

NPL REPORT MS 31

NMS 2018-2021 LIFE-SCIENCES AND HEALTHCARE PROJECT  
“DIGITAL HEALTH: CURATION OF HEALTHCARE DATA” - FINAL  
REPORT

NADIA SMITH, MARINA ROMANCHIKOVA, IGNACIO PARTARRIEU,  
ELIZABETH COOKE, AGNIESZKA LEMANSKA AND SPENCER  
THOMAS

MARCH 2021



NMS 2018-2021 Life-Sciences and Healthcare Project  
“Digital Health: Curation of Healthcare Data”  
Final Report

Nadia Smith, Marina Romanchikova, Ignacio Partarrieu, Elizabeth  
Cooke, Agnieszka Lemanska and Spencer Thomas

Data Science Department

## ABSTRACT

NPL has been running an inter-disciplinary project on Digital Health data curation and analysis for the last three years. The project aims to demonstrate how a ‘metrological mindset’ applied to the curation and analysis of healthcare data and metadata could help solve some of the important and emerging challenges of utilising healthcare data, addressing at the same time one of the key challenges of the UK’s Measurement Strategy: to provide confidence in the intelligent and effective use of data. This report summarises the main findings of the project.

© NPL Management Limited, 2021

ISSN 1754-2960

<https://doi.org/10.47120/npl.MS31>

National Physical Laboratory  
Hampton Road, Teddington, Middlesex, TW11 0LW

This work was funded by the UK Government's Department for Business, Energy and Industrial Strategy (BEIS) through the UK's National Measurement System programmes.

Extracts from this report may be reproduced provided the source is acknowledged and the extract is not taken out of context.

Approved on behalf of NPLML by  
Louise Wright  
Head of Science  
Data Science Department



## CONTENTS

<b>TERMINOLOGY.....</b>	<b>I</b>
<b>ABBREVIATIONS .....</b>	<b>I</b>
<b>EXECUTIVE SUMMARY .....</b>	<b>III</b>
<b>1 INTRODUCTION .....</b>	<b>1</b>
1.1 MEASUREMENT CHALLENGE SUMMARY .....	1
1.2 MULTI-MODAL DATA AND METADATA .....	2
1.3 ONTOLOGIES AND DATA CURATION .....	3
1.4 THE LONG-TERM VALUE OF DATA .....	4
1.5 STATE-OF-THE-ART .....	5
1.6 PROJECT ACHIEVEMENTS .....	9
<b>2 CASE STUDY 1: VOXTOX .....</b>	<b>11</b>
2.1 CASE STUDY PROPOSAL .....	11
2.2 CASE STUDY TECHNICAL WORK .....	13
2.3 CASE STUDY OUTCOMES/HIGHLIGHTS .....	16
2.4 FUTURE DIRECTION OF CASE STUDY .....	17
<b>3 CASE STUDY 2: ECHO .....</b>	<b>20</b>
3.1 CASE STUDY PROPOSAL .....	20
3.2 CASE STUDY TECHNICAL WORK .....	22
3.3 CASE STUDY OUTCOMES/HIGHLIGHTS .....	32
3.4 FUTURE DIRECTION OF CASE STUDY .....	33
<b>4 CASE STUDY 3: COMPUTERISED MEDICAL RECORDS (CMR) .....</b>	<b>36</b>
4.1 CASE STUDY PROPOSAL .....	36
4.2 CASE STUDY TECHNICAL WORK .....	37
4.3. OTHER RELATED PROBLEMS .....	42
4.4 CASE STUDY OUTCOMES/HIGHLIGHTS .....	48
4.5 FUTURE DIRECTION OF CASE STUDY .....	49
<b>5 CASE STUDY 4: SYNTHETIC DATA .....</b>	<b>50</b>
5.1 CASE STUDY PROPOSAL .....	50
5.2 CASE STUDY TECHNICAL WORK .....	51
5.3 CASE STUDY OUTCOMES/HIGHLIGHTS .....	60
5.4 FUTURE DIRECTION OF CASE STUDY .....	60
<b>6 CASE STUDY 5: MENTAL HEALTH AND COVID-19 .....</b>	<b>62</b>
6.1 CASE STUDY PROPOSAL .....	62
6.2 CASE STUDY TECHNICAL WORK .....	62
6.3 CASE STUDY OUTCOMES/HIGHLIGHTS .....	67
6.4 FUTURE DIRECTION OF CASE STUDY .....	67

<b>7</b>	<b>MAIN OUTPUTS AND LEARNINGS.....</b>	<b>68</b>
7.1	LIST OF PEER-REVIEWED PUBLICATIONS .....	68
7.2	WHAT HAS BEEN ACHIEVED OVER THE THREE YEARS .....	68
7.3	LEARNING POINTS FOR FUTURE WORK.....	69
<b>8</b>	<b>ACKNOWLEDGMENTS.....</b>	<b>70</b>
<b>9</b>	<b>REFERENCES .....</b>	<b>71</b>

## TERMINOLOGY

Data Curation	Organization and integration of data collected from various sources, annotation of the data, and publication and presentation of the data such that the value of the data is maintained over time, and the data remains available for reuse and preservation.
Data Interoperability	Addresses the ability of systems and services that create, exchange and consume data to have clear, shared expectations for the contents, context and meaning of that data.
Data Integration	Aggregation of datasets from heterogeneous sources by linking, combination or fusion.
Data Provenance	A history of the dataset that captures its origin, purpose and all modifications it underwent since its creation.
FAIR principles	A set of guiding principles to make data Findable, Accessible, Interoperable and Re-usable.
Measurement	The process of experimentally obtaining one or more quantity values that can reasonably be attributed to a quantity.
Metadata	A set of descriptors of the dataset that make the dataset easier to locate, understand and use. Metadata can include information about the purpose of the dataset creation, the experimental setup or the data formats. Depending on the nature of the investigation, metadata can be data in its own right.
Ontology	A formal representation of a knowledge domain that comprises a vocabulary of terms, as well as relations between them. An ontology can be presented as a knowledge graph where terms are depicted as vertices and relationships as edges.
Reproducibility	Condition of measurement, out of a set of conditions that includes different locations, operators, measuring systems, and replicate measurements on the same or similar objects.
Traceability	Property of a measurement result whereby the result can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty.
Uncertainty of Measurement	The parameter characterizing the dispersion of the quantity values being attributed to a measurand

## ABBREVIATIONS

AI	Artificial Intelligence
ANCOVA	Analysis of Covariance
CMR	Computerised Medical Records
CNN	Convolutional Neural Network
COV	Coefficient of Variations
CT	Computed Tomography
DAE	Deep Autoencoder
DICOM	Digital Imaging and Communications in Medicine
ECHO	Effect of cancer and its treatment on health outcomes
FAIR	Findable, Accessible, Interoperable and Re-usable
FOV	Field of View
GAN	Generative Adversarial Network
GDPR	General Data Protection Regulation
GP	General Practice
HU	Hounsfield Units
ISIC	International Skin Imaging Collaboration

KNN	k-nearest neighbours
LDA	Linear Discriminant Analysis
ML	Machine Learning
MRI	Magnetic Resonance Imaging
MVCT	Mega-voltage Computed Tomography
NHS	National Health Service
NHSX	Unit within the NHS that is driving forward the digital transformation of health and social care
NHSD	NHS Digital
NHSE/I	NHS England/Improvements
NLST	National Lung Screening Trial
NMI	National Measurement Institute
NMS	National Measurement System
NPL	National Physical Laboratory
PEG	Programme Experts Group
PET	Positron Emission Tomography
PHE	Public Health England
RCGP RSC	Royal College of General Practitioners Research and Surveillance Centre
ROI	Region of Interest
RT	Radiotherapy
STAC	Science and Technology Advisory Council
SQL	Structured Query Language
SVM	Support Vector Machines
US	Ultrasound

## EXECUTIVE SUMMARY

Healthcare is increasingly and routinely generating large volumes of data from different sources, which are difficult to handle and integrate. Confidence in data can be established through the knowledge that the data is validated, well curated and free of bias or errors.

NPL has been running an inter-disciplinary project on Digital Health data curation and analysis for the last three years. The project aims to demonstrate how a 'metrological mindset' applied to the curation and analysis of healthcare data and metadata could help solve some of the important and emerging challenges of utilising healthcare data, addressing at the same time one of the key challenges of the UK's Measurement Strategy: to provide confidence in the intelligent and effective use of data.

Establishing confidence in the data is necessary for establishing trustworthiness and traceability, not only in the data itself, but also in the analysis and interpretation methods. This is essential in the current era where clinical decisions are increasingly being supported by a range of data from different sources.

In the first year of the project, a landscaping activity was carried out and a stakeholder workshop was held to decide the areas to work on, resulting in a peer-reviewed publication in the British Journal of Radiology<sup>1</sup>.

During its second year, the project focussed on four pilot case studies that were carried out in collaboration with clinical partners: 1) integration of equipment calibration data with medical images; 2) ontology-based linkage of primary and secondary care datasets for prostate cancer patients; 3) development of data-driven models to identify key prognostic markers in healthcare records; and 4) generation of synthetic datasets with associated data quality metrics. With each of these case studies the aim was to identify the key data streams and associated measurement modalities, as well as current practice and areas for improvement.

The plans for the final year had to be adapted due to the Covid-19 pandemic. Half of the resource from this project was reallocated to help with the pandemic response work. Nevertheless, in the final year the main project deliverables were achieved and a new case study related to mental health during the first lockdown was developed. The project now has its own website <https://www.npl.co.uk/research/data-science/digital-health>, which is used to promote the work carried out as well as to engage with new collaborators and extend the project's network.

This report is a compilation of the work carried out over the three years of the project. Section 1 introduces the topic area; Sections 2 to 6 describe the work carried out for each of the case studies and summarises the findings as well as future work; and Section 7 draws the main conclusions of this work.



## 1 INTRODUCTION

NPL has been running an inter-disciplinary project on Digital Health for the last three years. The project aims to understand the role NPL can play within the current Digital Health national landscape and to outline how a 'metrological mindset' applied to the curation and analysis of healthcare data and metadata could help solve some of the important and emerging challenges of utilising healthcare data, as will be described in what follows.

At the same time this project addresses one of the key challenges of the UK's Measurement Strategy, to provide confidence in the intelligent and effective use of data. Establishing confidence in the data is necessary for establishing trustworthiness and traceability, not only in the data itself, but also in the analysis and interpretation methods. This is essential in the current era where clinical decisions are increasingly being supported by a range of data from different sources.

### 1.1 MEASUREMENT CHALLENGE SUMMARY

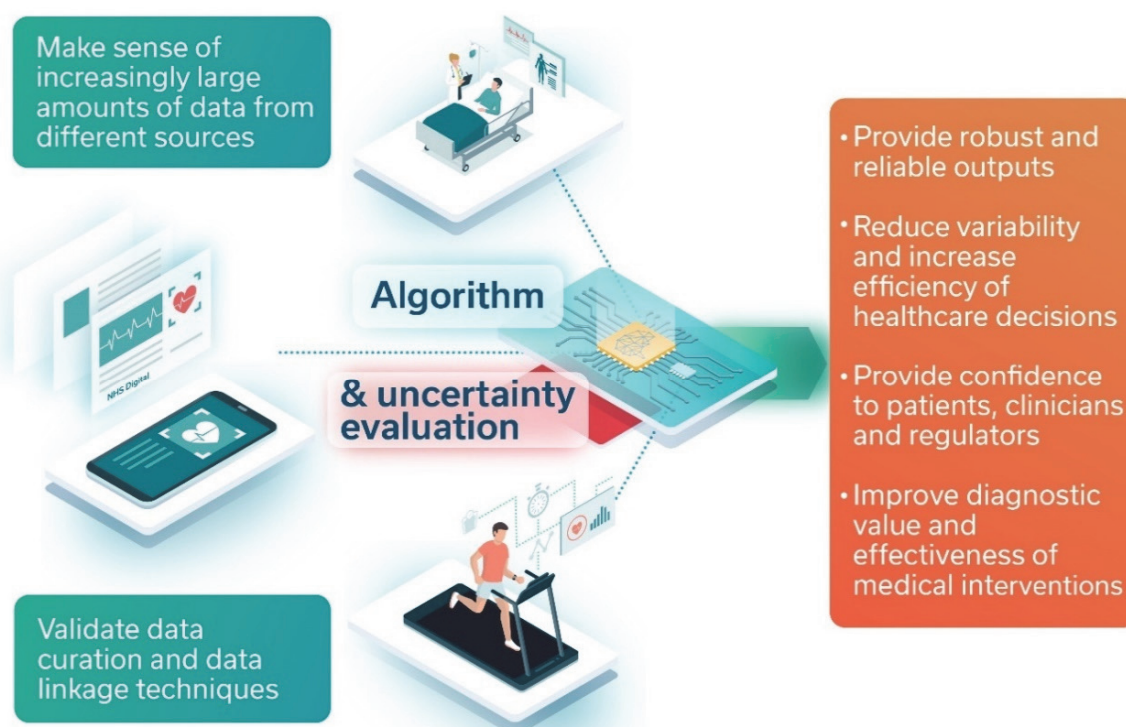
The digital revolution and new measurement modalities within the medical area are creating increasingly vast amounts of high-dimensional data coming from disparate sources and of varying quality. These data may include genetics, genomics, imaging, blood tests, electronic healthcare records, and data from wearable devices. Accessing and linking different sources of medical data is a difficult task due to the multi-modal and confidential nature of the data. However, healthcare increasingly relies on the integration of large datasets and trusted and robust tools to analyse the data.

The integration of large volumes of healthcare data coming from multiple sources presents several metrology challenges. Initially, there is the definition of integration, which may include, but not be limited to, the data being combined (additively, concatenated, fused) or linked (association, either directionally or hierarchically). The context of this integration can pose some measurement issues such as combining images from different scanners or settings and any pre-processing or normalisation that may be required prior to this. A further challenge is the management of associated metadata (from multiple sources) which, if done correctly, should enable the data to comply with the FAIR principles<sup>2,3</sup>.

The context of the integration will impact on the management of the metadata. Including assessments of data quality through calibration information and quality metrics will allow for better data analysis. Ensuring the data are trusted and interoperable, as well as including quality and uncertainty measures will improve confidence in analysis of healthcare data as well as expanding the possibilities for what analyses can be performed. Furthermore, if the quality of the data is known and considered, it could lead to improved quality of healthcare at the population and individual patient levels. For example, in multi-centre medical imaging studies, secure data sharing between centres must include a pre-agreed definition of mandatory metadata to capture data provenance and acquisition protocols. These annotations result in clearer interpretability and comparability of datasets, enabling uncertainty estimation and future analysis. It must be noted data fusion and variations in data quality of individual datasets present important challenges to be addressed in the future.

Recent publications have recognised the pivotal role of metrology in increasing confidence in research results and ensuring reproducibility, saving resources and accelerating bench-to-bedside transition<sup>4,5,6</sup>. Metrology offers multiple tools to support data curation, increase re-use, improve linkage, and guarantee consistent quality with growing data volumes. These tools include inter-comparisons, reference datasets and standards with an emphasis on traceability and measures of uncertainty<sup>7,8</sup>.

Applying metrology tools to healthcare data curation will enable objective assessment of data quality, extend data lifetime, increase its utility, and ensure its compliance with FAIR principles. Furthermore, metrology can also help in providing confidence in the outputs of an analysis of the data, ensuring traceability and reliability, and by quantifying the uncertainty. Implementation of best practices in data curation and analysis will support advancement of new tools and help identify predictors of clinical outcomes; reduce variability in decision making; provide confidence to patients, clinicians and regulators; and ultimately, improve the diagnostic value and effectiveness of medical interventions (see Figure 1.1).



**Figure 1.1 Measurement challenge summary**

## 1.2 MULTI-MODAL DATA AND METADATA

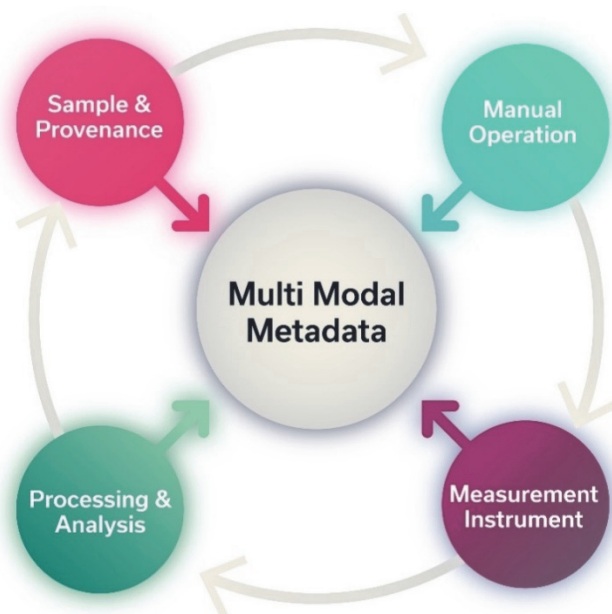
Linking multi-modal datasets introduces a new challenge of combination of metadata from these datasets. A combination is required if the linked data are to be part of a curated database for long term storage, otherwise potentially vital pieces of information may be missing for future analysis or attempts to reproduce data.

Curated data is an essential requirement for reproducible science as it allows datasets to be correctly grouped or distinguished based on their attributes prior to analysis. This information enables analysis of metadata to determine whether a difference in the data is the result of a real difference or from inappropriately comparing ‘apples’ and ‘oranges’. For problems that are less precise, such as noise measurements or biological systems, this may also provide a means to add confidence or uncertainty bounds on the input data and to link this with uncertainty propagation through the data analysis pipeline. Appropriately curated data allows the validation of any differences observed in endpoint data analysis arising from the system rather than differences in provenance by comparing the traceable metadata.



The challenge of data curation in measurement science is far from trivial and is often a multi-source problem. In the healthcare and biomedical sciences context, metadata are usually multi-modal. The sources of multi-modal metadata can be described by four categories, each with possibly multiple sources. In some specific contexts there may be no sources for a given category but conceptually they exist. These sources are metadata relating to the:

- 1) Sample (e.g. patient, provenance of the sample up to measurement);
- 2) Manual operational metadata (e.g. sample preparation, operator influence on the measurement, deviation from protocols, etc);
- 3) Instrument performing the measurement (type, model, calibration records, settings, etc);
- 4) Post-measurement processing, handling and analysis (conversion, normalisation, noise reduction, data mining, software versions and libraries used, etc).



**Figure 1.2 Sources of multi-modal metadata**

These sources are summarised in Figure and a complete record of all these multi-modal sources of metadata are required in order to replicate a scientific study and assess the reproducibility of the results. Differences in any part could lead to erroneously differing outputs from analyses, and a well-curated integration of these sources would ensure traceability to identify whether the source of variance in the data are real or a result of changes in the multimodal metadata.

### 1.3 ONTOLOGIES AND DATA CURATION

The lack of usage and inconsistency of usage of healthcare domain ontologies often makes medical data less usable and their access less systematic and non-replicable. An ontology can provide a systematic replicable method for clinical case definition by supplying domain researchers with a common vocabulary, thus aiding data retrieval, interoperability and curation. Ontological approaches have been successfully developed to identify cases of particular diseases from routine primary care data<sup>9</sup>.

Some efforts to ensure that the data curation complies with FAIR principles have been implemented<sup>10,11</sup> but there are still necessary future developments when applying these principles to medical datasets: issues around data accessibility and data privacy make the

findable and accessible principles difficult to comply with; the lack of use of common vocabulary, standards and information models such as HL7 FHIR<sup>12</sup> (Fast Healthcare Interoperability Resources) make healthcare data less interoperable; and multi-modal metadata is not always recorded or captured.

