, from https://www.timeshighereducation.com/news/triple-miracle-sees-huge-rise-in-eu-funds-for-frontier-research/416952.article?storycode=416952

Gülcü, C. (2012). log4j Class Level. Retrieved October 17, 2016, from https://logging.apache.org/log4j/1.2/apidocs/org/apache/log4j/Level.html

Guthrie, L. B., Oken, E., Sterne, J. A. C., Gillman, M. W., Patel, R., Vilchuck, K., … Martin, R. M. (2012). Ongoing monitoring of data clustering in multicenter studies. *BMC Medical Research Methodology*, *12*, 29. https://doi.org/10.1186/1471-2288-12-29

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? - A literature review of empirical studies on gamification. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 3025–3034. https://doi.org/10.1109/HICSS.2014.377

Hamine, S., Gerth-Guyette, E., Faulx, D., Green, B. B., & Ginsburg, A. S. (2015). Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. *Journal of Medical Internet Research*, *17*(2), e52. https://doi.org/10.2196/jmir.3951

Harris, P. a, Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, *42*(2), 377–81. https://doi.org/10.1016/j.jbi.2008.08.010

Hentschel, S., Pritzkuleit, R., Schmid-Höpfner, S., & Katalinic, A. (2011). Epidemiologische krebsregistrierung in Deutschland: Aufgaben und aktueller status. *Onkologe*, *17*, 97–106. https://doi.org/10.1007/s00761-010-1939-y

Herger, M. (2014). *Enterprise Gamification: Engaging People by Letting Them Have Fun*. CreateSpace.

Herzberg, S., Rahbar, K., Stegger, L., Schäfers Michael, M., & Dugas, M. (2011). Concept and implementation of a computer-based reminder system to increase completeness in clinical documentation. *International Journal of Medical Informatics*, *80*(5), 351–358. https://doi.org/10.1016/j.ijmedinf.2011.02.004

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, *520*(7548), 429–431. https://doi.org/10.1038/520429a

Hills, R. A., Revere, D., Altamore, R., Abernethy, N. F., Lober, W. B., & State, W. (2010). Timeliness and Data Element Completeness of Immunization Data in Washington State in 2010 : a Comparison of Data Exchange Methods, 340–349.

Hirdes, J. P., Poss, J. W., Caldarelli, H., Fries, B. E., Morris, J. N., Teare, G. F., … Jutan, N. (2013). An evaluation of data quality in Canada's Continuing Care Reporting System (CCRS): secondary analyses of Ontario data submitted between 1996 and 2011. *BMC Medical Informatics and Decision Making*, *13*(1), 27. https://doi.org/10.1186/1472-6947-13-27

Hofstädter, F., & Klinkhammer-Schalke, M. (2011). Aufgaben und Strukturen für die klinische Krebsregistrierung. *Der Onkologe*, *17*(2), 121–125. https://doi.org/10.1007/s00761-010-1938-z

Holle, R. (1995). *Methoden zur Konstruktion und Evaluierung klinischer Scores*. Inst. für Med. Biometrie u. Informatik, Abt. Med. Biometrie.

Hong, M. K. H., Yao, H. H. I., Pedersen, J. S., Peters, J. S., Costello, A. J., Murphy, D. G., … Corcoran, N. M. (2013). Error rates in a clinical data repository: lessons from the transition to electronic data transfer--a descriptive study. *BMJ Open*, *3*(5), 1–8. https://doi.org/10.1136/bmjopen-2012-002406

Huang, H. M., & Liaw, S. S. (2005). Exploring users' attitudes and intentions toward the web as a survey tool. *Computers in Human Behavior*, *21*(5), 729–743. https://doi.org/10.1016/j.chb.2004.02.020

Huang, Z., Dong, W., Ji, L., Gan, C., Lu, X., & Duan, H. (2014). Discovery of clinical pathway patterns from event logs using probabilistic topic models. *Journal of Biomedical Informatics*, *47*, 39–57. https://doi.org/10.1016/j.jbi.2013.09.003

Ibukun, Y. (2014). Nigeria Uses Android App With Facebook to Beat Ebola. Retrieved October 22, 2016, from http://www.bloomberg.com/news/2014-10-07/nigeria-usesandroid-app-with-facebook-to-beat-ebola.html

ICD. (2012). *Maximizing Mobile for Development*. Retrieved from http://siteresources.worldbank.org/EXTINFORMATIONANDCOMMUNICATIONANDTECHNOLOGIES/Resources/IC4D_Infographic-1.png

International Conference on Harmonization. (1996). *E6 Guideline for Good Clinical Practice* (Vol. 1996). Retrieved from http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E6_R1/Step4/E6_R1__Guideline.pdf

International Human Genome Sequencing Consortium. (2004). Finishing the euchromatic sequence of the human genome. *Nature*, *431*(7011), 931–45. https://doi.org/10.1038/nature03001

JCAHO. (1990). *Accreditation Manual for Hospitals*. (Joint Commission on Accreditation of Healthcare Organizations, Ed.). Chicago.