#### 1.4 THE LONG-TERM VALUE OF DATA

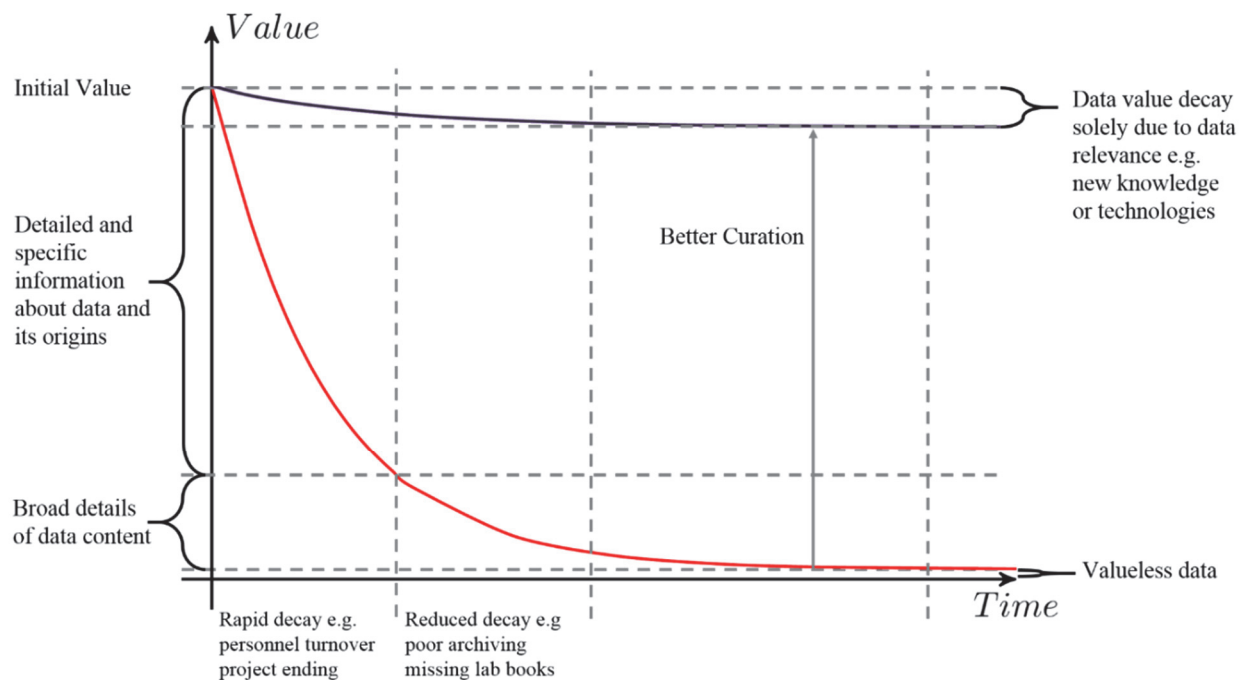
The value of data is determined not only by its contents, but also by the knowledge about its generation and by the number of possibilities for re-use. In this age of Big Data, large volumes of complex and heterogeneous data can be obtained from many domains, and the quality of these data is extremely variable. Examples of such data in the context of digital health data are electronic medical records, medical imaging data, genomic data, physiological test results, etc. The impact that can be derived from these data elevate their value. However, the lack of structure, standardisation, provenance or knowledge of source reduce these enormous volumes of data to small subsets of 'useable' data.

In the absence of knowledge of the content or context of data we are limited to system level metadata (number and size of files, format etc.), making them 'valueless' outside of the originating domain. In the business context, the term "data half-life" (borrowed from nuclear physics) is used to describe the data depreciation, as data relevance decreases (e.g., in response to stock prices or media coverage). This process is illustrated in Figure 1.3. In healthcare settings, the data life timescales are large due to the costs (time and monetary) in planning and acquiring these data, and even though the data's relevance decreases over time because changing practices and standards mean it is no longer up to date or captures the best possible information, it is at a comparatively slow rate of reduction to other industry areas. For example, a large collection of labelled oncology data is extremely valuable, and many institutions in research and industry are using them to develop methods to improve detection and diagnosis. Furthermore, such a dataset will continue to be valuable for years, if not decades, as a benchmark dataset.

Several benchmark datasets are often used to evaluate the performance of new algorithms against the current state of the art in computer science, such as the MNIST dataset<sup>13</sup> (27,300 results in Google Scholar). Examples of data for healthcare applications are the Camelyon Grand Challenge<sup>14,15</sup> consisting initially of 400 annotated whole slide images of annotated pathology data, taking around 1 hour per slide for annotation by experienced pathologists<sup>16</sup>. A Google Scholar search returns 143 results for "Camelyon" that use or cite the available data, 135 of which also include machine or deep learning. Due to the importance and size of this dataset, now 1399 annotated whole slide images of pathology data<sup>17</sup>, the original paper was been viewed more than 38,000 times in less than 2 years and the second grand challenge, Cameylon17, had 444 participants<sup>18</sup>, and many users since the challenge ended.

In a healthcare context, information about data provenance is likely to get lost or degrade over time. The common causes for the loss of data provenance include end of project lifetime, focus shifting to other areas, personnel change and turnover, loss or corruption of records due to incompatible file formats (see, e.g., the famous Covid-19 excel scandal<sup>19</sup>) etc. For example, a dataset of equal numbers of healthy patients, early-stage disease and late-stage disease patients with a full history as well as outcome data is extremely valuable. If we remove the patient history and outcomes, the data remains valuable (albeit less so). However, if we remove the disease staging information, we lose the information on disease stage indicators and differentiators – two of three subgroups of patients are no longer differentiable. While the data would remain valuable, we would have lost a significant portion of its utility. Having well-curated data prevents the decay in data value, as the metadata includes the traceable history of the data maintaining this value for future use and information retrieval. The data then are only limited by the relevance and usefulness to the context, e.g.

resolution, quality, experimental design, knowledge of system, technological advancements etc.



**Figure 1.3 Data value decay due to poor curation**

## 1.5 STATE-OF-THE-ART

Over the last decade the way we treat and think about medical data has changed dramatically. The digital revolution and new measurement modalities in healthcare are creating increasingly vast amounts of high-dimensional data which in turn given rise to big data, allowing in-depth analysis leading to predictive modelling and artificial intelligence (AI) for precision medicine and the advancement of healthcare and healthcare productivity across the globe.

In the context of healthcare, large quantities of such readily available data presents three key challenges:

1. Identification of dataset fitness-for-purpose to answer a clinical question with respect to its content and quality;
2. Safeguarding sensitive patient information from misuse and malpractice;
3. Linkage and sharing of this information in a standardized manner to fully realise the benefits for research and development.

Health systems across the world are stretched beyond their limits due to a growing and ageing population accompanied by a decrease in the available healthcare funding<sup>20</sup>. New technologies and alternative data sources such as in-home monitoring devices for isolated patients could provide a much-needed advancement to improve health services on a large scale<sup>21,22</sup>. However, cohesive and secure structures must be put in place for the protection of patients and the ease of access for clinicians, policy makers and multidisciplinary researchers to avoid issues such as hacking attacks on healthcare service providers<sup>23,24</sup> or malicious data tampering of medical equipment<sup>25</sup>. Furthermore, the easy availability and high variety of data from wearable devices is accompanied by issues related to the data quality. Firstly, these data lack any form of standardisation. Secondly, only a small proportion of

wearable device measurements are validated by the vendor, and vanishingly few are accredited by a National Measurement Institute (NMI)<sup>26</sup>.

### 1.5.1 The approach to Digital Healthcare taken by the NHS

The Secretary of State for Health and Social Care commissioned “*The Topol Review: Preparing the healthcare workforce to deliver the digital future*”<sup>27</sup>, as part of the draft health and care Workforce Strategy for England to 2027 – Facing the Facts, Shaping the Future. The Topol Review was published in February 2019 and makes recommendations that will enable NHS staff to make the most of innovative technologies such as genomics, digital medicine, artificial intelligence and robotics to improve services. These recommendations support the aims of the NHS Long-Term Plan<sup>28</sup>, and the workforce implementation plan, helping to ensure a sustainable NHS.

In addition to the Topol Review, in July 2019, building on the NHS Long-Term Plan, NHSX<sup>29</sup> was launched to deliver the Secretary of State for Health and Social Care’s Tech Vision<sup>30</sup>. The role of NHSX is to speed up the digital transformation of the NHS and social care allowing patients and staff to benefit from the latest digital systems and technology. NHSX’s responsibilities include:

- setting national policy and developing best practice for NHS technology, digital and data - including data-sharing and transparency;
- setting standards – developing, agreeing and mandating clear standards for the use of technology in the NHS;
- ensuring that NHS systems can talk to each other across the health and care system;
- helping to improve clinical care by delivering agile, user-focused projects;
- supporting the use of new technologies by the NHS, both by working with industry and via its own prototyping and development capability;
- ensuring that common technologies and services, including the NHS App, are designed so that trusts and surgeries don’t have to “reinvent the wheel” each time;
- making sure that all source code is open by default so that anyone who wants to write code for the NHS can see what is needed;
- reforming procurement – helping the NHS buy the right technology through the application of technology standards, streamlined spend controls and new procurement frameworks that support their standards;
- setting national strategy and mandating cyber security standards, so that NHS and social care systems have security designed in from the start;
- championing and developing digital training, skills and culture so NHS staff are digital-ready;
- delivering an efficient process for technology spend, domain name management and website security.

In October 2019, NHSX published policy documents about the use of AI in healthcare<sup>31</sup>, giving a considered and cohesive overview of the current state of play of data-driven technologies within the health and care system, covering everything from the local research environment to international frameworks in development. In order to put policy into practice, the government has invested in the creation of the NHS AI Lab<sup>32</sup>, run collaboratively by NHSX and the NHS Accelerated Access Collaborative (AAC)<sup>33</sup>, which has been set up to support innovation in an open environment where innovators, academics, clinicians and others can develop, learn, collaborate and build technologies at scale to deliver maximum impact in health and care safely and effectively.

Alongside NHSX, there are other national NHS agencies with digital responsibilities: NHS Digital<sup>34</sup> and the NHS England/Improvement regional teams<sup>35</sup>, all of which existed prior to NHSX. The role of NHS Digital, as the national information and technology and data partner for the health and care system, has increased significantly since 2017<sup>36</sup>. Their vision is to harness the power of information and technology to make health and care better for the care professionals and people who work in the service, and for the people who receive them, by:

- ensuring that every citizen's data is protected;
- establishing shared architecture and standards so everyone benefits;
- implementing national services to meet national and local need;
- supporting care organisations to get the best out of technology, data and information;
- enabling much better use of health and care information.

The NHS England/Improvement regional teams are responsible for the quality, financial and operational performance of all NHS organisations in their region, drawing on the expertise and support of their corporate teams to improve services for patients and support local transformation. They also support the identity and development of sustainability and transformation partnerships and integrated care systems.

In July 2020 the Secretary of State for Health and Social Care commissioned<sup>37</sup> a high-level review of the digital responsibilities of different NHS agencies, led by NHS Digital chair, Laura Wade-Gery with consultants McKinsey & co. The intentions of the review are to determine the capabilities and digital operating model across NHS Digital, NHSX and NHS England/Improvement regional teams needed to drive the wider system transformation envisaged in the NHS long-term plan<sup>38</sup>. The findings of the review seem to be delayed so far, probably due to the ongoing pressure related to Covid-19 and the new vaccination programme<sup>39</sup>. Regardless of the distribution of responsibilities, NHSX and NHS Digital are playing vital roles during the Covid-19 pandemic, both strategically<sup>40</sup> and in a delivery support role<sup>41</sup>.

It is clear that a great deal of attention is currently being given to NHS policy to adjust it in response to the rapidly growing use of data. A recent government publication<sup>42</sup> identified key challenges in:

- ensuring the use of legacy technology and commercial arrangements;
- complexity of delivery structures;
- reducing risk;
- limited resources;
- gaining public trust.

In response to such issues, guiding principles were presented to map out priorities that should:

- encompass the storage and handling of data: user need;
- maintain privacy and security good practice following guidance from GDPR<sup>43</sup>, National Data Guardian for health and social care<sup>44</sup>, NHS DH Toolkit<sup>45</sup>, and guide to good practice for digital and data-driven health technologies<sup>46</sup>;
- achieve interoperability and openness - standardization will reduce cost and can take advantage of large infrastructure by sharing knowledge in a more granular and dynamic manner;
- ensure inclusion and accessibility of systems.

In summary, the NHS has already set in motion policies to keep abreast of the rapidly growing development of data collection and data analysis. It has therefore never been more imperative to obtain a clear understanding of how we can best utilize data in research whilst maintaining the integrity of the patient data.

## 1.5.2 Digital Health initiatives within the UK

There are currently a wide range of Digital Health initiatives and Health Data networks and programmes in the UK. A non-exhaustive list is the following:

- Health Data Research UK (HRDUK)<sup>47,48</sup> - the national institute for health data science. HDRUK are an independent, non-profit organisation that bring together over 20 research institutes across the UK. They are building a data innovation engine for the NHS, academia and industry to accomplish their vision for large scale data and advanced analytics to benefit every patient interaction, clinical trial, biomedical discovery and enhance public health. HDRUK have set up the Research Alliance<sup>49</sup> (an independent alliance of leading healthcare and research organisations united to establish best practice for the ethical use of UK health data for research at scale), the Research Hubs<sup>50</sup> (to curate disease focussed clinical trials and gather real-world evidence), and the Research Innovation Gateway<sup>51</sup> to find and access data.
- The Academic Health Science Networks (AHSNs)<sup>52</sup> – the role of the AHSNs is to connect the NHS and academic organisations, local authorities, the third sector and industry, to act as catalysts that create the right conditions to facilitate change across whole health and social care economies, with a clear focus on improving outcomes for patients.
- Digital Health London<sup>53</sup> – a programme aiming to speed up the development and scaling of digital innovations across health and care and pioneer their adoption by the NHS. They match innovators with NHS need, and support them to navigate the UK health environment.
- OneLondon<sup>54</sup> - one of the country's first Local Health and Care Record Exemplars (LHCRE), designated by NHS England. The OneLondon LHCRE is a partnership of NHS organisations and local government across London, working together with citizens to transform London's health and care services by joining up information to support fast, effective and safe care.
- MedCity<sup>55</sup> – an initiative launched in 2014 to promote and grow the world-leading life sciences cluster of England's Greater South East by promoting life sciences investment, entrepreneurship and industry in the region.
- Innovate UK centres for excellence of digital pathology and imaging<sup>56</sup>– funded through the Industrial Strategy Challenge Fund (ISCF), five new centres of excellence for digital pathology and imaging, including radiology, using AI medical advances were selected in November 2018. The centres will use AI, an area the government is backing in its modern Industrial Strategy, to find new ways to speed up diagnosis of diseases to improve outcomes for patients. The centres are:
  - The London Medical Imaging & AI Centre for Value-Based Healthcare<sup>57</sup> applies AI technologies to medical imaging delivery, interpretation, and reporting in MRI, CT, PET and ultrasound scans.
  - I-CAIRD, Glasgow (Industrial Centre for AI Research in Digital Diagnostics)<sup>58</sup> brings together clinicians, health planners, and industry to work with innovative SMEs to answer clinical questions and solve healthcare challenges more quickly and efficiently.
  - NCIMI (National Consortium of Intelligent Medical Imaging)<sup>59</sup> in Oxford brings together world-leading clinicians, NHS hospitals, industry experts in clinical imaging and artificial intelligence, charities and patient groups. Their goal is to build a pipeline for innovation to allow new medical imaging AI tools to be developed, tested, validated and adopted into the NHS.
  - The Northern Pathology Imaging Collaborative (NPIC)<sup>60</sup> located in Leeds is a unique collaboration between NHS, Academic and Industry partners that deploys digital pathology across hospitals in the North of England and will develop artificial intelligence tools to help diagnose cancer and other diseases

- The Pathology image data Lake for Analytics, Knowledge and Education (PathLAKE)<sup>61</sup> based in Coventry comprises some of the nation's leading digital and computational innovators from NHS and academia. It is working on the digitisation of five major NHS laboratories and the formation of a computational pathology hub, to drive AI innovation in pathology for the UK.
- Digital Health and Care Institute (DHI)<sup>62</sup> - is part of the Scottish Funding Council's Innovation Centre Programme, which is designed to support transformational collaboration between universities and businesses. Their networks, reach and capabilities are able to bring the right people together and provide them with the means to identify, design, evaluate and invest in new solutions to the country's prioritised health and care challenges.
- Alan Turing Institute (ATI) – Health and medical sciences programme<sup>63</sup> - The ATI's health programme will deliver a major programme of research into the theory and methods of AI, statistics, and data analytics underpinning medical and health applications that will enable scientists to do better science, without compromising respect for privacy and patient trust.
- Ada Lovelace Institute<sup>64</sup> - although not uniquely dedicated to healthcare problems, the Ada Lovelace Institute is worth mentioning as their mission of ensuring that data and AI work for people and society is relevant.
- Covid-19 related initiatives: during 2020, a wealth of initiatives to manage data and facilitate research activity on Covid-19 nationally and globally were set up. Most of the initiatives listed above have a separate Covid-19 dedicated area, such as the NHSX Covid-19 response<sup>65</sup>, the NHS Digital response to coronavirus<sup>66</sup>, HDRUK's Covid-19 data strategy<sup>67</sup>, and the Covid Corpus initiative<sup>68</sup>.

## 1.6 PROJECT ACHIEVEMENTS

### 1.6.1 Project plans and deliverables

In the first year of the project, a landscaping activity was carried out and an external stakeholder workshop was held to decide the areas to work on. The one-day stakeholder workshop was held at NPL's Cambridge hub in the Maxwell Centre on Wednesday 6th February 2019. Over 30 external stakeholders attended, including NHS professionals, scientists, academics, and industry partners. The aims of the day were to:

- listen to issues and experiences of stakeholders relating to curation of healthcare datasets
- listen to stakeholders' thoughts on what software tools NPL could help develop towards a standardised platform for curation and linkage of datasets
- identify pilot case studies and datasets that NPL could start working on

The workshop opened with presentations from three keynote speakers: Richard Irvine (Head of Data Management at NHS Digital), Sara Faithfull (Professor of Cancer Nursing Practice, Lead for Clinical Innovation at University of Surrey) and Harry Witchel (Discipline Leader in Physiology, Trafford Centre, Brighton and Sussex Medical School). The talks were followed by four breakout group discussions on the data issues, potential partnerships and case studies. The workshop's findings were refined in a roundtable discussion with all participants. All the discussed topics resonated well with the stakeholders, and the outcomes of the day provided new insights and ideas for the next year of the project as well as for defining NPL's role within the Digital Health umbrella, resulting in a peer-reviewed positioning paper in the British Journal of Radiology<sup>69</sup>.

During its second year, the project focussed on four case studies that were carried out in collaboration with clinical partners:



1. integration of equipment calibration data with medical images;
2. ontology-based linkage of primary and secondary care datasets for prostate cancer patients;
3. development of data-driven models to identify key prognostic markers in healthcare records;
4. generation of synthetic datasets with associated data quality metrics.

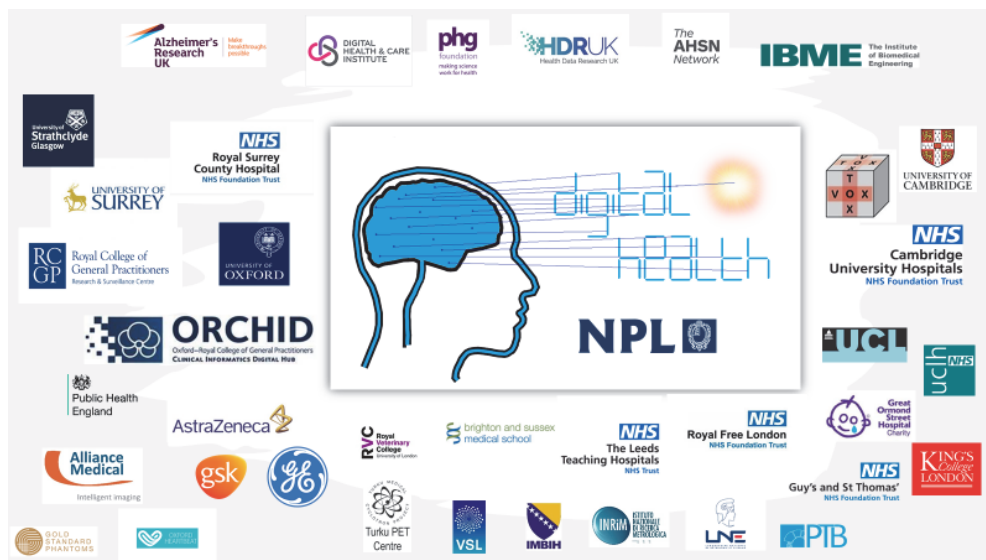
With each of these case studies, the aim was to identify the key data streams and associated measurement modalities, as well as current practice and areas for improvement, focusing on increasing traceability and reducing uncertainty.

The plans for the final year had to be adapted due to the Covid-19 pandemic. Half of the resource from this project was reallocated to help with the pandemic response work<sup>70,71</sup>. Nevertheless, in the final year the main project deliverables were achieved and a new case study related to mental health during the first national lockdown was developed. The project now has its own website<sup>72</sup>, which is used to promote the work carried out as well as to engage with new collaborators and extend the project's network.

This report is a compilation of the work carried out over the three years of the project. Sections 2 to 6 describe the work carried out for each of the case studies and summarises the findings as well as future work; and Section 7 draws the main conclusions of this work.

### 1.6.2 Collaborators and Stakeholders

An important role for NPL in this area is to act as a facilitator between institutions as an independent organisation and we have focussed efforts in maintaining and extending our networks throughout the year (see Figure 1.4 for a summary of partnerships and other relationships). We are also working with our network of contacts on other projects in the healthcare area through other NMS-funded projects as well as via other sources of funding (such as EMPIR, A4I), which leverages the funding from this area.