Jensen, P. B., Jensen, L. J., & Brunak, S. (2012). Mining electronic health records: towards better research applications and clinical care. *Nature Reviews. Genetics*, *13*(6), 395–405. https://doi.org/10.1038/nrg3208

Joos, O., Silva, R., Amouzou, A., Moulton, L. H., Perin, J., Bryce, J., … L.C., M. (2016). Evaluation of a mhealth data quality intervention to improve documentation of pregnancy outcomes by health surveillance assistants in Malawi: A Cluster randomized trial. *PLoS ONE*, *11*(1), e0145238. https://doi.org/10.1371/journal.pone.0145238

Juran, J. M., & De Feo, J. A. (2012). *Juran's quality handbook* (6th ed.). New York: McGraw-Hill Professional. Retrieved from http://accessengineeringlibrary.com.ezp.lib.unimelb.edu.au/browse/jurans-quality-handbook-the-complete-guide-to-performance-excellence-sixth-edition/Preface03

Kapoor, A., Morris, T., & Rebello, R. (2011). Guidelines for the management of the incidentally discovered adrenal mass. *Canadian Urological Association Journal*, *5*(4), 241–247. https://doi.org/10.5489/cuaj.11135

Katalinic, A., & Hentschel, S. (2008). Epidemiologische Krebsregistrierung in Deutschland. In *Das Manual der epidemiologischen Krebsregistrierung* (pp. 1–8).

Kebebew, E., Reiff, E., Duh, Q.-Y., Clark, O. H., & McMillan, A. (2006). Extent of disease at presentation and outcome for adrenocortical carcinoma: have we made progress? *World Journal of Surgery*, *30*(5), 872–8. https://doi.org/10.1007/s00268-005-0329-x

Kelling, S., Hochachka, W. M., Fink, D., Riedewald, M., Caruana, R., Ballard, G., & Hooker, G. (2009). Data-intensive Science: A New Paradigm for Biodiversity Studies. *BioScience*, *59*(7), 613–620. https://doi.org/10.1525/bio.2009.59.7.12

Khatib, F., DiMaio, F., Cooper, S., Kazmierczyk, M., Gilski, M., Krzywda, S., … Baker, D. (2012). Crystal structure of a monomeric retroviral protease solved by protein folding game players. *Nature Structural & Molecular Biology*. https://doi.org/10.1038/nsmb0312-364b

Khoury, M. J., Gwinn, M., Yoon, P. W., Dowling, N., Moore, C. a, & Bradley, L. (2007). The continuum of translation research in genomic medicine: how can we accelerate the appropriate integration of human genome discoveries into health care and disease prevention? *Genetics in Medicine*, *9*(10), 665–674. https://doi.org/10.1097/GIM.0b013e31815699d0

Knepper, D., Fenske, C., Nadolny, P., Bedding, A., Gribkova, E., Polzer, J., … Lawton, A. (2016). Detecting Data Quality Issues in Clinical Trials: Current Practices and Recommendations. *Therapeutic Innovation & Regulatory Science*, *50*(1), 15–21. https://doi.org/10.1177/2168479015620248

Koivisto, J., & Hamari, J. (2014). Demographic differences in perceived benefits from gamification. *Computers in Human Behavior*, *35*, 179–188. https://doi.org/10.1016/j.chb.2014.03.007

Köpcke, F., Kraus, S., Scholler, A., Nau, C., Schüttler, J., Prokosch, H.-U., & Ganslandt, T. (2013). Secondary use of routinely collected patient data in a clinical trial: an evaluation of the effects on patient recruitment and data acquisition. *International Journal of Medical Informatics*, *82*(3), 185–92. https://doi.org/10.1016/j.ijmedinf.2012.11.008

Köpcke, F., Trinczek, B., Majeed, R. W., Schreiweis, B., Wenk, J., Leusch, T., … Prokosch, H.-U. (2013). Evaluation of data completeness in the electronic health record for the purpose of patient recruitment into clinical trials: a retrospective analysis of element presence. *BMC Medical Informatics and Decision Making*, *13*, 37. https://doi.org/10.1186/1472-6947-13-37

Krzych, L. J., Lees, B., Nugara, F., Banya, W., Bochenek, A., Cook, J., … Flather, M. D. (2011). Assessment of data quality in an international multi-centre randomised trial of coronary artery surgery. *Trials*, *12*(1), 212. https://doi.org/10.1186/1745-6215-12-212

Kuchinke, W., Aerts, J., Semler, S. C., & Ohmann, C. (2009). CDISC standard-based electronic archiving of clinical trials. *Methods of Information in Medicine*, *48*(5), 408–13. https://doi.org/10.3414/ME9236

Kuchinke, W., Ohmann, C., Yang, Q., Salas, N., Lauritsen, J., Gueyffier, F., … Lejeune, S. (2010). Heterogeneity prevails: the state of clinical trial data management in Europe - results of a survey of ECRIN centres. *Trials*, *11*, 79. https://doi.org/10.1186/1745-6215-11-79

Kumar, S., Nilsen, W. J., Abernethy, A., Atienza, A., Patrick, K., Pavel, M., … Swendeman, D. (2013). Mobile health technology evaluation: The mHealth evidence workshop. *American Journal of Preventive Medicine*, *45*(2), 228–236. https://doi.org/10.1016/j.amepre.2013.03.017

Laberge, M., & Shachak, A. (2013). Developing a tool to assess the quality of socio-demographic data in community health centres. *Applied Clinical Informatics*, *4*(1), 1–11. https://doi.org/10.4338/ACI-2012-10-CR-0041

Lane, S. J., Heddle, N. M., Arnold, E., & Walker, I. (2006). A review of randomized controlled trials comparing the effectiveness of hand held computers with paper methods for data collection. *BMC Medical Informatics and Decision Making*, *6*, 23. https://doi.org/10.1186/1472-6947-6-23

Larsen, I. K., Småstuen, M., Johannesen, T. B., Langmark, F., Parkin, D. M., Bray, F., & Møller, B. (2009). Data quality at the Cancer Registry of Norway: An overview of comparability, completeness, validity and timeliness. *European Journal of Cancer*, *45*(7), 1218–1231. https://doi.org/10.1016/j.ejca.2008.10.037

Le Jeannic, A., Quelen, C., Alberti, C., & Durand-Zaleski, I. (2014). Comparison of two data collection processes in clinical studies: electronic and paper case report forms. *BMC Medical Research Methodology*, *14*, 7. https://doi.org/10.1186/1471-2288-14-7

Ledford, H. (2011). Translational research: 4 ways to fix the clinical trial. *Nature*, *477*(7366), 526–8. https://doi.org/10.1038/477526a

Luke, D. A., Carothers, B. J., Dhand, A., Bell, R. A., Moreland-Russell, S., Sarli, C. C., & Evanoff, B. A. (2015). Breaking Down Silos: Mapping Growth of Cross-Disciplinary Collaboration in a Translational Science Initiative. *Clinical and Translational Science*, *8*(2), 143–149. https://doi.org/10.1111/cts.12248

Luxton, D. D., McCann, R. A., Bush, N. E., Mishkind, M. C., & Reger, G. M. (2011). mHealth for Mental Health: Integrating Smartphone Technology in Behavioral Healthcare. *Professional Psychology-Research and Practice*, *42*(6), 505–512. https://doi.org/10.1037/a0024485

Macefield, R. C., Beswick, A. D., Blazeby, J. M., & Lane, J. A. (2013). A systematic review of on-site monitoring methods for health-care randomised controlled trials. *Clinical Trials (London, England)*, *10*(1), 104–24. https://doi.org/10.1177/1740774512467405

Martin-Sanchez, F., & Verspoor, K. (2014). Big data in medicine is driving big changes. *Yearbook of Medical Informatics*, *9*(1), 14–20. https://doi.org/10.15265/IY-2014-0020

Mathers, C. D., Fat, D. M., Inoue, M., Rao, C., & Lopez, A. D. (2005). Counting the dead and what they died from: an assessment of the global status of cause of death data. *Bulletin of the World Health Organization*, *83*(3), 171–7. https://doi.org//S0042-96862005000300009

McGonigal, J. (2011). *Reality is broken: Why games make us better and how they can change the world*. Penguin.

McShane, L. M., Cavenagh, M. M., Lively, T. G., Eberhard, D. a, Bigbee, W. L., Williams, P. M., … Conley, B. a. (2013). Criteria for the use of omics-based predictors in clinical trials. *Nature*, *502*(7471), 317–20. https://doi.org/10.1038/nature12564

Moen, R. D., & Norman, C. L. (2010). Circling Back: Clearing up the myths about the Deming cycle and seeing how it keeps evolving. *Quality Progress*, (November), 22–28.