**Figure 1.4 Map of NPL's collaborators and stakeholders in the Digital Health area**



## 2 CASE STUDY 1: VOXTOX

**Full title:** VOXTOX: IMPROVING DATA QUALITY AND COMPARABILITY BY LINKING PATIENT IMAGE WITH IMAGING DEVICE CALIBRATION DATA

**Technical lead:** Marina Romanchikova

**Web:** <https://www.npl.co.uk/research/data-science/digital-health/case-study/voxtox>

Variations in medical image quality have been shown to impact image analysis and classification, regardless of whether these are performed by human observers, or by conventional image processing or machine learning methods<sup>73</sup>. We believe that by using imaging device quality assurance data it is possible both to reduce these variations and to improve the image traceability to a known physical property, which would lead to improved accuracy of radiotherapy treatment and patient outcomes.

### 2.1 CASE STUDY PROPOSAL

Megavoltage computed tomography (MVCT) images are a standard clinical utility in image-guided radiation therapy (IGRT) on TomoTherapy and Radixact devices<sup>74</sup>. Intensity values in these images represent electron densities in the imaged tissue. These values are a subject to equipment- and environment-induced fluctuations that can result in the variations of up to 5%<sup>75</sup>. To account for variations due to the medical device, clinical medical physicists undertake regular equipment calibration procedures to track the changes and ensure the stability of the imaging system<sup>76</sup>.

Calibration data are presented and stored as calibration curves reflecting the relation between the measurand (i.e. the tissue electron density) and the observation. The calibration history for each device is evaluated periodically by imaging physicists and stored for the device lifetime.

Although calibration data are vital for intra- and inter-device image comparability and adaptive dose re-calculation<sup>77</sup>, there are no mechanisms in place to link it to the medical images. This shortcoming was reported by physicists nearly a decade ago: “The analysis of MVCT images in terms of Hounsfield units cannot be done conveniently with the standard TomoTherapy software”<sup>78</sup>. However, even in newer Radixact systems with updated vendor software that superseded older TomoTherapy systems, the lack of the calibration-related metadata in medical images persists. Current dose re-computations on MVCT images either define analytical models to account for Hounsfield units (HU) variations between scanners or within a scanner over time<sup>79</sup> or conduct additional measurements alongside the imaging process<sup>80</sup>.

#### 2.1.1 Pilot Study Data

The MVCT images data to be linked with the scanner calibration data in this case study originate from the VoxTox study<sup>81</sup>. The VoxTox study was designed to investigate the relationship between the delivered radiation dose and observed normal tissue toxicity. The dataset comprises over 800 patient cases totalling 23,000 scans acquired between 2007 and 2017 on two TomoTherapy devices at Cambridge University Hospitals.

#### 2.1.2 Tomotherapy Calibration Data

The two TomoTherapy units underwent daily, weekly, and monthly imaging calibrations. Daily scanner calibrations are conducted via a so-called “air scan”. More extensive weekly and monthly calibrations are obtained by imaging of a test object called “cheese phantom” (see Figure 2).



**Figure 2.1 The cheese phantom with Gammex density plugs.**

For monthly calibrations, the cheese phantom was imaged with inserted Gammex (Gammex, Middleton, WI) density plugs that are designed to mimic the electron densities of the human body tissues such as air, lung, adipose tissue, cartilage and bone. The plug set also includes a distilled water vial and two plugs with holes of different diameters to assess the scanner resolution. The purpose of the liquid water plug is to provide a reference measurand with known HU value of 0 in every phantom image to control for subtle variations in the phantom density, composition of the density plugs and their non-equivalence to anatomical tissues in terms of the electron density.

Once a scan is acquired, the DICOM image is exported to be analysed by imaging physicists. During the analysis the density plugs are automatically segmented, and a mean HU value and standard deviation are recorded for each density. The measured HU values are used to construct a scanner calibration curve, i.e., the function that translates the HU values from the image into electron density. Imaging physicists compare the monthly measurements with the past values, look for trends and assess whether the observed deviations are within prescribed tolerances<sup>82</sup>. If the latter is not the case, service engineers are called to address the underlying technical issues.

### 2.1.3 Proof-of-Concept Implementation in DICOM

The Digital Imaging and Communications in Medicine (DICOM) standard was introduced to enable medical data exchange and curation<sup>83</sup>. Since its introduction in 1993, the standard has been widely adopted for a range of medical imaging devices including all forms of CT, magnetic resonance imaging, ultrasound devices and radiotherapy treatment planning systems<sup>84</sup>.

The standard provides controlled key-value dictionaries, file format and services. Key features of DICOM-compliant files or “objects” are:

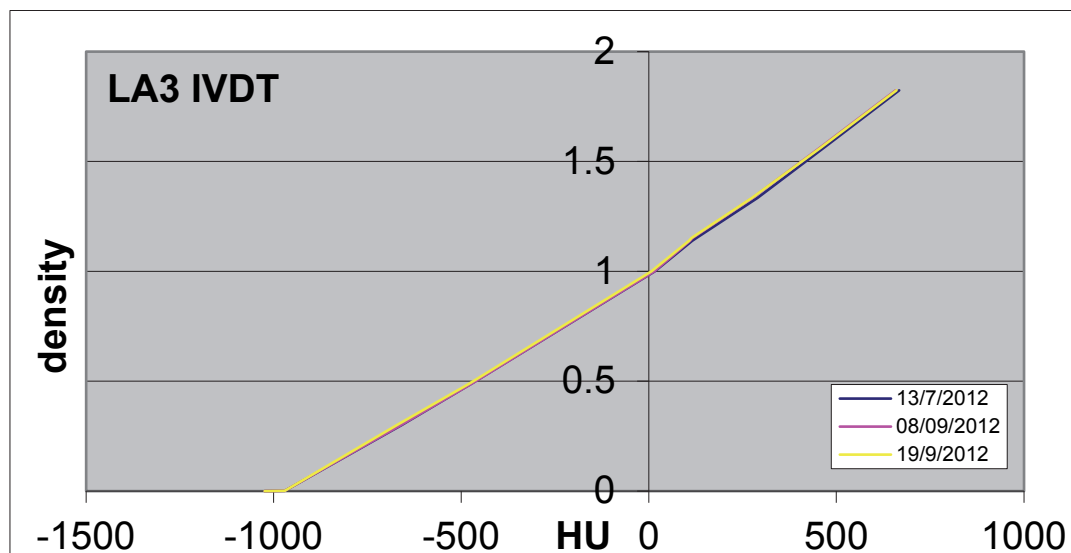
- Unique digital identification of each object via Service-Object Pair Instance Unique Identifier (SOP Instance UID).
- Aggregation of image data and metadata to ensure the data provenance and traceability. It is particularly important in the context of images acquired using ionising radiation, to avoid data losses and to track the radiation-induced tissue damage.
- Maintenance of links to other DICOM objects via Referenced SOP Instance UIDs.

Using the [pydicom](#) Python library, we aimed to integrate calibration data from TomoTherapy devices into the patient MVCT images using DICOM private tags and Referenced SOP Instance UID mechanisms. The integration of calibration data into medical images is expected to provide a quantitative estimator of HU uncertainty which could be utilised for image normalisation. The calibration-corrected MVCT images could then improve the adaptive radiotherapy dosimetry and provide comparable dataset for AI-based image analysis software.

## 2.2 CASE STUDY TECHNICAL WORK

### 2.2.1 Harmonisation of Imaging Device Calibration Data

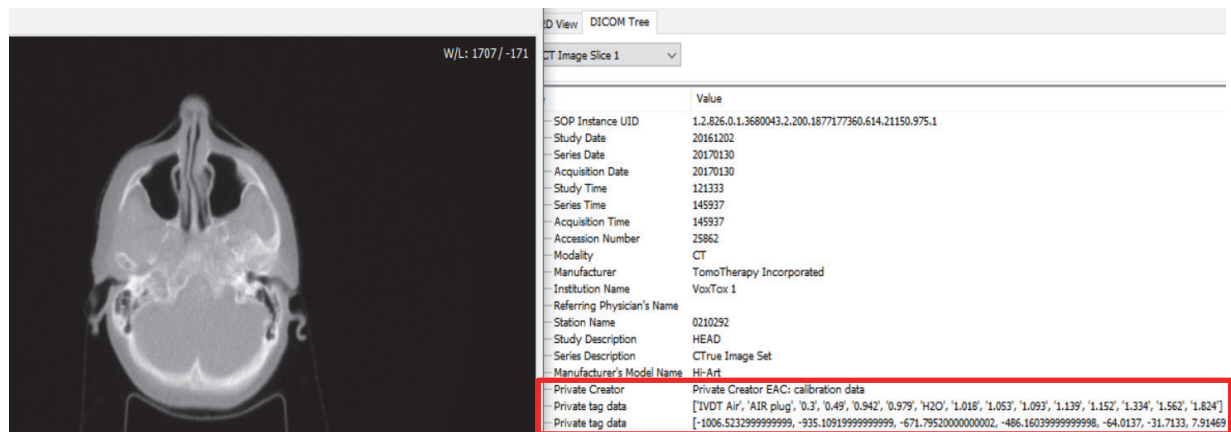
A total of 479 monthly cheese phantom calibration measurements for two TomoTherapy treatment units acquired between 2007 and 2018 were reviewed and collated into a single document. The measurements contain the HU values for 14 anatomical tissue substitute materials and translate the image grey value to the tissue electron density. An example of a single calibration record is shown in Figure 2.2.



**Figure 2.2 Image-value-to-density (IVDT) table acquired from MVCT cheese phantom images.**

### 2.2.2 Linkage of Patient Imaging Data and Device Calibration Data

A software procedure was implemented to match the MVCT DICOM images of 873 patients to the most recent calibration record with respect to the image acquisition date. The DICOM headers were enhanced by private DICOM tags to include the calibration information. Twenty-six CNS patients had a total of 793 MVCT scans, all of which have had private tags with said calibration information included. There were also 319 H&N patients, for which 7771 out of 9213 MVCT scans (~84%) have had calibration information included. Finally, the 528 prostate cancer patients had a total of 16262 MVCT scans, of which 11274 have had calibration information added (~69%).



**Figure 2.3 DICOM image (left) and DICOM header (right) with inserted calibration information (red frame).**

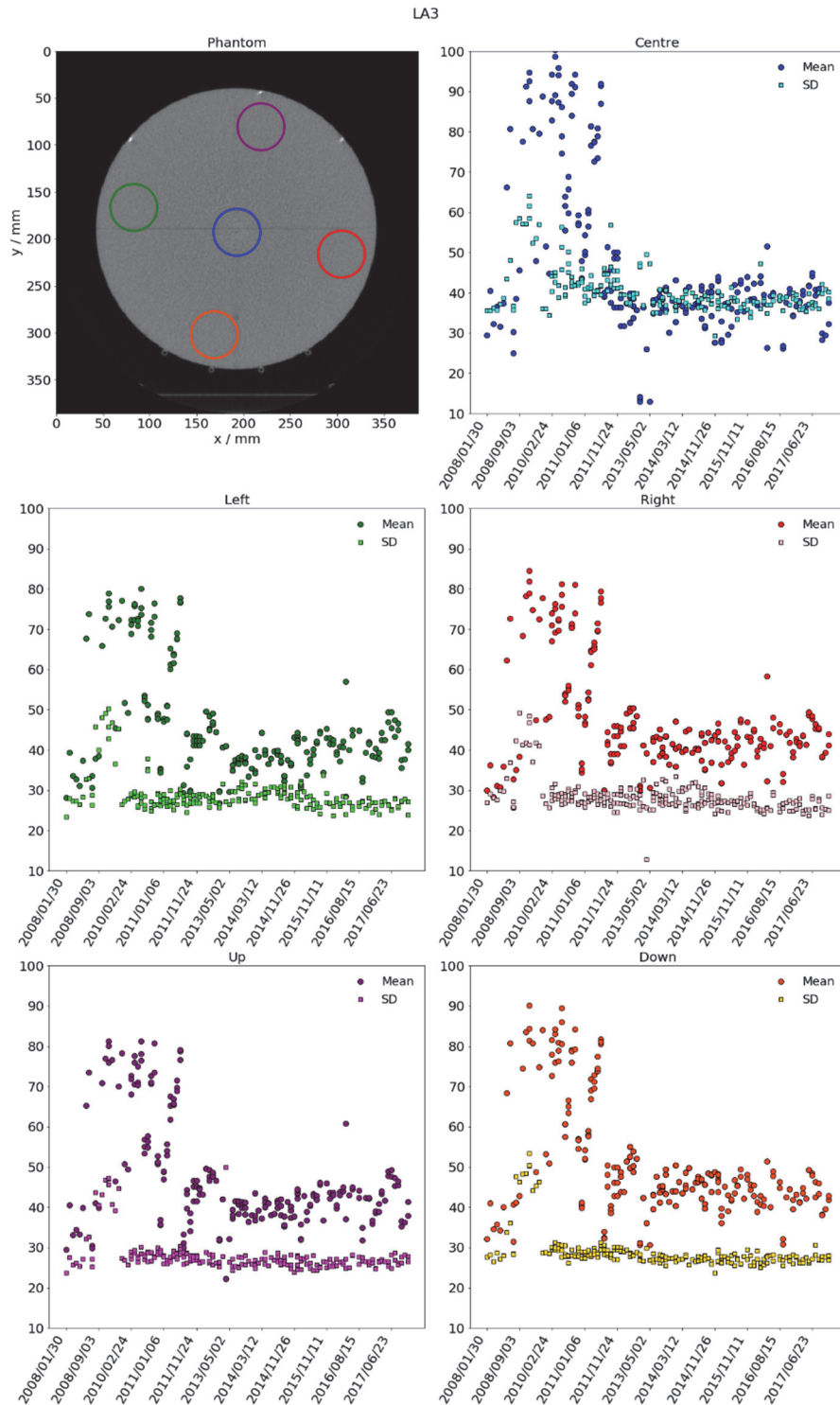
### 2.2.3 Image Quality Assessment

To explore potential variations between the two scanners and over time, the cheese phantom images were re-analysed using Matlab (version 9.5.0 R2018b, Natick, Massachusetts: The MathWorks Inc.) software developed in-house. Key image characteristics such as image uniformity and daily mean HU value variations were computed to re-evaluate the metrics recorded by the hospital medical physics team and to enhance them by image noise measurements.

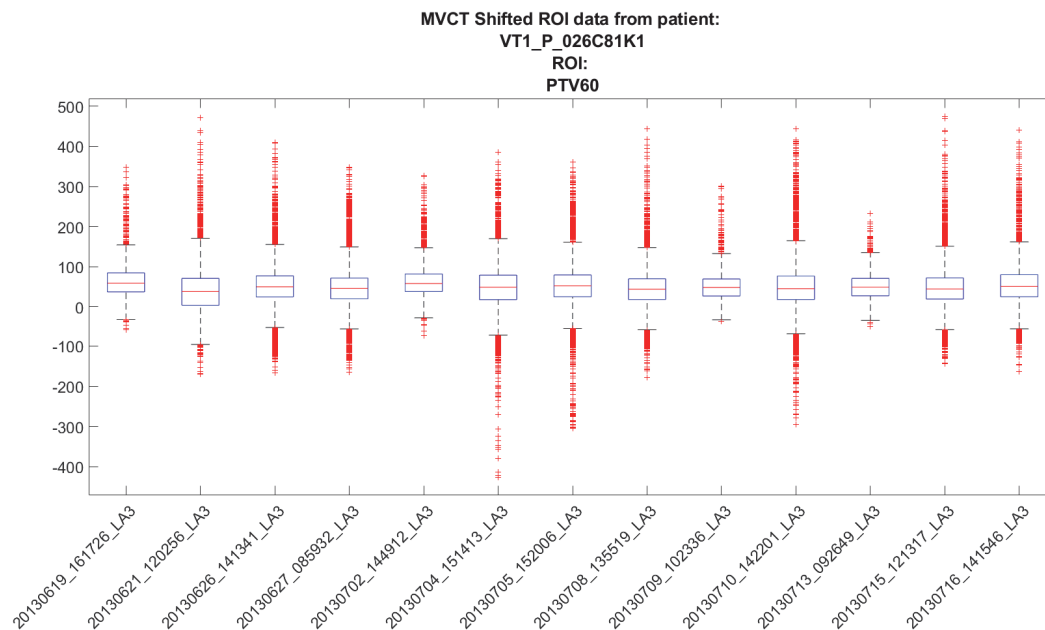
Five regions of interest (ROIs) were placed in the uniform parts of the cheese phantom body as illustrated in Figure 2.4. Mean and standard deviation HU values were computed for each calibration image.

Figure 2.4 demonstrates a clear difference in the standard deviation of HU values observed in peripheral (~30 HU) and central ROIs (~40 HU). This difference can be attributed to beam hardening, whereby the incident beam spectre shifts towards higher energies as the photons are attenuated by larger amounts of material in the centre of the bore compared to the peripheral regions. Similar effects can be observed for patients and are well documented in the literature<sup>85</sup>. The inter-region variations between the four peripheral ROIs are very low and do not exceed 35 HU within a single scan, though the regions vary maximally from 0 HU to 65 HU over time. The temporal variations show marked improvement in 2011, as the vendor introduced a new daily scanner self-calibration method called “air scan”, whereby the daily values captured by scanning the empty scanner bore are used to normalise the subsequent images of the day.

Figure 2.5 shows daily image variations. In this box plot, we see that although median values remain consistent over time, the spread of the values can vary substantially (from a range of ~ 400 HU to 800 HU) between patient acquisitions.



**Figure 2.4 Cheese phantom with homogeneous ROIs (top left). Mean and standard deviations of HU values in the individual ROIs during the scanner lifetime.**



**Figure 2.5 HU values in the soft tissue region at different days of radiotherapy treatment. Plots use MATLAB default parameters where the central mark is the median, bottom and top edges are the 25<sup>th</sup> and 75<sup>th</sup> percentiles, and whiskers extend to the most extreme data points not considered outliers ( $\pm 2.7\sigma$  or 99.3% coverage is data is normally distributed)**

## 2.3 CASE STUDY OUTCOMES/HIGHLIGHTS

### 2.3.2 Linkage of patient and device calibration data

Data about scanner grey-value fluctuations was collected from phantom calibrations and included into the DICOM header information as a private field if a suitable calibration scan could be found. This successfully included device calibration data into all CT scans and a majority of MCVT scans. Having such data will improve future analyses as they will now be able to correct for scanner induced grey-level variations. Additionally, we have shown that variation in the phantom and patient images can differ over time, despite an outwardly stable median/mean.

### 2.3.3 Learning from the case study

While the DICOM standard is widely used in medical imaging, large variations exist on what metadata are supplied with the images. Frequently, this lack of consensus leads to the absence of data that are crucial for image quantification. Scanner calibration data is one such example. Calibration data are not routinely included with medical images and are not easily accessible for image analysis. We believe that calibration information has further use for image grey-level corrections that could result in more comparable images with grey values traceable to the ground truth. Such corrections then allow for a better quantitative description of the image data. In particular, once corrected for known sources of drift, it is possible to detect and investigate unknown sources of drift. These might be indicative of component failure, and useful indicators that a scanner might need maintenance. Such techniques are commonly used in component manufacturing processes and are known as **predictive maintenance**. They can reduce the occurrence of failure events and associated expenses. In a medical imaging and radiation oncology context, predictive maintenance would help to increase the scanner and treatment unit availability, improve the utilisation of hospital resources and ultimately save lives.



### 2.3.4 Workshop presentations

The case study was presented alongside the other Digital Health case studies as a poster at the workshop “Artificial Intelligence Developments in Healthcare Imaging” held at Cambridge Institute for Mathematical Sciences on 23-24<sup>th</sup> of October 2019. A dedicated poster was presented during the Scientific Technology Advisory Council visit to NPL Teddington in November 2019.

## 2.4 FUTURE DIRECTION OF CASE STUDY

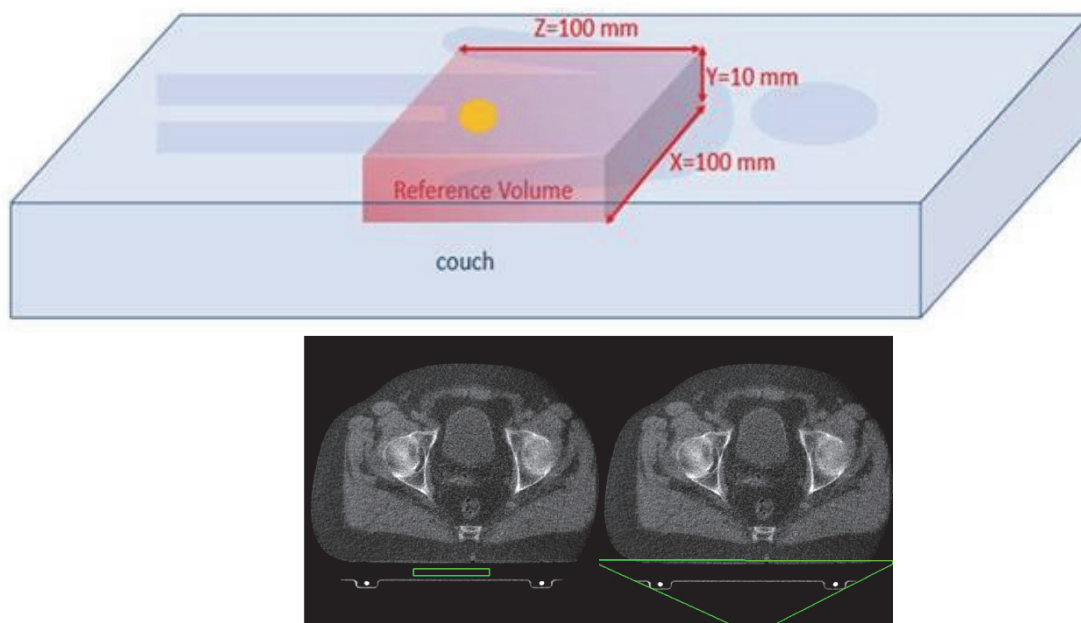
The results demonstrated variations in image properties both between the two imaging devices and over time. While these differences may be insignificant for radiation therapy dosimetry<sup>86</sup>, they may have considerable impact for assessment of subtle tissue changes during the treatment course.