Mongkolwat, P., Kleper, V., Talbot, S., & Rubin, D. (2014). The National Cancer Informatics Program (NCIP) Annotation and Image Markup (AIM) Foundation Model. *Journal of Digital Imaging*, *27*(6), 692–701. https://doi.org/10.1007/s10278-014-9710-3

Moorhead, S. A., Hazlett, D. E., Harrison, L., Carroll, J. K., Irwin, A., & Hoving, C. (2013). A New Dimension of Health Care: Systematic Review of the Uses, Benefits, and Limitations of Social Media for Health Communication. *Journal of Medical Internet Research*, *15*(4), e85. https://doi.org/10.2196/jmir.1933

Nahm, M. (2012). Data Quality in Clinical Research. In R. L. Richesson & J. E. Andrews (Eds.), *Clinical Research Informatics* (pp. 175–201). London: Springer. https://doi.org/10.1007/978-1-84882-448-5_10

Nahm, M., Pieper, C. F., & Cunningham, M. M. (2008). Quantifying data quality for clinical trials using electronic data capture. *PloS One*, *3*(8), e3049. https://doi.org/10.1371/journal.pone.0003049

Nair, B. G., Newman, S. F., Peterson, G. N., Wu, W. Y., & Schwid, H. a. (2010). Feedback mechanisms including real-time electronic alerts to achieve near 100% timely prophylactic antibiotic administration in surgical cases. *Anesthesia and Analgesia*, *111*(5), 1293–1300. https://doi.org/10.1213/ANE.0b013e3181f46d89

Naumann, F., Leser, U., & Freytag, J. C. (1999). Quality-driven integration of heterogeneous information systems.

Nielsen, G., Peschel, L., & Burgess, A. (2014). Essential Documentation Elements. *Advanced Emergency Nursing Journal*, *36*(2), 199–205. https://doi.org/10.1097/TME.0000000000000020

Nieman, L. K. (2010). Approach to the patient with an adrenal incidentaloma. *The Journal of Clinical Endocrinology and Metabolism*, *95*(9), 4106–13. https://doi.org/10.1210/jc.2010-0457

Nonnemacher, M., Nassek, D., & Stausberg, J. (2014). *Datenqualität in der medizinischen Forschung. Leitlinie zum adaptiven Management von Datenqualität in Kohortenstudien und Registern.* (2nd ed.). Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft.

Nonnemacher, M., Weiland, D., & Stausberg, J. (2007). *Datenqualität in der medizinischen Forschung: Leitlinie zum adaptiven Management von Datenqualität in Kohortenstudien und Registern*. Berlin: MWV Medizinisch Wiss. Ver.

Nurse, J. R., Agrafiotis, I., Goldsmith, M., Creese, S., & Lamberts, K. (2014). Two sides of the coin: measuring and communicating the trustworthiness of online information. *Journal of Trust Management*, *1*(1), 5. https://doi.org/10.1186/2196-064X-1-5

OECD. (2016). Gross domestic spending on R&D (indicator). https://doi.org/10.1787/d8b068b4-en

Ohmann, C., Canham, S., Cornu, C., Dreß, J., Gueyffier, F., Kuchinke, W., … Wittenberg, M. (2013). Revising the ECRIN standard requirements for information technology and data management in clinical trials. *Trials*, *14*, 97. https://doi.org/10.1186/1745-6215-14-97

Ohmann, C., & Kuchinke, W. (2009). Future Developments of Medical Informatics from the Viewpoint of Networked Clinical Research. *Methods of Information in Medicine*, 45–54. https://doi.org/10.3414/ME9137

Ohmann, C., Kuchinke, W., Canham, S., Lauritsen, J., Salas, N., Schade-Brittinger, C., … Torres, F. (2011). Standard requirements for GCP-compliant data management in multinational clinical trials. *Trials*, *12*, 85. https://doi.org/10.1186/1745-6215-12-85

Ottevanger, P. B., Therasse, P., van de Velde, C., Bernier, J., van Krieken, H., Grol, R., & De Mulder, P. (2003). Quality assurance in clinical trials. *Critical Reviews in Oncology/hematology*, *47*(3), 213–35. https://doi.org/10.1016/S1040-8428(03)00028-3

Parkin, D. M., & Bray, F. (2009). Evaluation of data quality in the cancer registry: principles and methods Part II. Completeness. *European Journal of Cancer (Oxford, England : 1990)*, *45*(5), 756–64. https://doi.org/10.1016/j.ejca.2008.11.033

Paulsen, A., Overgaard, S., & Lauritsen, J. M. (2012). Quality of data entry using single entry, double entry and automated forms processing--an example based on a study of patient-reported outcomes. *PloS One*, *7*(4), e35087. https://doi.org/10.1371/journal.pone.0035087

Payne, P. R. O., & Embi, P. J. (2015). An Introduction to Translational Informatics and the Future of Knowledge-Driven Healthcare. In P. R. O. Payne, P. J. Embi, I. N. Sarkar, N. Shah, J. D. Tenenbaum, & A. B. Wilcox (Eds.), *Translational Informatics* (pp. 3–19). London: Springer.