Future work will focus on:

- a) quantification of image variations in a representative soft tissue region
- b) development of methods to correct the image for daily machine variations
- c) exploration of methods to predict machine behaviour from clinical images

### 2.4.2 Normalisation of Imaging Data Using Calibration Proxy

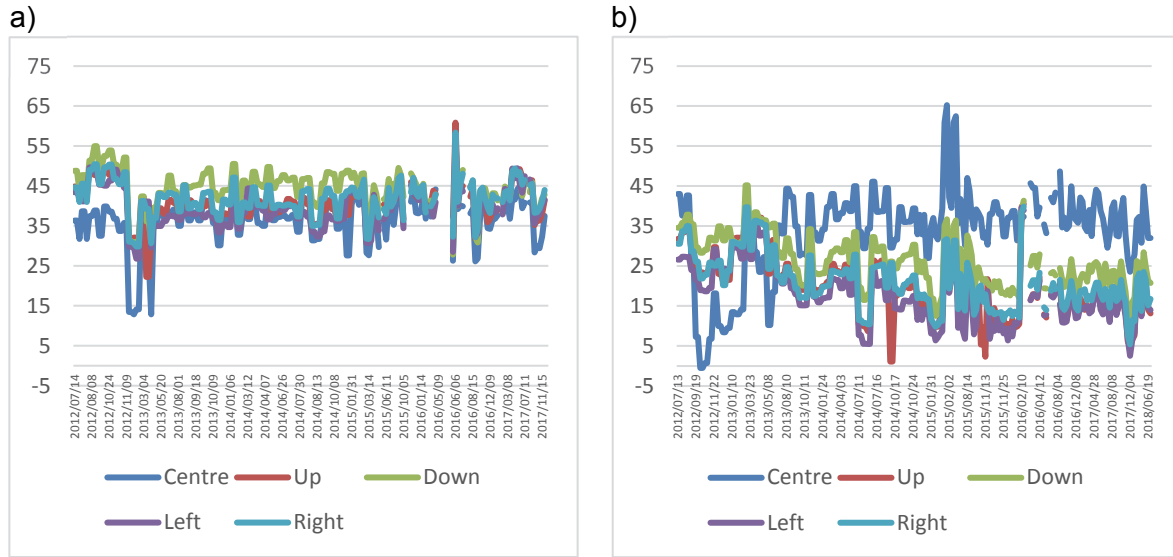
Phantom and liquid water measurements are only taken during monthly quality assurance tests. This temporal sparseness of reference measurements does not allow to analyse daily variations. To gain insight into the daily fluctuations, we plan to explore the use of the ROI defined inside of the treatment couch as a calibration proxy. A schematic representation of the proxy ROI is shown in Figure 2.6. This technique is possible on TomoTherapy machines, as the couch is the only invariable object between all images.



**Figure 2.6: An example ROI drawn in the couch region as potential calibration proxies**

## 2.4.2 Application of Process Control Techniques to Monitor Imaging Device Behaviour

The calibration data acquired are typically used to ensure that TomoTherapy machines measure the plugs and background HUs within a set range of values, but variation within this range tends to be ignored during typical clinical calibration. However, simple plots of the mean background value for both TomoTherapy units show marked differences in variations during their lifetimes as seen in Figure 7. The variations in Figure 2.7a) are tightly grouped suggesting this scanner drifts very little over time, whereas variations in Figure 2.7b) over time are much larger and also position dependant, suggesting it is not performing as expected (e.g. the HU values in the central ROI range from -5 to 65 HU). This information could potentially be used for anticipating scanner maintenance needs<sup>87</sup>.



**Figure 2.7 Variation of the mean background HU of the cheese phantom in locations shown in figure 2.4 over time for a) TomoTherapy unit 1; b) TomoTherapy unit 2.**

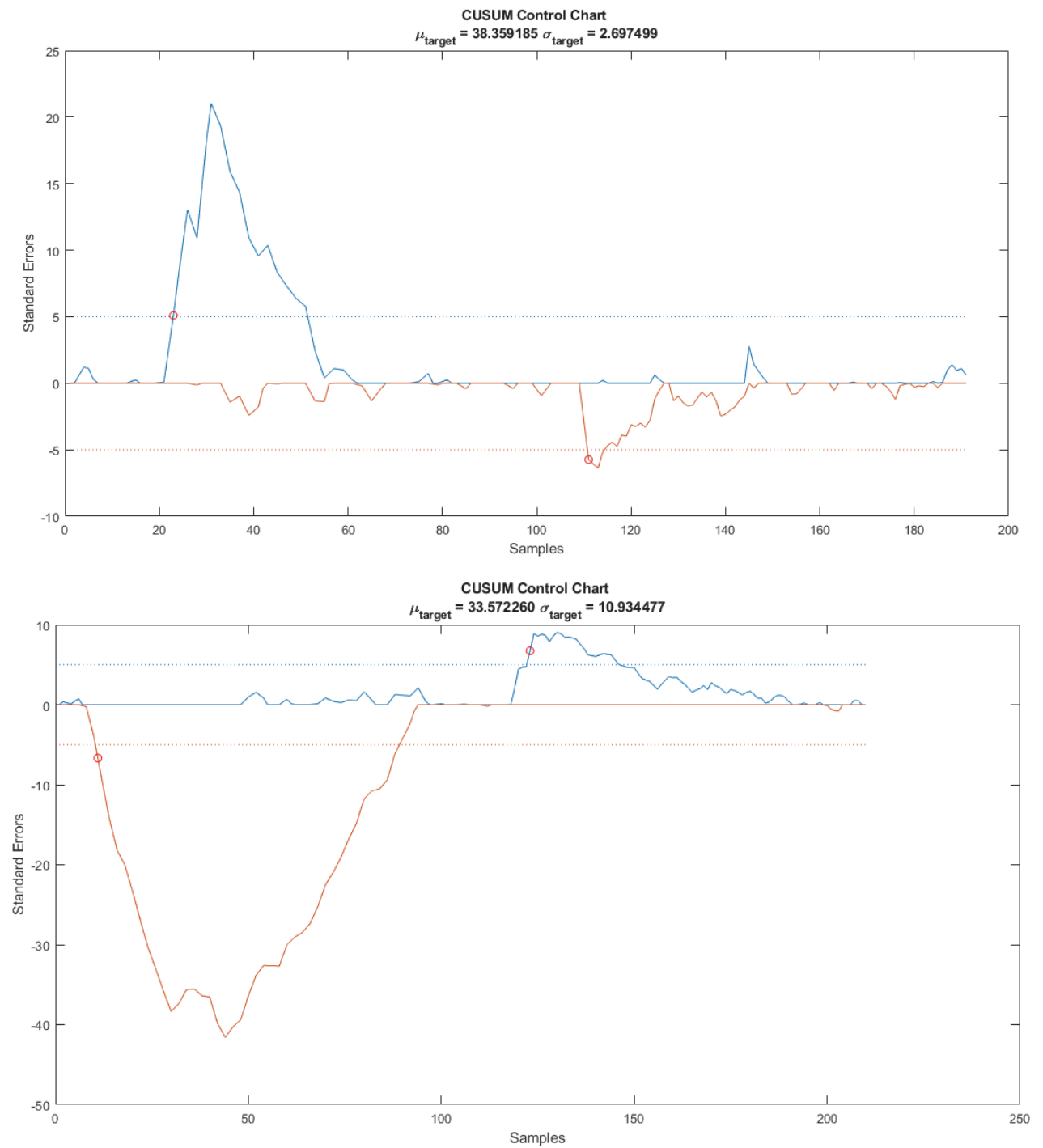
The simplest method to investigate this data further is known as the cumulative sum, or cusum<sup>88</sup>. This is an accumulated sum of the deviations of a process from the desired value. This value gets larger if there are multiple measurements with a substantial deviation from the desired value. In our case this could be the background mean in the centre of the phantom. So our cusum  $S$  after  $t$  samples would be:

$$S_{t+1} = \max(0, S_t + x_t - \omega_t) \quad (2.1)$$

with  $S_0 = 0$  and where  $x_t$  is a sampled mean and  $\omega_t$  is a weighting factor, often interpreted as the likelihood function, which determines what is considered an important deviation.

Application of the cusum method where the weight was set as the mean over the whole time series gave the charts in Figure 2.8, which clearly indicate the magnitude of the drift in the second TomoTherapy unit. Such charts, paired with more advanced techniques, could assist medical physicists in knowing when their machines require correcting, as large periods of drift such as those seen below are likely to be indicative of problems in the scanner.





**Figure 2.8** Cusum charts showing negative and positive drift in the scanners, with step changes indicated by red circles; top is Tomotherapy unit 1; bottom is Tomotherapy unit 2

### 3 CASE STUDY 2: ECHO

**Full title:** USING DATA LINKAGE TO IMPROVE EVIDENCE-BASED PRACTICE BEYOND THE EVIDENCE FROM CLINICAL TRIALS. IMPROVING THE EFFECT OF CANCER AND ITS TREATMENT ON HEALTH OUTCOMES (ECHO)

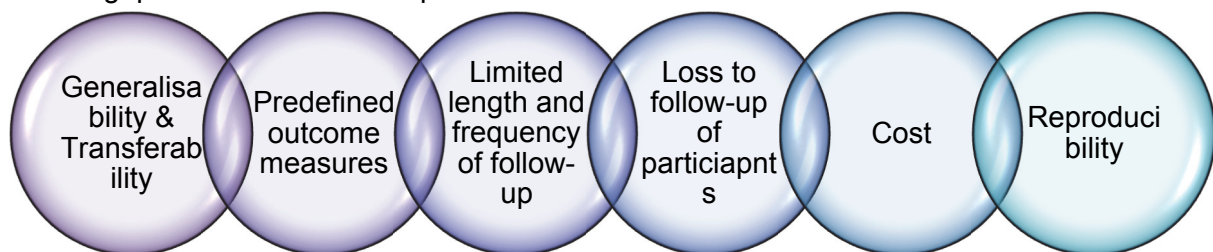
**Technical lead:** Agnieszka Lemanska

**Web:** <https://www.npl.co.uk/research/data-science/digital-health/case-study/echo>

#### 3.1 CASE STUDY PROPOSAL

##### 3.1.1 Introduction

Clinical trials are the gold standard that drive evaluation and improvements in healthcare. New treatments are compared with the standard of care treatments in a selected group of eligible patients. This provides evidence that can be used in clinical practice. However, clinical trials such as randomised controlled trials (RCTs) have limitations (Figure 3.) and this creates gaps in evidence based practice.



**Figure 3.1 Limitations of RCTs in informing evidence-based practice**

Although RCTs provide information about the effectiveness of a treatment, it is in selected populations (strict eligibility criteria), so the results of RCTs may not be **generalisable or transferable** to the clinical populations of patients requiring treatments. Evidence shows that trials recruit younger and healthier people than the true clinical populations (e.g. older people with more comorbidities, medications and higher risk of complications). In addition, RCTs have a **limited frequency and length of follow-up** as well as **loss of participants to follow-up**. RCTs also collect a **limited number of predefined variables**. However, side-effects from treatments can emerge many years post-treatment and new side-effects can emerge that were not considered at the outset of a trial. RCTs are **expensive to run and implement**, and a lack of standardised support, training and evaluation of RCTs implementation across different cancer centres and NHS Trusts can limit **reproducibility** in clinical setting.

Therefore, RCTs should not be used alone to evaluate treatments but supplemented with other data e.g. national registers and routinely-collected data (from both primary and secondary care). Currently it is not fully known how to best evaluate healthcare in populations of patients not in clinical trials. Utilisation of routinely collected NHS data can offer an effective and cost-effective solutions for the long-term follow-up and evaluation of treatments. These sources of data are not yet used extensively, but they should be included in addition to clinical trials to drive the evaluation and improvements in oncology and healthcare in general.

The levels of evaluation are:

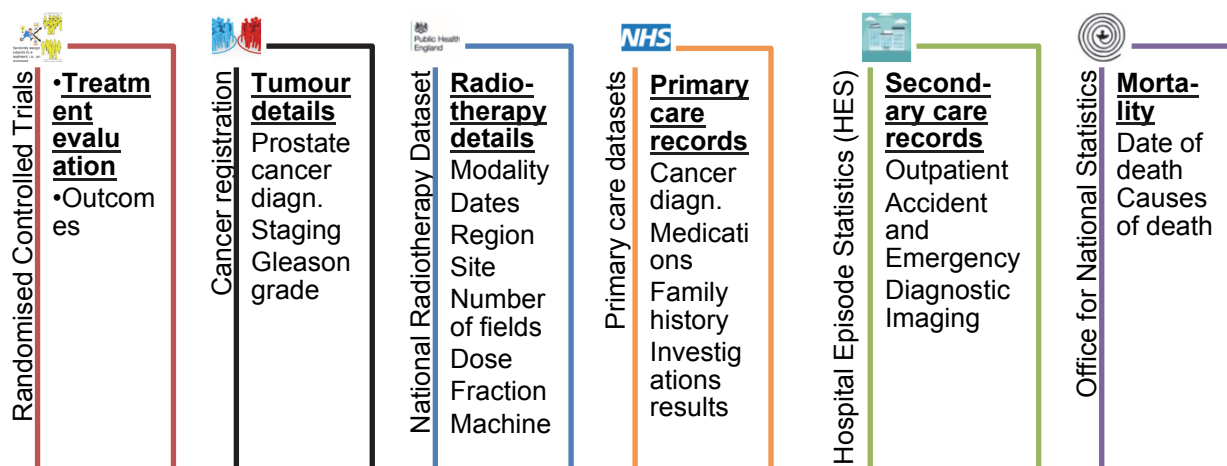
- ⇒ **Population level**: evaluate outcomes beyond RCTs using analysis of large NHS datasets

- Compare outcomes as published in RCT to the routine clinical setting to explore the reproducibility of results of RCTs after implementation
  - Compare outcomes of different treatments
  - Develop methodologies and standards for supplementation of RCTs with routinely collected NHS data
- ⇒ **Local level**: evaluate individual cancer centres and their outcomes
- Compare outcomes achieved in different NHS Trusts and cancer centres
- Individual level**: evaluate outcomes in the clinical setting
- Curate patient-centred outcomes and develop standards for measurement of outcomes from routine data
  - Identify risk factors e.g. age, comorbidity, lifestyle and their effect on outcomes
  - Investigate longitudinal outcomes e.g. BMI, secondary cancers, mortality of patients
  - Develop clinically useful risk prediction models for individual patient outcomes
  - Create tools to support personalised medicine

Large amounts of data are collected in the NHS, but a lack of standards in what attributes are included and how they are recorded, makes the utilisation of much of this data challenging. In addition, without standards for what and how to record, combining multiple sources of data becomes extremely challenging due to incompatible data formats, missing values, mismatch of measurement units and lack of metrological traceability. Mining such heterogeneous data requires extensive 'data cleaning' to remove errors, input the missing data or convert data into compatible types. For large datasets, such as population data, this becomes a daunting and impractical task. Furthermore, data mining and cleansing tools have to be re-developed every time any of the source data formats changes, adding to the strain on the clinical resources and increasing the likelihood of errors. Standardisation of data and metadata formats ensures that the data are collected and used in the best possible way, greatly increasing the 'data half-life' and long-term value.

### 3.1.2 Knowledge gaps and data challenges

Recommendations and standards are needed on how the evidence from clinical trials may be supplemented by other sources of data collected in the NHS. Clinical trials will remain an important source of evaluation evidence, but to provide further clinical benefits for patients it is important to take advantage of other sources of data such as the data routinely collected in primary and secondary care. Figure 3. demonstrates sources of data available in the NHS that could be used to evaluate oncology healthcare (subject to necessary data infrastructures).



**Figure 3.2 Data sources available in the NHS to support evidence-based practice in oncology.**

RCTs as gold standard for treatment evaluation can be supplemented with evidence from national registers (e.g. Cancer Registry, Radiotherapy Dataset [RTDS], mortality [ONS]) and routinely collected data (secondary and primary care). In this case study, the sources of healthcare information listed above will be accessed to address the issues related to data curation and data infrastructures. This case study is important because it will add to standards on how these data are curated and analysed.

### 3.1.3 Overarching objectives

Healthcare services are stretched to respond to the increasing demands as a result of increasing cancer diagnosis, treatment side-effects and other cancer consequences. It is therefore critical to understand the long-term healthcare needs of the ageing population of cancer survivors and the impact of comorbidities on the management of cancer side-effects. This case study will use data curation approaches to inform and support the development of new data standards, recommendations, methodologies and techniques in the area of evidence-based oncology healthcare. The study will form part of a bigger work package in the Data Science department undertaken to: 1) investigate how routinely-collected data can be used to contribute to evidence-based oncology practice (e.g. for more effective and cost-effective follow-up of patients beyond RCTs); and 2) to improve the Effect of Cancer and its treatment on long-term Health Outcomes (ECHO).

Data curation issues that will be explored:

- ✓ Understand the sources, nature, limitations and usability of different sources of NHS data for cancer outcomes research
- ✓ Develop standards for prostate cancer case identification in primary care using an ontological approach
- ✓ Link a prostate cancer RCT called CHHiP to a primary care database
- ✓ Report on the feasibility of the linkage process
- ✓ Investigate the value of the linked resource in providing evidence for evaluation of cancer outcomes

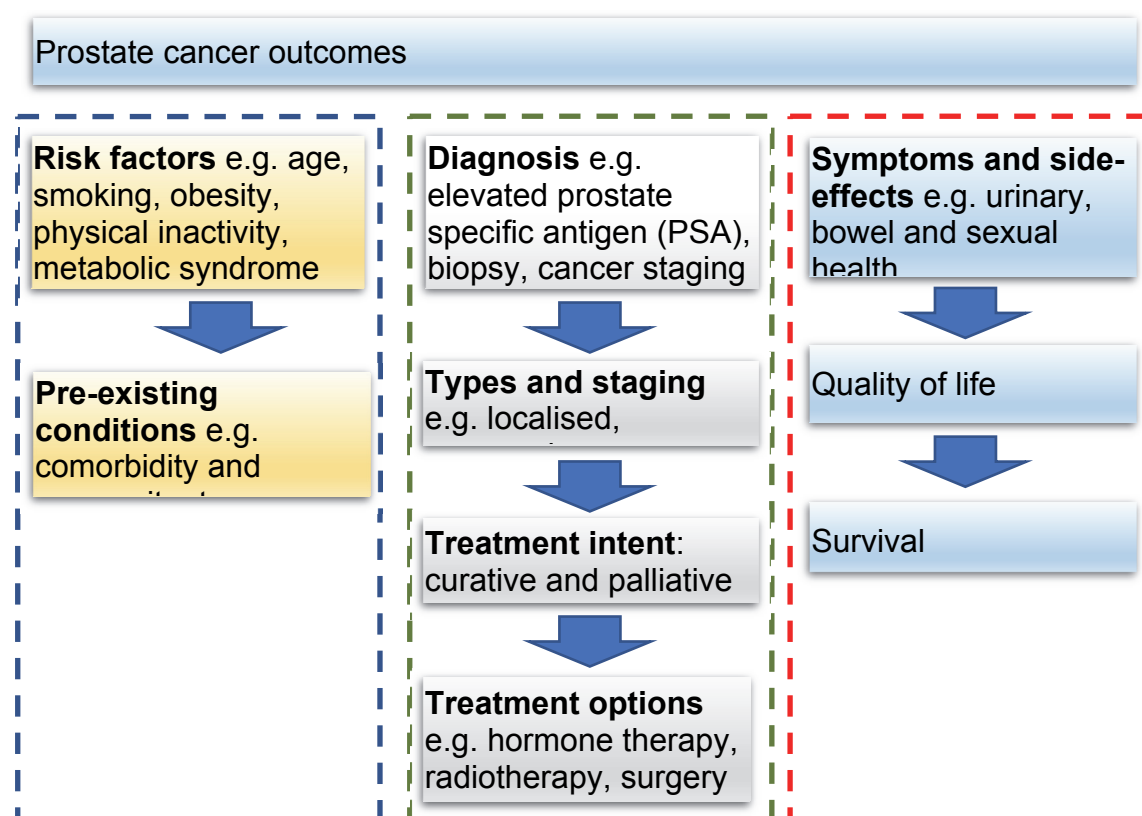
## 3.2 CASE STUDY TECHNICAL WORK

**Title:** Curation of prostate cancer diagnosis, symptoms and treatment side-effects: an ontology-based approach for case identification from routine primary care data.