Payne, P. R. O., Embi, P. J., & Sen, C. K. (2009). Translational informatics: enabling high-throughput research paradigms. *Physiological Genomics*, *39*(3), 131–40. https://doi.org/10.1152/physiolgenomics.00050.2009

Payne, P. R. O., Johnson, S. B., Starren, J. B., Tilson, H. H., & Dowdy, D. (2005). Breaking the Translational Barriers: The Value of Integrating Biomedical Informatics and Translational Research. *Journal of Investigative Medicine*, *53*(4), 192. https://doi.org/10.2310/6650.2005.00402

Payne, P. R. O., Lussier, Y., Foraker, R. E., & Embi, P. J. (2016). Rethinking the role and impact of health information technology: informatics as an interventional discipline. *BMC Medical Informatics and Decision Making*, *16*(1), 40. https://doi.org/10.1186/s12911-016-0278-3

Penno, M. a S., Couper, J. J., Craig, M. E., Colman, P. G., Rawlinson, W. D., Cotterill, A. M., … Harrison, L. C. (2013). Environmental determinants of islet autoimmunity (ENDIA): a pregnancy to early life cohort study in children at-risk of type 1 diabetes. *BMC Pediatrics*, *13*(1), 124. https://doi.org/10.1186/1471-2431-13-124

Phillips, B., Ball, C., Sackett, D., Badenoch, D., Straus, S., Haynes, B., … Howick, J. (2009). *Levels of*

Evidence (March 2009). Oxford.

Pichai, S. (2016). *Google I/O 2016 Keynote*. Mountain View. Retrieved from https://www.youtube.com/watch?v=862r3XS2YB0

Porcheret, M., Hughes, R., Evans, D., Jordan, K., Whitehurst, T., Ogden, H., & Croft, P. (2004). Data Quality of General Practice Electronic Health Records: The Impact of a Program of Assessments, Feedback, and Training. *Journal of the American Medical Informatics Association*, *11*, 78–86. https://doi.org/10.1197/jamia.M1362

Porgo, T. V., Moore, L., & Tardif, P.-A. (2016). *Evidence of Data Quality in Trauma Registries*. *Journal of Trauma and Acute Care Surgery* (Vol. Publish Ah). https://doi.org/10.1097/TA.0000000000000970

Prokscha, S. (2007). *Practical Guide to Clinical Data Management* (2nd ed.). Boca Raton: CRC Press Taylor & Francis Group.

Puttkammer, N., Baseman, J. G., Devine, E. B., Valles, J. S., Hyppolite, N., Garilus, F., … Barnhart, S. (2016). An assessment of data quality in a multi-site electronic medical record system in Haiti. *International Journal of Medical Informatics*, *86*, 104–116. https://doi.org/10.1016/j.ijmedinf.2015.11.003

Quilici, J., Fugon, L., Beguin, S., Morange, P. E., Bonnet, J.-L., Alessi, M.-C., … Cuisset, T. (2013). Effect of motivational mobile phone short message service on aspirin adherence after coronary stenting for acute coronary syndrome. *International Journal of Cardiology*, *168*(1), 568–569. https://doi.org/10.1016/j.ijcard.2013.01.252

Ramsey, B. W., Davies, J., McElvaney, N. G., Tullis, E., Bell, S. C., Dřevínek, P., … Elborn, J. S. (2011). A CFTR Potentiator in Patients with Cystic Fibrosis and the G551D Mutation. *New England Journal of Medicine*, *365*(18), 1663–1672. https://doi.org/10.1056/NEJMoa1105185

Rashid, A. M., Ling, K., Tassone, R. D., Resnick, P., Kraut, R., & Riedl, J. (2006). Motivating participation by displaying the value of contribution. *ACM Conference on Human Factors in Computing Systems*, 955–958. https://doi.org/10.1145/1124772.1124915

Reeves, B., & Read, J. L. (2013). *Total engagement: How games and virtual worlds are changing the way people work and businesses compete*. Harvard Business Press.

Reisch, N., Walz, M. K., Erlic, Z., & Neumann, H. P. H. (2009). [Pheochromocytoma - still a challenge]. *Der Internist*, *50*(1), 27–35. https://doi.org/10.1007/s00108-008-2196-7

Richesson, R. L., & Andrews, J. E. (2012). *Clinical Research Informatics*. (R. L. Richesson & J. E. Andrews, Eds.). London: Springer.

Richesson, R. L., & Krischer, J. (2007). Data standards in clinical research: gaps, overlaps, challenges and future directions. *Journal of the American Medical Informatics Association : JAMIA*, *14*(6), 687–96. https://doi.org/10.1197/jamia.M2470

Rock, E. P., Molloy, V. J., & Humphrey, J. S. (2010). GCP data quality for early clinical development. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research*, *16*(6), 1756–63. https://doi.org/10.1158/1078-0432.CCR-09-3267

Rostami, R., Nahm, M., & Pieper, C. F. (2009). What can we learn from a decade of database audits? The Duke Clinical Research Institute experience, 1997--2006. *Clinical Trials (London, England)*, *6*(2), 141–50. https://doi.org/10.1177/1740774509102590

Rothman, K. J., Greenland, S., & Lash. (2008). *Modern Epidemiology*. *Book* (3rd ed.). Lippincott Williams & Wilkins.

Rothman, K. J., & Monson, R. (1973). Survival in trigeminal neuralgia. *Journal of Chronic Diseases*, *26*(5), 303–9. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/4718951

Rui, J. (2016). *Exploring Data Quality and Usage Patterns of the International Niemann-Pick Disease Registry*. University of Melbourne.

Rui, J., & Sinnott, R. (2016). *Niemann-Pick Data Analytics*. University of Melbourne, Melbourne.

Ryan, R. M., & Deci, E. (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology*, *25*(1), 54–67. https://doi.org/10.1006/ceps.1999.1020

Sandman, L., Mosher, A., Khan, A., Tapy, J., Condos, R., Ferrell, S., & Vernon, A. (2006). Quality assurance in a large clinical trials consortium: the experience of the Tuberculosis Trials Consortium. *Contemporary Clinical Trials*, *27*(6), 554–60. https://doi.org/10.1016/j.cct.2006.06.003

Sarkar, I. N. (2010). Biomedical informatics and translational medicine. *Journal of Translational Medicine*, *8*, 22. https://doi.org/10.1186/1479-5876-8-22

Say, P., Stein, D. M., Ancker, J. S., Hsieh, C.-K., Pollak, J. P., & Estrin, D. (2015). Smartphone Data in Rheumatoid Arthritis - What Do Rheumatologists Want? *AMIA ... Annual Symposium Proceedings. AMIA Symposium*, *2015*, 1130–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/26958252

Schirpenbach, C., & Reincke, M. (2007). Primary aldosteronism: current knowledge and controversies in Conn's syndrome. *Nature Clinical Practice. Endocrinology & Metabolism*, *3*(3), 220–227. https://doi.org/10.1038/ncpendmet0430

Schuchman, E. H. (2007). The pathogenesis and treatment of acid sphingomyelinase-deficient Niemann-Pick disease. *Journal of Inherited Metabolic Disease*, *30*(5), 654–663. https://doi.org/10.1007/s10545-007-0632-9

Shadish, W. R., Cook, T., & Campbell, D. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Houghton: Mifflin and Company.

Shah, N. H., & Tenenbaum, J. D. (2012). The coming age of data-driven medicine: translational bioinformatics' next frontier. *Journal of the American Medical Informatics Association*, *19*(e1), e2–e4. https://doi.org/10.1136/amiajnl-2012-000969

Shiloach, M., Frencher, S. K., Steeger, J. E., Rowell, K. S., Bartzokis, K., Tomeh, M. G., … Hall, B. L. (2010). Toward Robust Information: Data Quality and Inter-Rater Reliability in the American College of Surgeons National Surgical Quality Improvement Program. *Journal of the American College of Surgeons*, *210*(1), 6–16. https://doi.org/10.1016/j.jamcollsurg.2009.09.031

Sim, I., & Niland, J. C. (2012). Study Protocol Representation. In R. L. Richesson & J. E. Andrews (Eds.), *Clinical Research Informatics* (pp. 155–174). London: Springer. https://doi.org/10.1007/978-1-84882-448-5_9

Sinnott, R., Han, J., Hu, W., Ma, X., & Yu, K. (2015). Application of Mobile Games to Support Clinical Data Collection for Patients with Niemann-Pick Disease. *2015 IEEE 28th International Symposium on Computer-Based Medical Systems*, 1–6. https://doi.org/10.1109/CBMS.2015.9

Sinnott, R., & Stell, A. J. (2011). Towards a Virtual Research Environment for International Adrenal Cancer Research. In *Procedia Computer Science* (Vol. 4, pp. 1109–1118). https://doi.org/10.1016/j.procs.2011.04.118

Sinnott, R., Stell, A., & Jiang, J. (2011). Classifying Data Sharing Architectural Models for e-Health

Collaborations. In *9th International HealthGrid Conference*.

Society for Clinical Data Management. (2005). *Good Clinical Data Management Practices. Version 4*. Retrieved from https://ncisvn.nci.nih.gov/WebSVN/filedetails.php?repname=ctms-forum&path=%252F2-Analyst_folders%252Fmichele_working%252Fgcdmp_v4.pdf

Souza, T., Kush, R., & Evans, J. P. (2007). Global clinical data interchange standards are here! *Drug Discovery Today*, *12*(3–4), 174–81. https://doi.org/10.1016/j.drudis.2006.12.012

Stausberg, J., Nonnemacher, M., Weiland, D., Antony, G., & Neuhäuser, M. (2006). Management of data quality--development of a computer-mediated guideline. *Studies in Health Technology and Informatics*, *124*, 477–482.