### 3.2.1 Introduction: Prostate cancer as a proof of concept

Cancer survivorship is increasing owing to continual improvements in cancer treatments and earlier diagnosis<sup>89</sup>. Increasingly, cancer is perceived as a chronic condition with good prognosis but long-term consequences and related health needs<sup>90</sup>. Symptoms and toxicity that are most affected by prostate cancer and its treatment are in the three main health domains: bladder, bowel and sexual (e.g. urinary urgency, rectal incontinence or erectile dysfunction). Cancer late-effects are a key concern to oncologists in developed countries because 80% of patients survive at least ten years post treatment and have the potential to develop late-effects alongside multiple comorbidities.

The consequences of cancer and its treatment, health outcomes and comorbidities can affect treatment choices, quality of life and healthcare services usage (contributing to NHS burden and costs). It is therefore critical to understand the long-term healthcare needs of the growing and ageing population of cancer survivors and the impact of comorbidities on the management of cancer side-effects. Figure shows the complexity of prostate cancer outcomes. Outcomes will vary depending on cancer stage, treatment selected and risk factors such as older age or comorbidity.



**Figure 3.3 Complexity of health outcomes in prostate cancer. They vary depending on cancer stage, treatment and risk factors such as older age or comorbidity.**

In the UK, all primary care patient-clinician encounters are recorded as episodes and prescriptions using clinical codes. Prostate cancer diagnosis should be recorded as an episode using a Read code B46 (malignant neoplasm of prostate). However, this relies on a clinician knowing what Read code to select from a possible list of codes in the system. In addition, some clinicians may fail to code prostate cancer altogether, so the diagnosis may need to be derived from a series of relevant Read codes as described above. An ontological

approach can be used to systematically define and map the clinical concepts such as prostate cancer and to relate them across different sources of information. This can then be used for standardised and reproducible identification of patient cohorts that can serve for epidemiological and translational research.

We use prostate cancer in the project as an exemplar to investigate how routine healthcare data and data linkage can be used to supplement the evidence from randomised control trials (RCTs) and improve oncology treatments and patient care. When validated, the methodology from this project could be adapted and generalised so that it could be applied to other treatments and in other diseases. For example, this research could be extended to neurodegenerative diseases such as Parkinson's and other long-term conditions to establish how RCTs can be supplemented with other sources of data.

### 3.2.2 Case study aims

Currently, primary care datasets are not routinely used for (prostate) cancer identification and follow-up. This work is a first step to increase the utility of primary care data for oncology by a) creating a knowledge base of data sources, b) mapping out the required integration efforts and c) developing a practical ontology-based method for systematic and reproducible prostate cancer case identification and validating this method on real-world datasets. The developed ontology will be used to standardise the identification and retrieval of prostate cancer cases from primary care data. Future work will extend the methods developed within this study to analyse patient outcomes at a national, local, and individual level and compare it to the results from clinical trials.

## Methods

To develop the ontology for prostate cancer case identification in primary care databases, we conducted research on previous studies in the area and analysed their cohort extraction and case identification approaches i.e. lists of clinical codes used for extraction. The existing ontologies were investigated and re-purposed. The ontology developed within this case study was tested on the Royal College of General Practitioners (RCGP) Research Surveillance Centre (RSC)<sup>91</sup> and published at the Medical Informatics Europe conference<sup>92,93</sup>.

## Ontology development

Two existing prostate cancer ontologies were found and repurposed in this case study:

- Prostate Cancer Lifestyle Ontology (PCaLiOn)  
(<https://bioportal.bioontology.org/ontologies/PCALION>)
- Prostate Cancer Ontology (PCaO)  
(<https://bioportal.bioontology.org/ontologies/PCAO>)

Using the two above ontologies, a new, bespoke ontology was developed in this case study for case identification in the routinely collected primary care data within the Royal College of General Practitioners (RCGP) Research Surveillance Centre (RSC)<sup>94</sup> dataset to define and improve methods for prostate cancer case identification.

The developed ontology maps clinical concepts (such as prostate cancer diagnosis and processes of care that are indicative of the diagnosis) across different sources of information. This is important because prostate cancer is not diagnosed in primary care, and because of this, a detailed understanding of data and integration efforts is required. The main clinical coding systems used in UK primary care are Read Version 2 (Read V2), Clinical Terms Version 3 (CTV3) and Systematised Nomenclature of Medicine Clinical Terms (SNOMED CT). We investigated the usability of clinical codes to identify and describe a cohort of

prostate cancer patients from a large English electronic database. The focus is on prostate cancer, but the results are relevant to other cancers and diseases.

The four steps of ontology-building are:

**Step 1: creating the ontology.** Defining the concepts of relevance to prostate cancer. This may include diagnoses, symptoms, investigations, treatments and other 'processes of care'.

**Step 2: identifying clinical codes.** Identifying the most appropriate clinical codes that describe the concepts. This step is specific to Read V2, CTV3 and SNOMED CT coding systems.

**Step 3: data extraction testing phase.** Using the RCGP RSC dataset extract prostate cancer cases to ensure that the extracted data are consistent with anticipated outputs.

**Step 4: validating the ontology.** Investigating reasons for false positive and false negative cases.

## Data extraction

Following data sharing and research governance approvals, the RCGP RSC, a nationally representative network database, was accessed. RCGP RSC has been collecting primary care data in England, managing the database and monitoring disease trends for over 50 years<sup>95</sup>. For testing the ontology, and for the primary results presented in this report we used the dataset from December 2016, which comprised 164 practices and 1,529,922 electronic health records (EHRs) for male patients who were registered with a practice within the RCGP RSC network between 2004 and 2016. As the network of practices is constantly expanding and currently includes over 530 practices with a total number of about 4 million active patients (>5% of the English population), the future publications will use the latest RCGP RSC dataset that will be available.

Prostate cancer diagnosis and treatment information including radiotherapy and hormone therapy was extracted using the developed ontology and lists of clinical codes. Demographic information included age at diagnosis, date of death, locality and ethnicity. Locality was based on Lower Layer Super Output Areas (LSOAs) which are small geographic areas (mean population of 1500) used in England and Wales to report statistics and protect patient identity. Prostate cancer diagnosis was extracted from primary care records using all diagnostic codes. This included B46 (*malignant neoplasm of prostate*), B834 (*carcinoma in situ of prostate*) and HNG0200 (*cancer of the prostate*, an old local code). Based on the ontology, it was also important to consider other codes e.g. for medical procedures, symptoms and investigations as indicative of prostate cancer diagnosis. This included codes for prostate biopsy, prostate-specific antigen (PSA) test or Gleason grading of prostate cancer. From this, in the primary results presented in this report, we have included 4M01 and Xalmc (*Gleason prostate grade 5-7 (medium)*), and 4M02 and Xalmd (*Gleason prostate grade 8-10 (high)*) because they can be directly related to prostate cancer.

The date and age at prostate cancer diagnosis was based on the first occurrence of the clinical code. Ethnicity was extracted in five categories using an ontology-based algorithm<sup>96</sup>. Radiotherapy was extracted using clinical codes, and hormone therapy was extracted using prescriptions. The list of medications used in hormone therapy, including gonadotropin-releasing hormone (GnRH) agonists and antiandrogens, was generated based on the British National Formulary (BNF, bnf.org).

Descriptive statistics were used including means and standard deviations for continuous variables and counts and percentages for categorical variables. To estimate the statistical significance, analysis of variance (ANOVA) was used. The ontology was implemented using the Protégé ontology editor version 5.5.0. Database extraction was performed using the Structured Query Language (SQL) Server Management Studio version 18.3.1. Statistical analyses were performed in R version 3.5.1 (R Development Core Team, Austria). This study



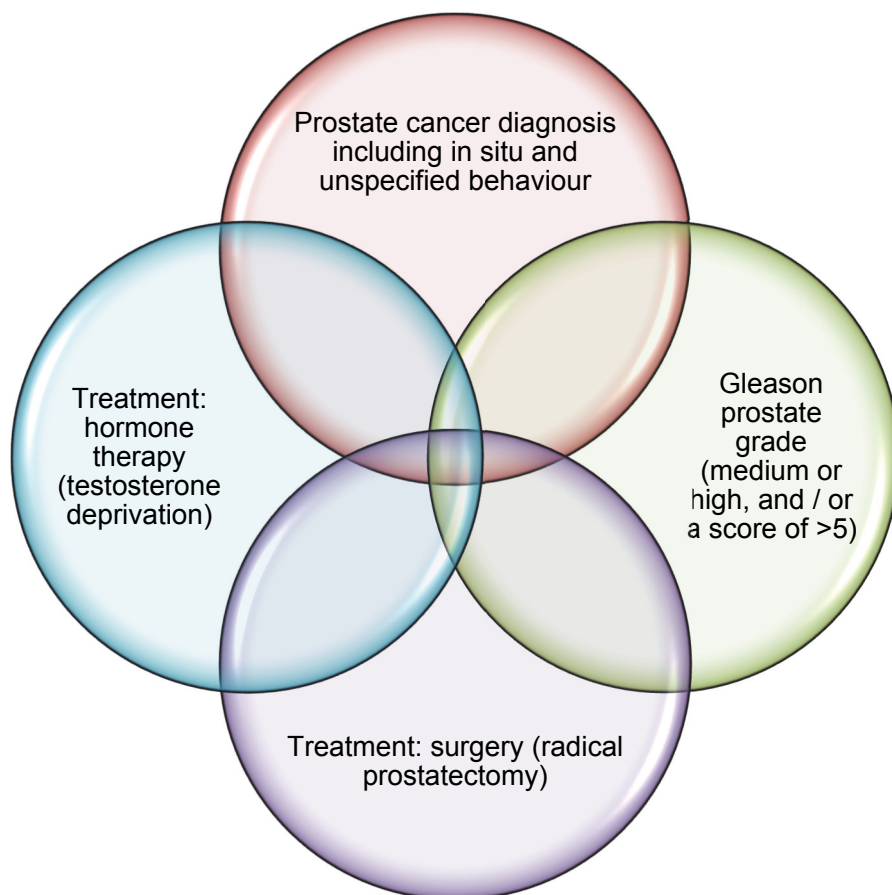
was approved by the RCGP RSC (data request RSC\_0118). Ethical approval was not required because we used pseudonymized and already collected data so no patient or clinician contact was needed. Patients consented to the research on an opt-out basis.

### 3.2.3 Results

#### The prostate cancer ontology

The ontology was developed for primary care data that systematically maps all the clinical concepts that can be used for prostate cancer case identification, cohort description and outcome evaluation. A screenshot of the ontology in the Protégé editor is presented in Figure 3.5.

The ontology will be published in the Biportal repository (<https://biportal.bioontology.org/>) and can be used for standardised and reproducible identification of patient cohorts for epidemiological research. The clinical concepts included in the algorithm for case identification are presented in Figure 3..



**Figure 3.4. Clinical concepts contributing to the prostate cancer case identification. The diagram is based on the ontology developed for primary care.**



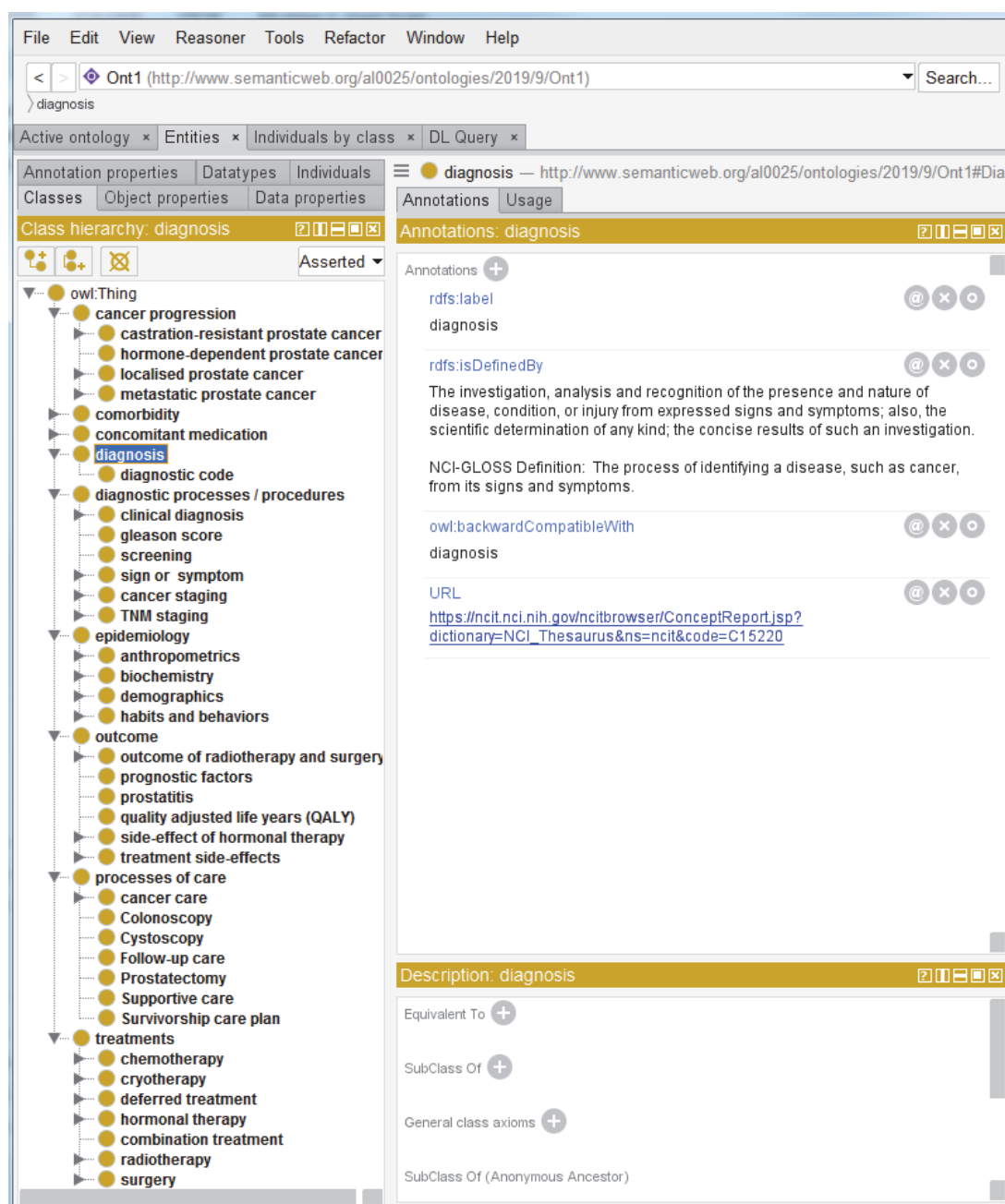


Figure 3.5 The prostate cancer ontology in the Protégé editor.

### The prostate cancer cohort

Prostate cancer diagnosis was established for 19,619 (1.3%) men who had at least one of the clinical codes that indicated prostate cancer. The three diagnostic codes: B46, B834 and HNG0200 were used in 18,356 (93.6%), 1,549 (7.9%) and 666 (3.4%) men respectively (some men had multiple codes). The Gleason grading codes for medium grade prostate cancer were used in 1506 and 29 EHRs (4M01 and Xalmc respectively), and for high grade in 556 and 6 EHRs (4M02 and Xalmd). This enabled the identification of 94 (0.5%) new cases that would not be included based on the diagnostic codes.

The average age at diagnosis was 70.8 (SD = 9.5) and the majority ( $n = 11,586$ ), 94.1% of men for whom ethnicity was recorded were white (Table 3.1). Ethnicity was missing for 7,302

(37.2%) participants. The men identified with prostate cancer were spread evenly across the country (see Figure 3.6).

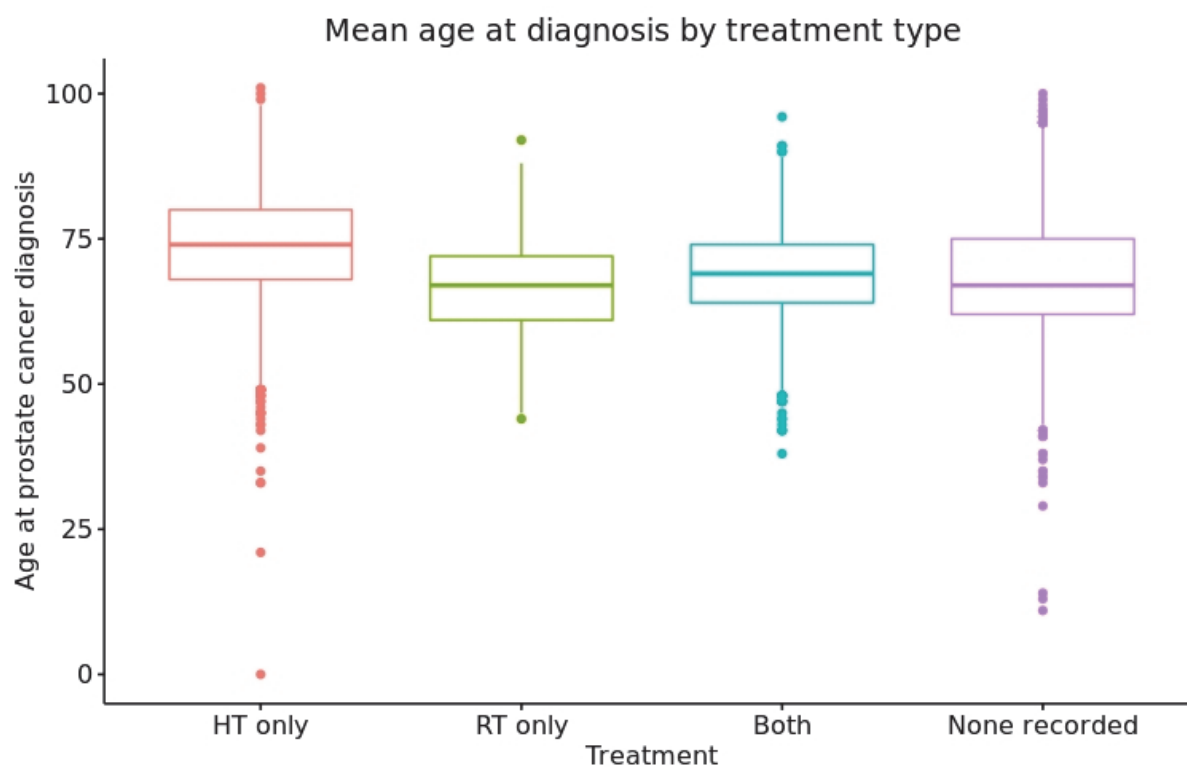
The most commonly-extracted treatment was hormone therapy. This analysis shows that 9324 (47.5%) men had hormone therapy alone and 1970 (10.1%) received it together with radiotherapy. Radiotherapy was extracted for 2604 (13.3%) and 7691 (39.2%) men did not have a clinical code for radiotherapy or a prescription for hormone therapy. Men who received hormone therapy were significantly older than those who received radiotherapy ( $P < 0.001$ ). The mean age at death was 80.6 (SD = 9.1). However, the date of death was only recorded for 1449 (7.4%) men.



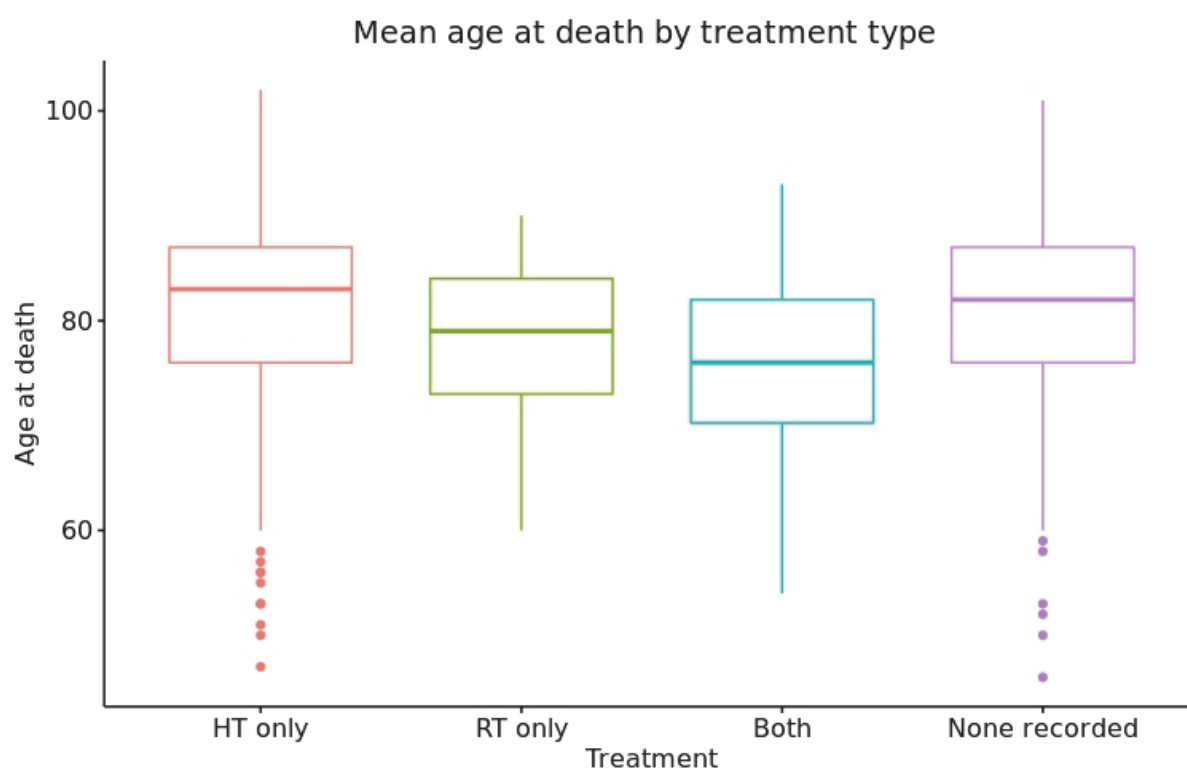
**Figure 3.6 Locations of study participants to demonstrate the spread of the sample across the country. This represents the need for integration of data across many GP practices in the country. Note that one patient was in Scotland. This is potentially because although they were registered with a GP in England, their home address was in Scotland.**

**Table 3.1. Demographic and treatment characteristics of the cohort**

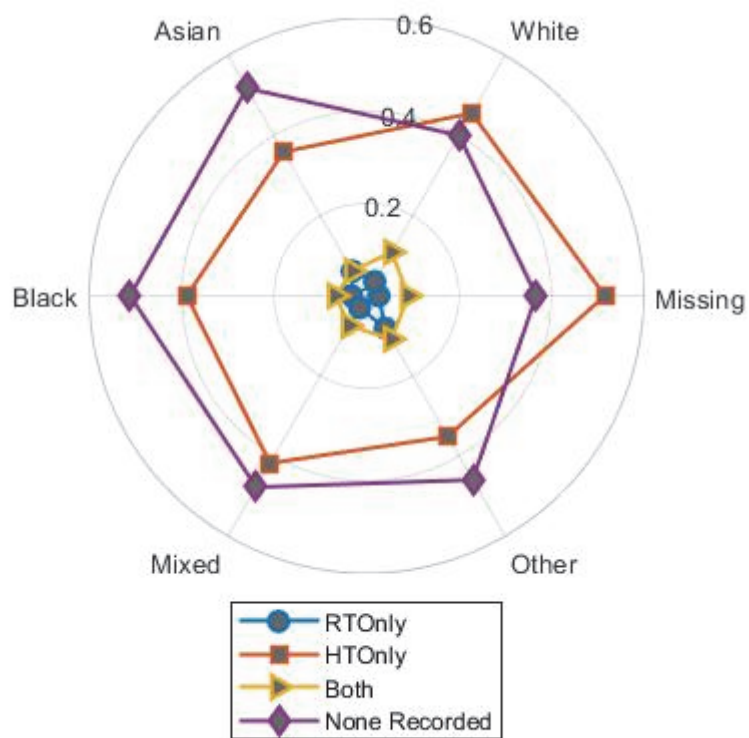
Characteristic	Mean (SD)	<i>n</i> (%)	<i>p</i> -value
Age at diagnosis (years)	70.8 (9.5)		
Missing		NULL	
Ethnicity			
White		11586 (59.1)	
Asian		195 (1.0)	
Black		432 (2.2)	
Mixed		67 (0.3)	
Other		37 (0.2)	
Missing		7302 (37.2)	
Age at diagnosis by ethnicity (years)			<0.001
White	70.3 (9.2)		
Asian	68.5 (8.9)		
Black	66.1 (9.8)		
Mixed	66.3 (11.1)		
Other	67.8 (8.9)		
Missing	72.1 (9.7)		
Treatment			
Radiotherapy		2604 (13.3)	
Radiotherapy (only)		634 (3.2)	
Hormone therapy		11294 (57.6)	
Hormone therapy (only)		9324 (47.5)	
Both treatments		1970 (10.0)	
None recorded		7691 (39.2)	
Age at diagnosis by treatment (years)			<0.001
Radiotherapy (only)	66.6 (7.7)		
Hormone therapy (only)	73.8 (9.0)		
Both treatments	68.7 (7.6)		
None recorded	68.1 (9.6)		
Age at the start of treatment (years)			
Radiotherapy	70.3 (8.0)		<0.001
Hormone therapy	74.8 (8.9)		
Age at death (years)	80.6 (9.1)		
None recorded		18170 (92.6)	
Age at death by treatment (years)			<0.001
Radiotherapy (only)	77.7 (7.3)		
Hormone therapy (only)	81.2 (8.6)		
Both treatments	76.0 (8.2)		
None recorded	81.3 (8.9)		



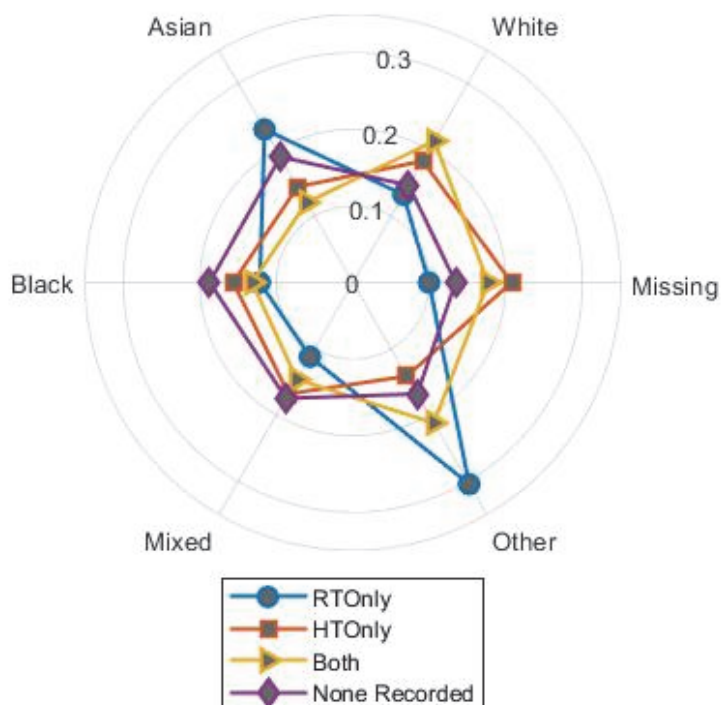
**Figure 3.7A Mean and standard deviation age at diagnosis of prostate cancer patients by treatment type.**



**Figure 3.7B Mean and standard deviation age at death of prostate cancer patients by treatment type.**



**Figure 3.8A Prostate cancer treatment modality grouped by ethnicity, normalised to ethnicity.**



**Figure 3.8B Prostate cancer treatment modality grouped by ethnicity, normalised to ethnicity and treatment modality.**

### 3.2.4 Key lessons learnt – lack of reproducibility in cohort studies

Primary care-based epidemiological studies use **varying subsets of clinical codes** to identify people with prostate cancer to be included into their study cohort. However, many studies do not report their lists of codes, and the lists of codes that are published differ from study to study. This leads to lack of comparability between studies and has impact on the study outcomes. In addition, “prostate cancer” describes a group of diagnoses (e.g. it can be localized or metastatic), and treatment and outcomes differ depending on cancer stage. Yet there are no established standards to retrieve this clinically relevant information from primary care databases, and relatively few studies have attempted this<sup>97</sup>.

### 3.2.5 Acknowledgements

This study was supported by the Royal College of General Practitioners (RCGP), Research and Surveillance Centre (RSC). This work uses data provided by patients and collected by the NHS as part of their care and support.

## 3.3 CASE STUDY OUTCOMES/HIGHLIGHTS

### 3.3.1 Prostate cancer ontology

An ontology was developed for routinely-collected primary care data to identify and describe the population of men affected by prostate cancer. It was a multidisciplinary effort as the ontology was reviewed by primary care practitioners, clinicians, prostate cancer experts and ontology experts. The ontology will be published in the Bioportal repository (<https://bioportal.bioontology.org/>).

Benefits of ontologies:

- Systematic and reproducible data extraction (e.g. used in cohort studies)
- Transparent data extraction
- Building of comprehensive lists of clinical codes
- More accurate than based on clinical codes alone case finding
- A more comprehensive and inclusive cohort than that defined using clinical codes only, leading to a better understanding of patient journey and symptoms

### 3.3.2 Prostate cancer cohort

At the time of writing this section of the report (Feb 2020) we are in the process of undertaking the final prostate cancer cohort extraction. The cohort will be based on the developed ontology and using the most recent RCGP RSC database (March 2019). This cohort will be used in the future for data curation projects such as the one suggested for next year to curate prostate cancer outcomes in primary care data.

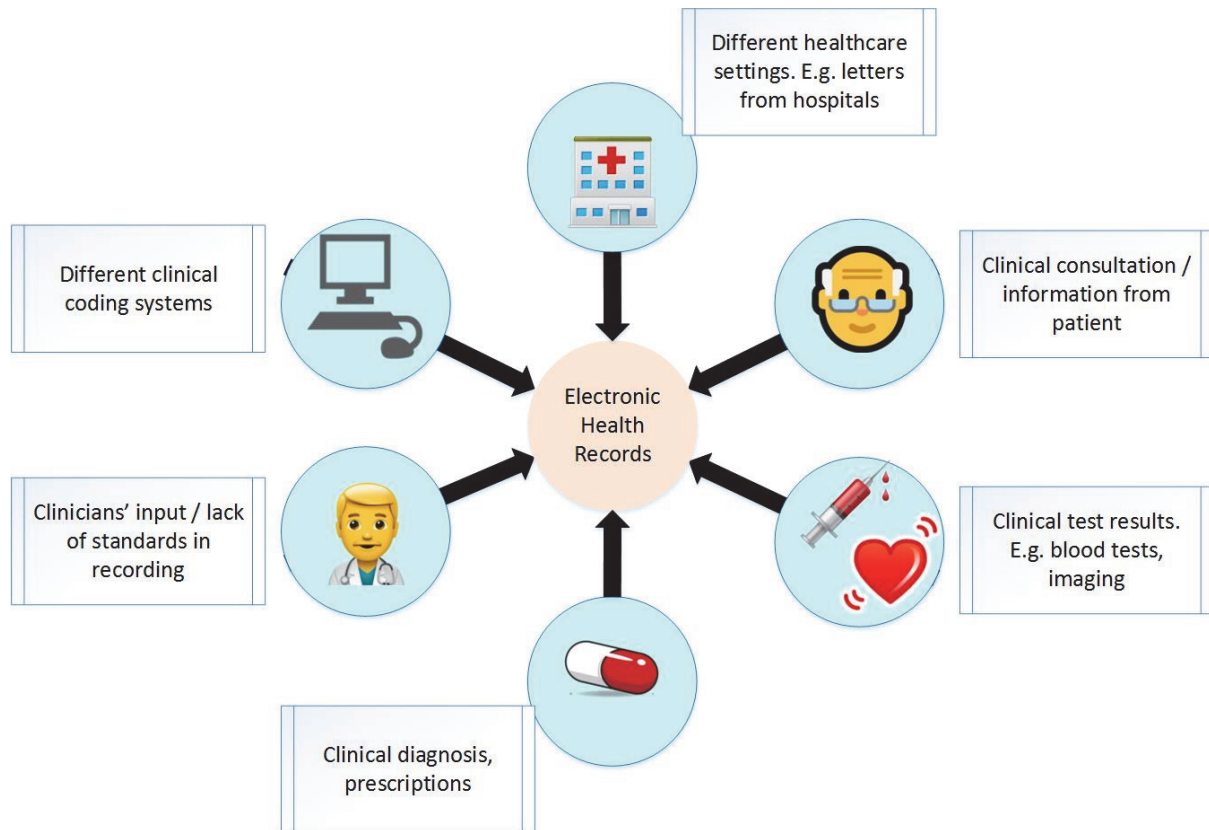
### 3.3.3 Conference presentation

An abstract was accepted and was published in the proceedings of Medical Informatics Europe 2020 in Geneva.

### 3.3.4 Learning from the case study

Although routine healthcare data are not collected for specific research purposes, they are increasingly being used in epidemiological research. The use of electronic health records (EHRs) is key for providing real-world evidence for improving healthcare by detecting and measuring consequences of diseases. However, this is hindered by the variable quality and

the lack of standards in recording clinical information. In this study, we identified and described a cohort of prostate cancer patients from an English primary care database to investigate the accessibility of data for epidemiological research. Routinely collected healthcare data can provide evidence for improving health and care of people living with and beyond cancer. In addition, the increasing digitalisation of healthcare records can facilitate large-scale epidemiological research in this area.



**Figure 3.6 Heterogeneity of primary care clinical data.**

### 3.4 FUTURE DIRECTION OF CASE STUDY

Future work could focus on treatment evaluation and improving the understanding of health outcomes that could be curated from primary care. Health outcomes in this context mean results of healthcare that patients receive and include patient's functional status, mortality, mental health, or satisfaction. Future work will develop standardised and reproducible methods to detect and measure key prostate cancer metrics such as symptoms and treatment side-effects in primary care data using ontological and data curation approaches. While prostate cancer is used here as an exemplar, the methodology will be applicable to other cancers and diseases. Figure 3.7 shows the potential health outcomes of prostate cancer and its treatment that could be curated. In the future, prostate cancer and treatment outcomes that are curated from routine data could be used to evaluate healthcare beyond RCTs.

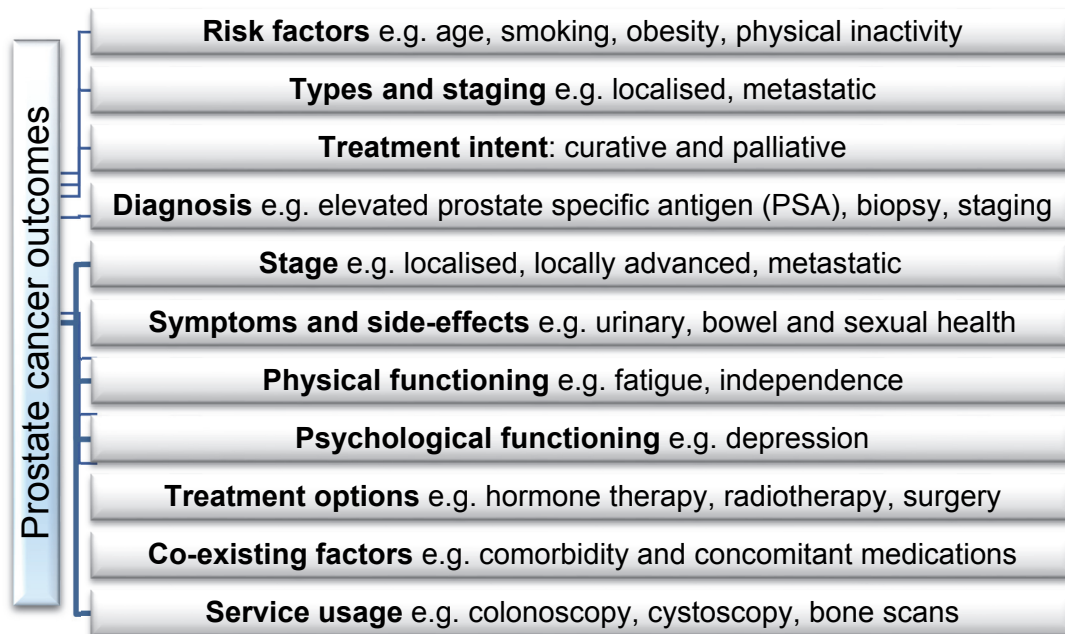
**Title:** An ontology-based approach for prostate cancer outcomes evaluation from routine primary care data: focus on patient centred symptoms and treatment side-effects.

**The aim** will be to curate, extract and measure key prostate cancer outcomes that could be used for the evaluation of healthcare.

Objectives



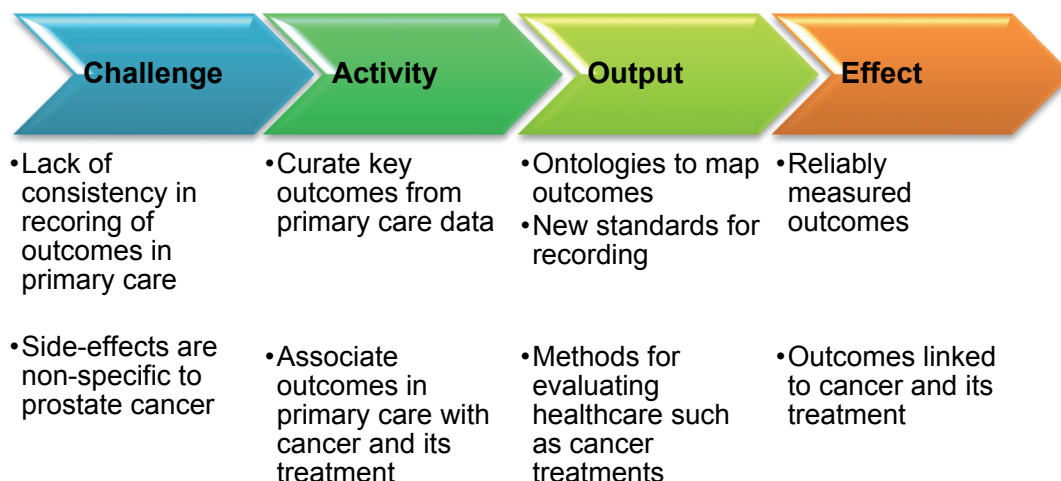
- Using an ontological approach developed in the current case study to extract prostate cancer cohort from RCGP RSC database
- Describe the incidence, demographics and clinical characteristics of the cohort
- Apply ontological approaches to establish key prostate cancer health outcomes
- Apply data curation approaches to detect and measure the key health outcomes
- Evaluate the quality of the routine data in this context.



**Figure 3.7 Potential prostate cancer outcomes that will be curated**

We will curate prostate cancer health outcomes from primary care. This is important for two reasons. Firstly, because side-effects of prostate cancer treatment are not reported consistently in primary care (there are no standards in reporting) and therefore they cannot be reliably measured. Secondly, because many of the side-effects are non-specific to prostate cancer. For example, urinary or sexual symptoms can develop as part of aging so cannot be directly linked to cancer and its treatment. Therefore, although it is challenging to find ways of associating outcomes in primary care with cancer and its treatment, it is necessary in order to evaluate healthcare such as radiotherapy. In trials, this is done by subtracting baseline symptoms (pre-cancer) from symptoms during and post-treatment. Some of this methodology could be used here but the approaches are not well developed for primary care data.





**Figure 3.8. Logic model for the proposed case study - curating prostate cancer outcomes from primary care data.**

In primary care practice, there are **variations in the recording of prostate cancer diagnosis and treatment**. This study contributes to improved understanding of the quality and accessibility of primary care data for research. The ontological approach provides a standardised and reproducible way of **measuring healthcare outcomes** and can facilitate **re-use** of routinely collected data for research.

## 4 CASE STUDY 3: COMPUTERISED MEDICAL RECORDS (CMR)

**Full title:** Data driven analysis of computerised medical records (CMR) with autoencoders

**Technical Lead:** Spencer Thomas

**Web:** <https://www.npl.co.uk/research/data-science/digital-health/case-study/cmr>

### 4.1 CASE STUDY PROPOSAL

Primary healthcare computerized medical records (CMR), also known as electronic health records, are a powerful source of information for decision making in healthcare settings. CMR data often contain highly detailed information. They are routinely collected during healthcare processes and can provide insight into individual level and population level statistics on healthcare. However, due to complexities in the data and a lack of standardisation, analysis of these data is extremely difficult<sup>98</sup>. As a result, CMRs are significantly under-utilised in research. In 2015, Campanella et al.<sup>99</sup> published a study evaluating the impact of using CMR on healthcare quality through a meta-analysis of articles in PubMed, Web of Knowledge, Scopus and Cochrane Library databases. Of the 23,398 entries relating to computerised or electronic records between 1994-2013, only 47 articles (0.2%) used the digital information to aid in decision making or assess benefits or improvements through using these data.

CMR data are often high dimensional and heterogenous and are a combination of multiple data types (text, numerical, categorical, etc) making tasks such as visualisation or data mining extremely challenging, and therefore hindering the insight that can be gained from these data. In addition, data may be missing or incorrect due to a lack of (or lack of adherence to) standards in recording the data<sup>100</sup>. This may be a consequence of resource limitation and time pressure on healthcare providers but creates challenges in uncertainty quantification.