Stead, W. W., Searle, J. R., Fessler, H. E., Smith, J. W., & Shortliffe, E. H. (2011). Biomedical informatics: changing what physicians need to know and how they learn. *Academic Medicine : Journal of the Association of American Medical Colleges*, *86*, 429–434. https://doi.org/10.1097/ACM.0b013e3181f41e8c

Stell, A., & Sinnott, R. (2012). The ENSAT registry: a digital repository supporting adrenal cancer research. *Studies in Health Technology and Informatics*, *178*, 207–12. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/22797043

Stell, A., Sinnott, R., & Jiang, J. (2010). Enabling secure, distributed collaborations for adrenal tumor research. *Studies in Health Technology and Informatics*, *159*, 259–63. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20543447

Stone, D. (1986). A method for the validation of data in a register. *Public Health*, *100*, 316–24.

Sung, N. S., Crowley, W. F., Genel, M., Salber, P., Sandy, L., Sherwood, L. M., … Rimoin, D. (2003). Central challenges facing the national clinical research enterprise. *JAMA*, *289*(10), 1278–87. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12633190

Swendeman, D., Ramanathan, N., Baetscher, L., Medich, M., Scheffler, A., Comulada, W. S., & Estrin, D. (2015). Smartphone Self-Monitoring to Support Self-Management Among People Living With HIV. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, *69*, S80–S91. https://doi.org/10.1097/QAI.0000000000000570

Tantsyura, V., Grimes, I., Mitchel, J., Fendt, K., Sirichenko, S., Waters, J., … Tardiff, B. (2010). Risk-Based Source Data Verification Approaches: Pros and Cons. *Drug Information Journal*, *44*(6), 745–756. https://doi.org/10.1177/009286151004400611

Tatara, N., Arsand, E., Skrøvseth, S. O., & Hartvigsen, G. (2013). Long-term engagement with a mobile self-management system for people with type 2 diabetes. *JMIR mHealth and uHealth*, *1*(1), e1. https://doi.org/10.2196/mhealth.2432

Teede, H., Deeks, A., & Moran, L. (2010). Polycystic ovary syndrome: a complex condition with psychological, reproductive and metabolic manifestations that impacts on health across the lifespan. *BMC Medicine*, *8*, 41. https://doi.org/10.1186/1741-7015-8-41

Tenenbaum, J. D. (2015). Personalized Medicine. In P. R. O. Payne, P. J. Embi, I. N. Sarkar, N. Shah, J. D. Tenenbaum, & A. B. Wilcox (Eds.), *Translational Informatics* (pp. 35–60). London: Springer.

Terzolo, M., Bovio, S., Pia, A., Reimondo, G., & Angeli, A. (2009). Management of adrenal incidentaloma. *Best Practice & Research. Clinical Endocrinology & Metabolism*, *23*(2), 233–43. https://doi.org/10.1016/j.beem.2009.04.001

The Jefferson Monticello. (2016). Information in the currency of democracy (Quotation). Retrieved December 18, 2016, from https://www.monticello.org/site/jefferson/information-currency-

democracy-quotation

Timmermans, C., Doffagne, E., Venet, D., Desmet, L., Legrand, C., Burzykowski, T., & Buyse, M. (2016). Statistical monitoring of data quality and consistency in the Stomach Cancer Adjuvant Multi-institutional Trial Group Trial. *Gastric Cancer*, *19*(1), 24–30. https://doi.org/10.1007/s10120-015-0533-9

Tomlinson, M., Rotheram-Borus, M. J., Swartz, L., & Tsai, A. C. (2013). Scaling Up mHealth: Where Is the Evidence? *PLoS Medicine*, *10*(2), e1001382. https://doi.org/10.1371/journal.pmed.1001382

Tudur Smith, C., Stocken, D. D., Dunn, J., Cox, T., Ghaneh, P., Cunningham, D., & Neoptolemos, J. P. (2012). The value of source data verification in a cancer clinical trial. *PloS One*, *7*(12), e51623. https://doi.org/10.1371/journal.pone.0051623

U.S. Department of Health and Human Services, Food and Drug Administration (FDA), Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), Center for Devices and Radiological Health (CDRH), Office of Good Clinical Practice (OGCP), & Office of Regulatory Affairs (ORA). (2013). *Guidance for Industry Oversight of Clinical Investigations — A Risk-Based Approach to Monitoring Guidance for Industry*.