There is a high risk in healthcare decision making based on bad quality data, but this can be addressed through improved reporting<sup>101</sup> or data correction methods<sup>102</sup>. Methods for analysing these data that are robust to these issues are starting to be developed that can identify clusters of patients<sup>103</sup> for visualisation or further analysis. However, standardised data will enable the use of a wider variety of analytical tools that may provide insight into healthcare systems. For decision making systems in healthcare, an understanding of the uncertainties on outputs of analysis is vital due to extremely high associated risks.

The potential in analysing these data is enormous as they can provide earlier indicators of individual, regional or national level healthcare issues enabling the appropriate planning and response or clinical decision making to be enabled as early as possible. For example, CMR data is used in epidemiology to identify potential disease outbreaks through population level health indicators<sup>104</sup> that can be used for estimations of disease incidence<sup>105</sup>. An important example of this is the Oxford-Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC)<sup>106</sup>, that collect CMR data for a number of respiratory conditions focusing on influenza.

The RSC is used to monitor influenza by the UK's Chief Medical Officer and deploy the necessary response<sup>107</sup>. The most reliable way of differentiating incident cases from prevalent cases is through the clinician assigning episode type on the patient's CMR<sup>108</sup>. This type of analysis can also be done on a regional level to provide spatially resolved healthcare information and insight. For individual or personalised medicine, the use of CMR has been shown to reduce the number of adverse drug effects<sup>109</sup>, improved glycaemic control and self-

care for diabetics<sup>110</sup>, optimal treatment planning such as improved palliative care<sup>111</sup>, early indicators of disease such as cancer trajectories<sup>112</sup>; Death, Intubation, or Transfer to ICU in patients with paediatric respiratory failure<sup>113</sup>.

For healthcare professionals, using CMR data has been shown to reduce documentation time and result in a lower number of medication errors<sup>114</sup>; improve resource management<sup>115</sup>; increase the utilisation of test results<sup>116</sup>, and improve administrative efficiency and reduce costs<sup>117</sup>. There is a clear benefit for utilising CMR data from both front-line and retrospective analysis perspectives.

This case study developed methods for analysing CMR data and showed how the results can be interpreted or further analysed. Specifically, we tested the generalisability of these methods for other datasets and the traceability of the data-driven clusters formed during the analysis. Records for patients with influenza like illnesses include an episode type attribute which are used distinguish between conditions that are new or ongoing. These data are vital for the monitoring of population level health and the identification of possible epidemics<sup>118</sup>. We aimed to develop a deep learning model that can predict episode type where missing, based on attributes within the CMR. Correcting data leads to a significant increase, up to 15%, of data that can be used for analysis, which is currently being discarded due to the missing episode type information.

We also aimed to develop another method for correcting CMR data entries based on their proximity to clusters in the population. This would enable all CMR data to be corrected or validated prior to further analysis, increasing the quality of the input data and hence the output of the subsequent analytical or forecasting model.

## 4.2 CASE STUDY TECHNICAL WORK

As a case study we use CMR data extracted from the Oxford Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC) <sup>119</sup>. All data are from English General Practices (GPs) and consisted of 2.4 million records from > 230 GPs at the time of extraction (2017) that was representative of the national population<sup>120</sup>. Data are uploaded to the RSC twice per week.

In this study we analyse anonymised data extracted from one of these uploads covering a six-week period in 2016 for 11,000 medical records each with attributes relating to the patient and condition. Deep learning methods for visualising these data have recently been introduced<sup>121</sup> and have been developed for prediction tasks. Specifically, we will develop and evaluate a deep learning model to correct erroneous and assign missing episode type entries in CMR data from the RSC based on other attributes in the CMR. Using a selection of data from the RSC we will train a deep learning model to predict the episode type based on entries in the CMR. We will validate this model with data that was not included when training the model with known episode type.

Furthermore, we will also develop methods to correct erroneous entries in the data generally based on the content of the CMR. Here we will take a subset of the RSC data and train a deep learning model to identify clusters in the data. We can evaluate the robustness of this method by adding known levels of perturbations to the CMR data prior to input to the model and comparing the corrected output to the original unperturbed data.

### 4.2.1 CMR Extract from the RGCP RSC Database

The data used here consist of a dataset extracted weekly from the RGCP RSC database of anonymised patients with influenza-like illness in 2016, as seen in Table 4.2. Each weekly

extract consists of the records from the previous week and any records that have been added or updated in the previous 5 weeks. Hence this typically consists of data over a 6-week period but can also include historical records outside of this window that are retrospectively updated and therefore included in the RGCP RSC weekly data upload. The data extract that was used in this study consists of 11,000 patients with the attributes as listed in Table 4.2 as well as pseudonymised patient ID, deprivation score and Registration Status. The latter three are excluded from the analysis.

**Table 4.2 Example of primary healthcare computerised medical record (CMR) data**

GP ID	Read Code	Read Term	Clinical System	Event Date	Episode	Age	Gender
129	H27z.	Flu like illness	EMIS Web	2016-10-31	O	33	F
147	H06z0	Chest infection	INPS Vision	2016-11-04		60	M
96	H06z0	Chest infection	EMIS Web	2016-10-28	F	37	M
133	H061.	Acute bronchiolitis	EMIS Web	2016-10-17	N	0	M
19	H060.	Acute bronchitis	EMIS Web	2016-09-30	F	59	F
2	XM0rz	Influenzalike illness	TPP SystemOne	2016-10-21	N	42	F
145	H061.	Acute bronchiolitis	EMIS Web	2016-11-03	F	0	M
120	H06z0	Chest infection	EMIS Web	2016-10-25		68	F
27	H06z0	Chest infection NOS	EMIS Web	2016-10-13	O	77	F
47	H27z.	Flu like illness	EMIS Web	2016-11-03	F	23	M

#### 4.2.2 Data Pre-processing

As inputs to the classification algorithms, we convert the CMR data into a numerical matrix using ASCII encoding as the data contain numerical, date, structured and unstructured text, and categorical data. Each specific data type is processed as follows. All numerical fields are converted to doubles. Text data (categorical, structured and unstructured) are first converted to integers using ASCII encoding for each character in a string, and each integer is converted to numerical doubles as above. For text data, where instances in the CMR differ in length, first the longest entry in the dataset is identified and all other instances are padded with whitespace characters at the end of the string to conform to the same length.

Encoding the text data in this way and using an autoencoder provides robustness against errors in the entries such as spelling mistakes and distinguishing different orders of the same words. As an autoencoder contains fully connected layers, entries with similar content and ordering will be identified, such as ‘back **pain**’ and ‘back **oain**’. For date variables, the attributes are read in as text strings in order to account for differing formats across datasets, and converted to the ISO 8601 format yyyy-mm-dd. These variables are then separated into three numerical values corresponding to the year, month, and day of the records to account for any regular patterns such as seasonal variation. Once the attributes have been converted into numerical data as described above, the values for an instance in the CMR is concatenated together to a single numerical vector of length  $M$ . This is performed for all  $N$  instances in the data yielding numerical matrix with dimensions  $N \times M$ . Specifically for the data used in this work we obtain a matrix of 11,000 instances each with 93 dimensions. We use this  $N \times M$  matrix for all our subsequent analysis and refer to this matrix as  $A$ .

#### 4.2.3 Autoencoder

An autoencoder is a special class of neural network that is structurally symmetric and is designed to first encode the input data and then decode it back to the input data. This feature enables the autoencoder to be trained in an unsupervised manner as the network learns an encoded representation of the data that can be decoded back to the input data. Typically, data are encoded to a lower dimensionality than the input data and it is possible to stack

these networks in order to obtain successively ‘deeper’ representations of the data. The original data  $X$  is a matrix of size  $N \times M$  ( $N$  instances and  $M$  dimensionality) and the encoded data  $Z$  is a matrix of  $N \times D$  where the reduced dimensionality  $D$  depends on the autoencoder architecture. The data are encoded through the activation function  $\sigma$ , weight matrix  $W$  and bias vector  $b$  via

$$Z = \sigma(WX + b).$$

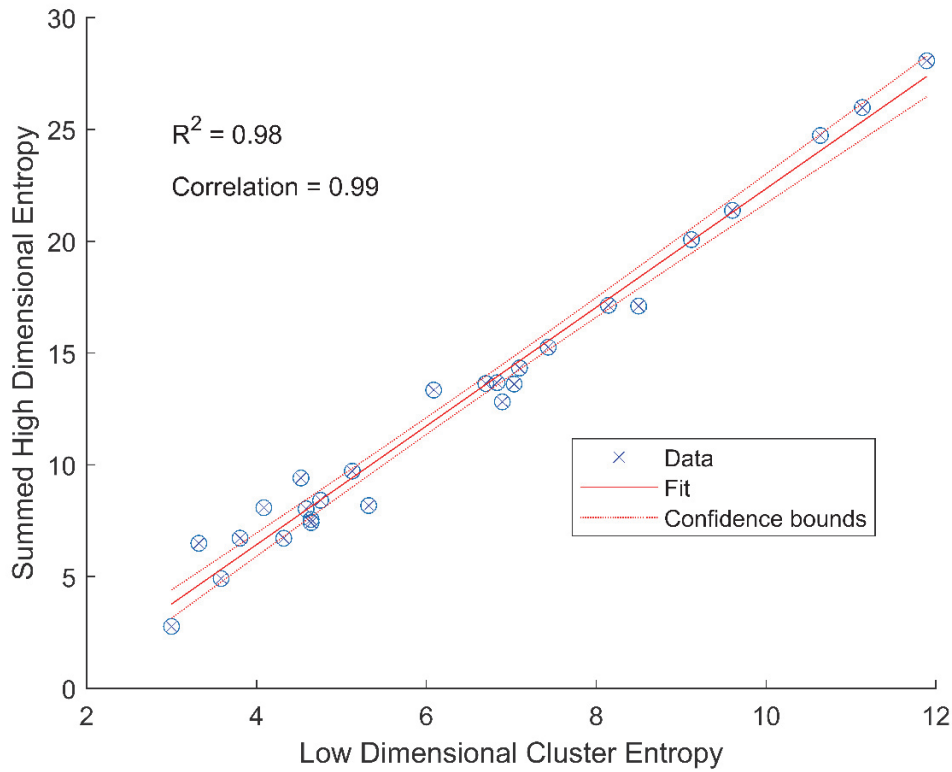
The weights of the network can be obtained by training the network using gradient descent in an unsupervised manner by reducing the mean squared error between the input data  $X_{ji}$  and the reconstructed output data  $X'_{ji}$ :

$$E = \frac{1}{N} \sum_j \sum_i^M (X_{ji} - X'_{ji})^2 + \lambda \Omega_W + \beta \Omega_S$$

The terms  $\lambda \Omega_W$  and  $\beta \Omega_S$  are regularisation terms for the weights and sparsity of the network to prevent overfitting. Once trained, the model will identify data-driven patterns such as clusters that can provide insight into sub-populations within the data without the need for priori knowledge or assumptions<sup>122</sup>. Interpreting the clusters as probability distributions we can make use of Shannon Entropy<sup>123</sup>,  $H(X)$ , defined as:

$$H(X) = - \sum_i^n P(x_i) \log_2 P(x_i)$$

where  $P(x_i)$  is the probability of  $x_i$ . This enables the calculation of  $H(X)$  when  $X$  is numerical, categorical, date or text data, by computing the number of possible outcomes and their frequencies within cluster. We can compare this in both the high dimensional input data and the low dimensional representation learned by the network. As can be seen in Figure 4.1 there is a direct correlation between these representations verifying that the autoencoder maintains the characteristic of the data in the low dimensional space.



**Figure 4.1 Correlation between entropy in the high dimensional input data and the low dimensional representation learnt by the autoencoder. The graph shows that the entropy characteristics of the data are maintained in this encoding. The confidence bounds are from fitting the model to the data.**

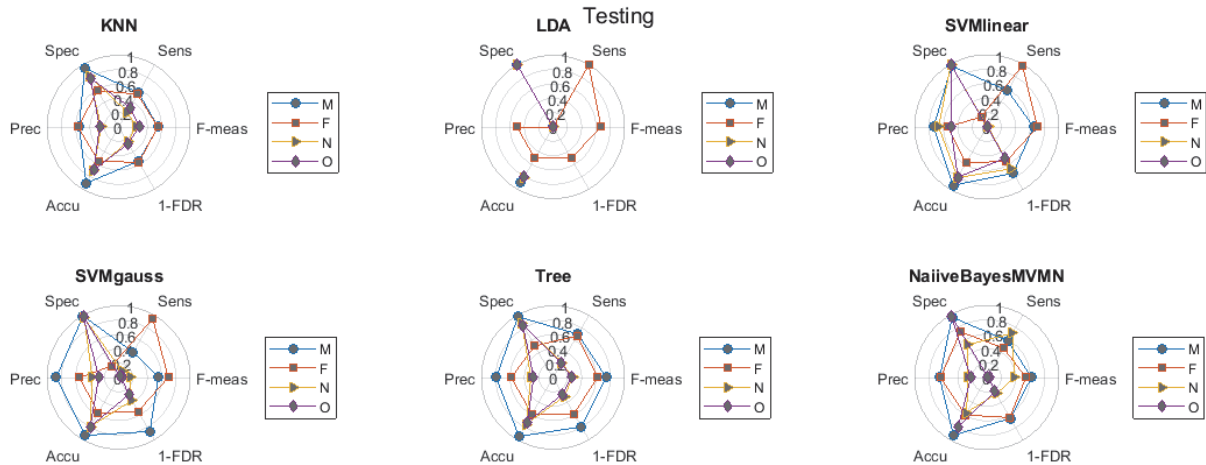
#### 4.2.4 Prediction of Episode Type

We investigate several machine learning algorithms for predicting episode type for the data in Table 4.2 following the data pre-processing specified above. Specifically, we use k-nearest neighbours (KNN), linear discriminate analysis (LDA), support vector machines (SVM) using a linear or Gaussian kernel, a decision tree (Tree) and naïve Bayes using a multivariate and multinomial class distribution. For details of these algorithms please refer to Thomas et al.<sup>124</sup>. We evaluate the classification performance using common classification metrics; accuracy (Accu), precision (Prec), specificity (Spec), sensitivity (Sens), F-measure (F-meas) and the false discovery rate (FDR). These are calculated per class and represented in polar plots, for further details please refer to Thomas et al.<sup>125</sup>

To build a classification model we split  $A$  into the episode type data (now a 1D vector) to use as labels for our data,  $L$ , and matrix  $A'$  which contains all remaining data in the  $A$ . Specifically

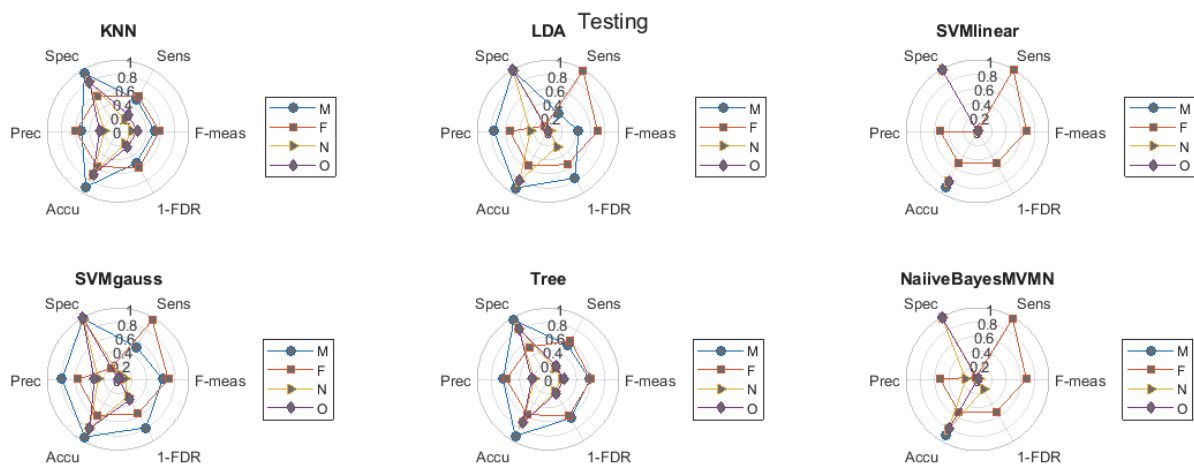
$$A = [A' L]$$

We use a random 90:10 split of the data for training and testing and use the subset of  $A'$  to train the models to predict the corresponding subset of  $L$ . When using  $A'$  directly we obtain the results seen in Figure 4.2 where no method is able to give robust performance (determined via the metrics) for all episode types.



**Figure 4.2 Testing performance for several machine learning models for predicting episode type from CMR data.**

Due to the high dimensionality of the data and the ability of deep autoencoders to identify patterns in data<sup>126</sup>, we consider classification of the data as in Table 4.2 and data reduced using an autoencoder<sup>127</sup>. These results are given in Figure 4.3 which overall are comparable with Figure 4.. KNN, SVM(Gauss) and Tree all show similar performance, whereas SVM (linear) and Naïve Bayes (MVMN) are drastically worse and LDA is significantly improved.



**Figure 4.3 Testing performance for several machine learning models for predicting episode type from CMR data reduced with an autoencoder.**

Both Figure 4. and Figure 4.3 show that some models are able to predict missing and first episode type to a high accuracy but moderate to low sensitivity. However, the models struggle with new and ongoing episode types. There is potential to repeat this with fewer classes, i.e. M, F, N+O, or as a binary problem i.e. missing and not missing. Note that in both, the full CMR and Deep Autoencoder (DAE) reduced data, we observed the same performance scores when repeatedly training the algorithms for different random splits for the training and testing data hence this is not a small batch effect and is reflective of the data. Using a larger dataset or more attributes in the CMR records would lead to better

performance. Further testing is required with a larger dataset that includes more patients and a longer time frame.

#### 4.2.5 Correcting Episode Type

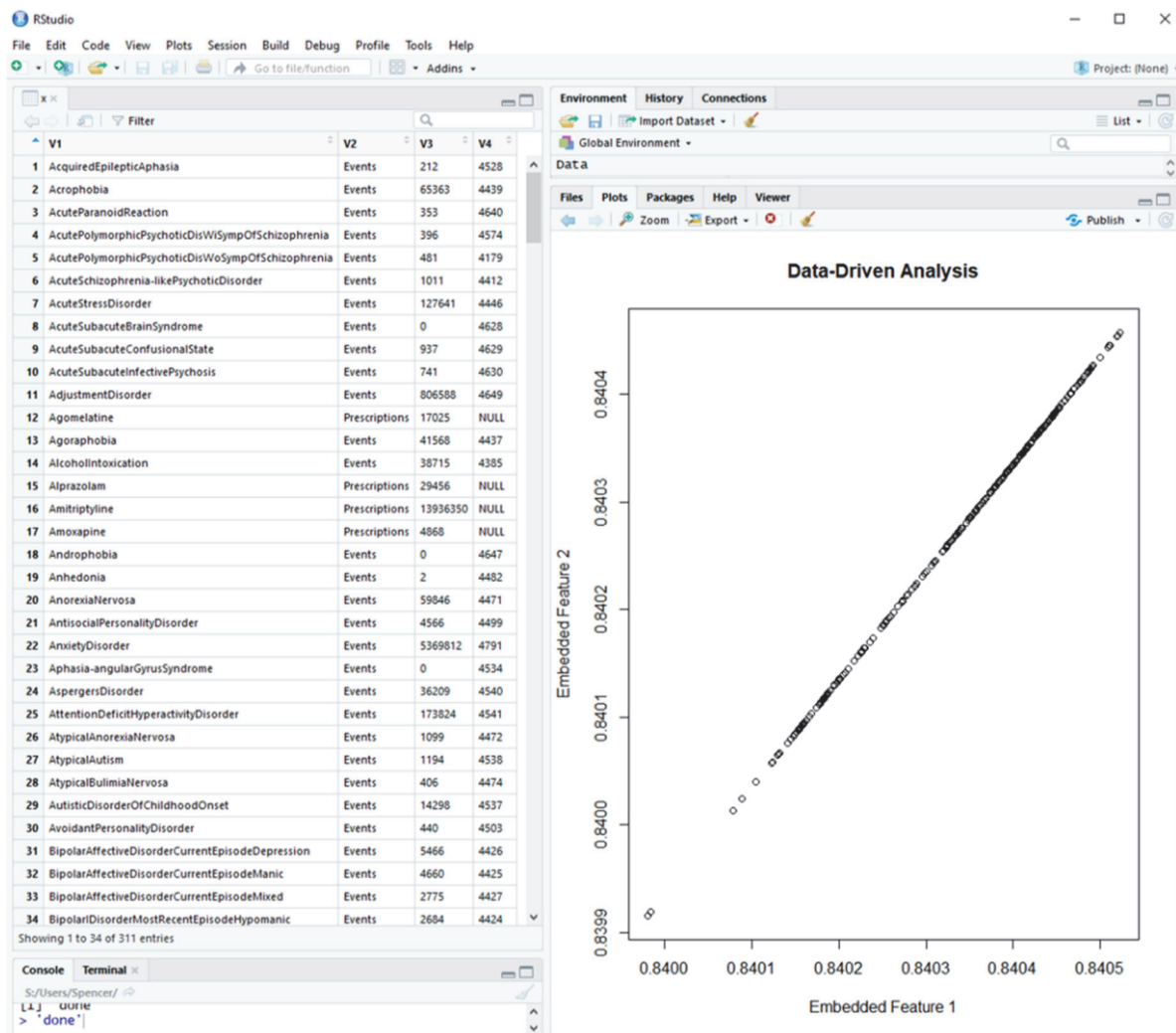
Although the machine learning algorithms are able to predict missing episode type to a high level of accuracy, the poor performance for the other episode types makes it difficult to develop a correction tool as the data is not sufficient to identify what inputs are common in the correct types. Further work with a larger dataset may help solve this issue.