Vaidyanathan, G. (2012). Redefining clinical trials: the age of personalized medicine. *Cell*, *148*(6), 1079–80. https://doi.org/10.1016/j.cell.2012.02.041

van Helden, P. (2013). Data-driven hypotheses. *EMBO Reports*, *14*(2), 104. https://doi.org/10.1038/embor.2012.207

Venet, D., Doffagne, E., Burzykowski, T., Beckers, F., Tellier, Y., Genevois-Marlin, E., … Buyse, M. (2012). A statistical approach to central monitoring of data quality in clinical trials. *Clinical Trials*, *9*(6), 705–713. https://doi.org/10.1177/1740774512447898

Ventrice, T. (2013). 6 Critical Game Mechanics to Consider in Leaderboards: Pt 1. Retrieved May 6, 2015, from https://badgeville.com/content/6-critical-game-mechanics-to-consider-in-leaderboards-part-1

Vogel, A. L., Hall, K. L., Fiore, S. M., Klein, J. T., Bennett, L. M., Gadlin, H., … Falk-Krzesinski, H. J. (2013). The Team Science Toolkit: enhancing research collaboration through online knowledge sharing. *American Journal of Preventive Medicine*, *45*(6), 787–9. https://doi.org/10.1016/j.amepre.2013.09.001

Walther, B., Hossin, S., Townend, J., Abernethy, N., Parker, D., & Jeffries, D. (2011). Comparison of electronic data capture (EDC) with the standard data capture method for clinical trial data. *PloS One*, *6*(9), e25348. https://doi.org/10.1371/journal.pone.0025348

Wand, Y., & Wang, R. Y. (1996). Anchoring data quality dimensions in ontological foundations. *Communications of the ACM*, *39*(11), 86–95.

Wang, R. W., & Strong, D. M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, *12*(4), 5. https://doi.org/10.2307/40398176

Weber, D., Presser, K., & Norrie, M. (2015). How to give Feedback on Data Quality: A Study in the Food Sciences. In *ECIS 2015 Proceedings* (pp. 1–14). ECIS 2015 Completed Research Papers. Retrieved from http://aisel.aisnet.org/ecis2015_cr/195

Weiskopf, N. G., & Weng, C. (2013). Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *Journal of the American Medical Informatics Association*, *20*(1), 144–151. https://doi.org/10.1136/amiajnl-2011-000681

Wild, C. P., Bucher, J. R., de Jong, B. W. D., Dillner, J., von Gertten, C., Groopman, J. D., … McLaughlin,

J. (2015). Translational cancer research: balancing prevention and treatment to combat cancer globally. *Journal of the National Cancer Institute*, *107*(1), 353. https://doi.org/10.1093/jnci/dju353

Woodcock, J., & Woosley, R. (2008). The FDA critical path initiative and its influence on new drug development. *Annual Review of Medicine*, *59*, 1–12. https://doi.org/10.1146/annurev.med.59.090506.155819

Woolf, S. H. (2008). The meaning of translational research and why it matters. *JAMA : The Journal of the American Medical Association*, *299*(2), 211–3. https://doi.org/10.1001/jama.2007.26

World Health Organization. (2011). mHealth: New horizons for health through mobile technologies. *Observatory*, *3*(June), 66–71. https://doi.org/10.4258/hir.2012.18.3.231

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science (New York, N.Y.)*, *316*(5827), 1036–9. https://doi.org/10.1126/science.1136099

Wurst, S. H. R., Lamia, G., Schlundt, J., Karlsen, R., & Kuhn, K. a. (2008). A Service-Oriented Architectural Framework for the Integration of Information Systems in Clinical Research. *2008 21st IEEE International Symposium on Computer-Based Medical Systems*, 161–163. https://doi.org/10.1109/CBMS.2008.91

Wynants, L., Timmerman, D., Bourne, T., Van Huffel, S., & Van Calster, B. (2013). Screening for data clustering in multicenter studies: the residual intraclass correlation. *BMC Medical Research Methodology*, *13*(1), 1–24. https://doi.org/10.1186/1471-2288-13-128

Yin, H. L., Gabrilove, J., Jackson, R., Sweeney, C., Fair, A. M., & Toto, R. (2015). Sustaining the Clinical and Translational Research Workforce. *Academic Medicine*, *90*(7), 861–865. https://doi.org/10.1097/ACM.0000000000000758

Zerhouni, E. a. (2007). Translational research: moving discovery to practice. *Clinical Pharmacology and Therapeutics*, *81*(1), 126–8. https://doi.org/10.1038/sj.clpt.6100029

Zerhouni, E. A. (2003). A New Vision for the National Institutes of Health. *Journal of Biomedicine and Biotechnology*, *2003*(3), 159–160. https://doi.org/10.1155/S1110724303306023

Zerhouni, E. A. (2005). Translational and clinical science--time for a new vision. *The New England Journal of Medicine*, *353*(15), 1621–3. https://doi.org/10.1056/NEJMsb053723

Zozus, M. N., Hammond, W. E., Green, B. B., Kahn, M. G., Richesson, R. L., Rusincovitch, S. a, … Smerek, M. M. (2014). *Assessing Data Quality for Healthcare Systems Data Used in Clinical Research*. Durnham. Retrieved from file:///C:/Users/anobles/Downloads/Assessing-data-quality_V1 0 (1).pdf

APPENDICES