### 4.3. OTHER RELATED PROBLEMS

#### 4.3.1 Implementation of data driven methods on a clinical system

We have implemented this data driven method (pre-processing and autoencoder) within the RCGP RSC secure server to allow its application to a pseudonymised patient level data within the database. Figure 4.4 is a screen shot from R studio within the RCGP secure network showing a demonstration of the data driven method on a test dataset. Here we use a dataset of events and prescriptions relating to a specific extract in the database (left) and provide a scatterplot from the data driven analysis method (right) Note that as these are all unique entries they appear as a line distinct points in the low dimensional space rather than clusters of subgroups (as there are none in this case). Points that are more similar to each other (e.g. letter content or ordering) are closer together and form areas of high density along the line.





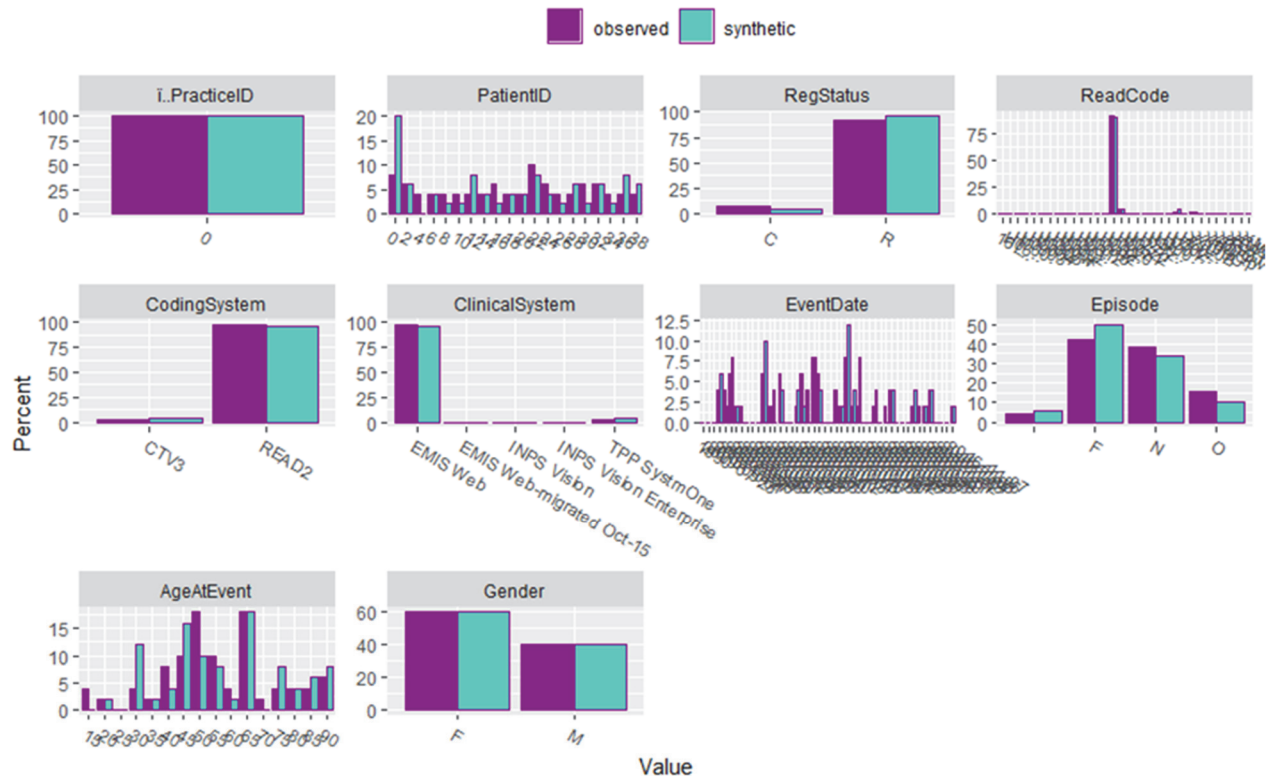
**Figure 4.4 Screenshot of data-driven autoencoder method implemented in R on the RCGP secure server. The data (left) is processed and analysed with the autoencoder to provide the scatter plot (right). Note the data is a list of events and prescriptions in an extract of the RCGP database hence no 'clusters' are expected.**

#### 4.3.2 Synthetic CMR Data

Primary care CMR are a rich source of data for population and individual level decision making in healthcare. However, they remain significantly under-utilised due to their complexity but also their limited availability. Restrictions on data availability are due to patient identification, data security, user training and regulations such as GDPR that limit the ability to train models to analyse these types of data and the impact on the healthcare system. Generating synthetic data provides the potential to access large volumes of data to train models such as deep learning, while avoiding these issues completely as the data are artificial and do not relate to real patients. The challenges are in the development synthetic datasets that are reflective of real datasets, which requires access to real and well-curated data. Here we use the computerized medical records (CMR) extracted from the RCGP RSC database<sup>128</sup> relating to influenza like illnesses.

We used the Synthpop package in R<sup>129</sup> to generate synthetic data. This is able to generate synthetic data of mixed types based on Bayesian statistics and is therefore an attractive tool for this purpose. This algorithm generates synthetic datasets based on original data by using its attributes as a posterior predictor.

The original data sample is made up of  $n$  units consisting of  $(x_{obs}; y_{obs})$  where  $x_{obs}$  (which may be null) is a matrix of data and  $y_{obs}$  is an  $n \times p$  matrix of the  $p$  variables to be synthesised. The observed data,  $x_{obs}$ , are assumed to be a sample from a population with parameters,  $\theta$ , that is to be estimated;  $y_{obs}$  is assumed to be a sample from  $f(Y|x_{obs}, \theta)$  where  $f$  is a vector of parameters. The synthesised data are generated from this distribution and obtained estimates of these parameters. An illustrative example of this method using the first 500 records in the CMR is shown in Figure 4.5.



**Figure 4.5 Distribution of original and synthetically generated data for some of the variables in the CMR data. Data shown are generated using the first 500 records for illustrative purposes.**

The synthetic data closely follow the data extracted from the RCGP database on an attribute level, some differences are observed but this is to be expected with a probabilistic model. Using a larger dataset may lead to improved results, though more work is required to look at not only the distribution of individual attributes, but also their relative distribution. That is, the individual attributes may be well modelled, but the entire synthetic record may not correspond to the RCGP data due to sample bias in small populations.

This can be tested using the CMR analysis techniques shown above and/or by using similarity metrics between the real and synthetic datasets. Here we estimate  $\theta$  for each attribute based on all other attributes and hence each attribute in a record is based on all other values. For complicated data, such as CMR, this may not be the case and some attributes may not have any relevance on others. Synthpop allows users to discard variables as predictors of other attributes in order to build a strong synthetic dataset and the key factors for each variable can be obtained via linear regression or optimisation methods. An example of this is given in Table 4.2. Much more work is required in order to identify the robustness and optimisation of this method and the suitability for generating large datasets to train machine learning models.

**Table 4.2 Prediction of attributes in the synthetic data based on specific attributes in the data**

Variable	Predictors	R <sup>2</sup>
Practice ID	Patient ID, Registration Status, Coding System, Clinical System, Age, Deprivation Score, Gender	2.61%
Patient ID	Practice ID, Registration status, Read code, Coding System, Clinical System, Age	10.6%
Registration Status	Practice ID, Read Code, Clinical System, Age	2.25%
Read Code	Practice ID, Coding System, Clinical System, Age, Gender,	53.94%
Read Term	Registration Status, Coding System, Clinical System, Age, Gender,	53.19%
Coding System	Practice ID, Patient ID, Clinical System, Deprivation Score, Episode, Read Term, Read Code	20.82%
Clinical System	Practice ID, Patient ID, Coding System, Deprivation Score, Age, Episode	20.38%
Age	Practice ID, Patient ID, Registration Status, Coding System, Deprivation Score, Gender, Episode, Read Term, Read Code	14.72%
Deprivation Score	Practice ID, Coding System, Clinical System, Age	0.9%
Episode	Coding System, Clinical System, Age,	11.89%
Gender	Practice ID, Age, Episode, Read Term, Read Code	1.3%
Date of Event	Registration Status	0.6%

#### 4.3.3 Generalisation of CMR and Medical Metadata Analysis

##### **International Skin Imaging Collaboration (ISIC)**

A large collection of digital skin images from the International Skin Imaging Collaboration (ISIC) Melanoma Project have been collated, processed and classified by expert dermatologists<sup>130</sup>. The data consist of thousands of skin images that have been made available for clinical support and the development of automated diagnosis systems such as deep learning techniques. A subset of this data set is the HAM1000 dataset which contains 10,015 digital images of skin lesions from eight classes of skin conditions that have been categorised by clinical experts. Each of the 10,015 images in the HAM10000 dataset has an associated metadata file in JSON format containing several fields relating to the clinical and acquisition aspects of the images. The clinical fields contain a small amount of patient information including diagnosis of the images. The acquisitions fields relate to the pixel dimensions and acquisition type for the images. The acquisition type is the same for all images and the number of X and Y pixels varies across the dataset.

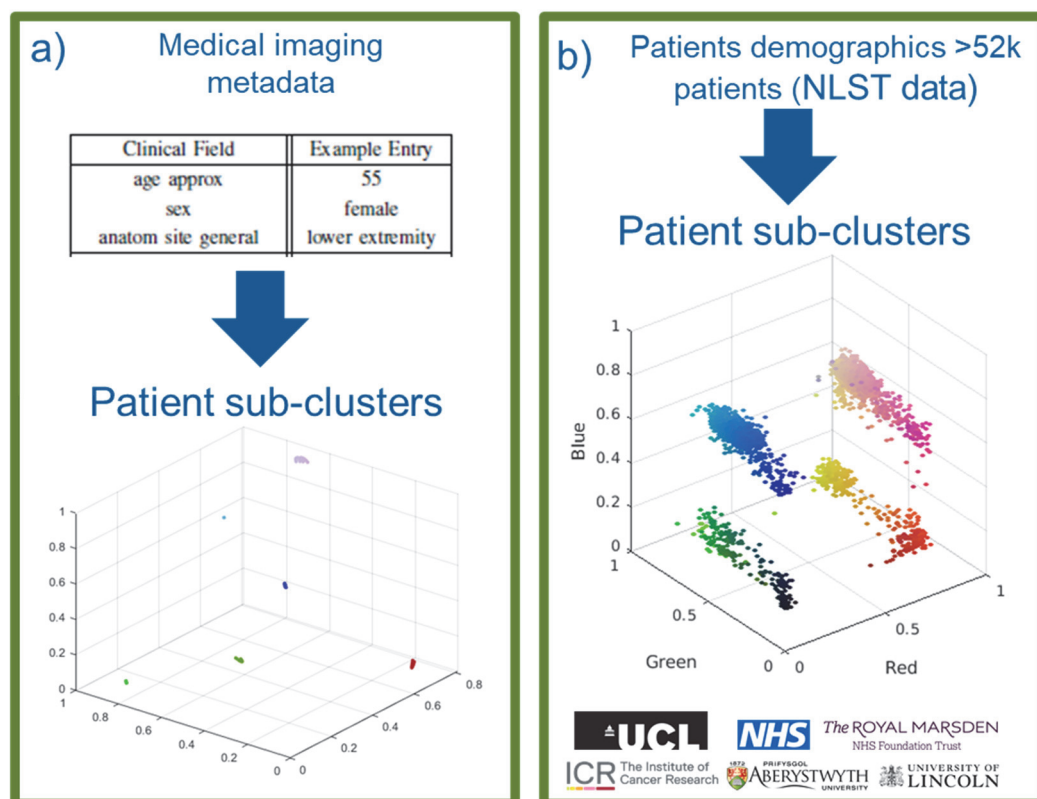
To test the generality of the method for analysing CMR data, we apply the same numerical transform as for the CMR data for the clinical fields, but not diagnosis related fields, from the ISIC HAM10000 image metadata. Specifically, the fields include approximate age, sex and anatomical site of the lesion. Once we obtain a numerical matrix asking similar to  $A$  for the CMR we again train a deep autoencoder to perform non-linear unsupervised reduction of these data, resulting in the formation of clusters of patients in the latent space of the DAE Figure a). The patterns driving these clusters and further interpretation of these data is still ongoing.

### National Lung Screening Trial (NLST)

This project resulted from a successful CRUK sandpit award to a consortium led by the Institute of Cancer Research (ICR) involving NPL, UCL, the Royal Marsden, University of Lincoln and Aberystwyth University. The goal of the project is to improve the prediction of cancerous lung nodules by combining the CT imaging and demographics data. NPL's role is to use the methodology developed in this project to analyse the demographics data.

The National Lung Screening Trial (NLST) is a publicly available dataset consisting of low dose CT images for ~25k patients each with demographic information for each patient. The number of images for each patient varies but typically they have 3 scans and potentially additional demographic data collected at each scan. Therefore, one patient may have multiple and differing numbers of scans and demographic information.

Here we focus on patient demographics. In total the demographics data consists of > 52k records with 322 fields. However, many these fields contain missing data, 201 fields are missing for > 50% of the records, 181 fields are missing for > 90% of records. We remove the 201 fields that are missing for >50% of the records and then apply a deep autoencoder as previously. This reveals patterns in the data as illustrated in 4.6b). The drivers behind the formation of these clusters and interpretation of these data is ongoing.

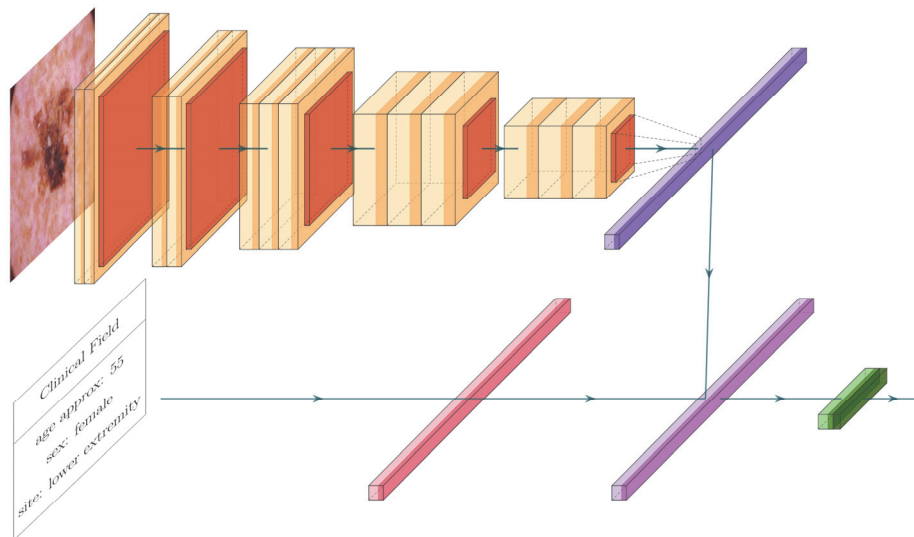


**Figure 4.6 Application of the DAE analysis of CMR data to a) metadata of medical imaging from ISIC<sup>131</sup> and b) patient demographics data from NLST<sup>132</sup>. In both cases**

**there are cluster observed in the low dimensional space that relate to patterns and subpopulations in the data.**

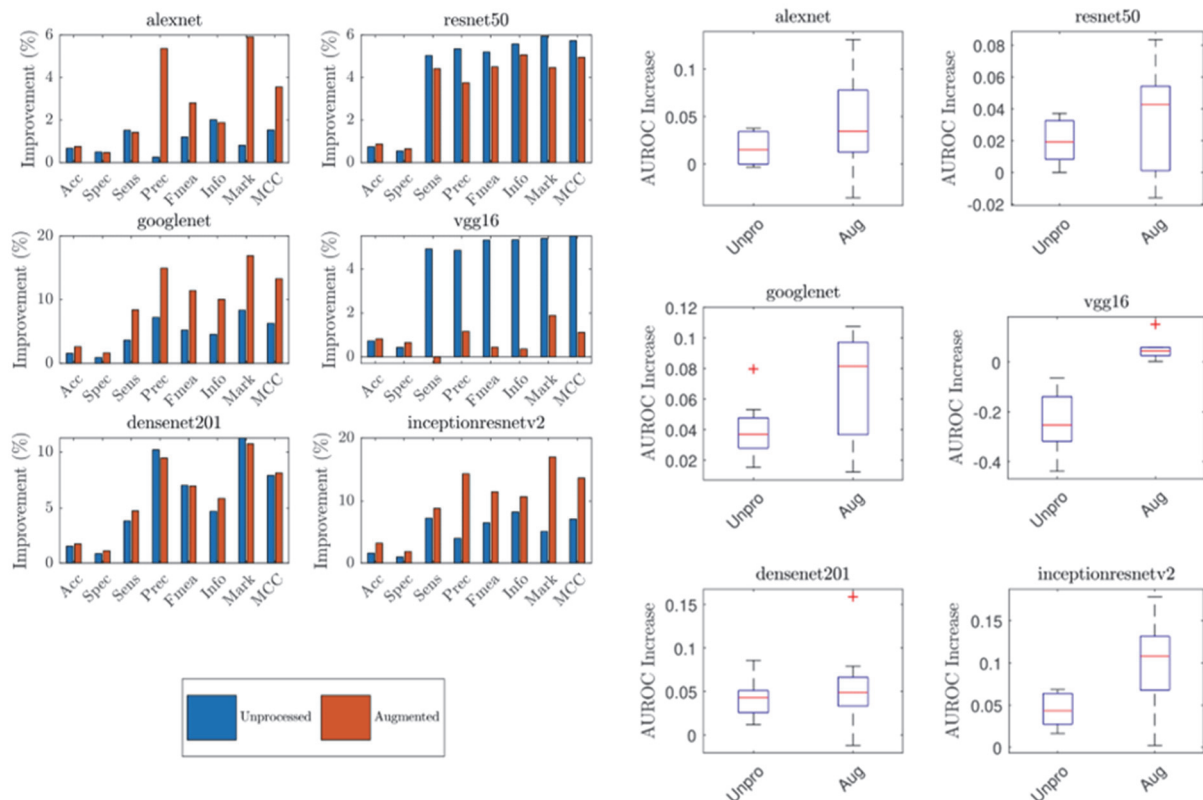
#### 4.3.4 Integrating Imaging and Non-Imaging Data

Clinicians and medical practitioners will often have access to additional patient information when diagnosing medical images hence training algorithms solely on the images potentially hinders their performance. Utilising the availability of the large medical imaging dataset and corresponding metadata from ISIC<sup>133</sup>, we investigated if combining the imaging and non-imaging data can improve classification of medical imaging data as illustrated in Figure .7.



**Figure 4.7 Combination of imaging and non-imaging data for medical classification tasks using deep convolutional neural networks.**

Here we use several popular convolutional neural network (CNN) deep learning architectures and utilise transfer learning based on a model pre-trained on the ImageNet dataset<sup>134</sup>. We compare the performance of classifying the data based only the image data, or a combination of image and metadata. Further, we also compare the use of image augmentation (here random rotation and translation) which is commonly used to improve the performance of CNN; and using the pretrained model directly via bottleneck feature extraction on the skin images<sup>135</sup>. The results of these different settings are shown in Figure 4.8 by averaging the performance per class in the ISIC HAM10000 dataset and averaging over the performance metrics shown. The inclusion of the metadata improves the classification performance regardless of the network or the use of image augmentation or not and is shown in all performance metrics. The aggregated improvement is up to 15 % and adds a negligible computational cost to the classification task, particularly when compared to the CNN. Further work is needed to try and improve these results further as well as to test the generality of this for other deep learning models and datasets.



**Figure 4.8 Aggregated performance of classification of the ISIC data by averaging over each class. Classification is based on images or images and metadata combined for several metrics (left) and class-wise boxplots of increase in area under the receiver operator characteristic curve (AUROC). Positive results indicate an improvement when including the metadata in the classification.**

#### 4.4 CASE STUDY OUTCOMES/HIGHLIGHTS

##### 4.4.1 Publications

The methodology for the analysis of CMR data using Deep Autoencoders was published in a peer reviewed journal, *Frontiers in Applied Mathematics and Statistics*<sup>136</sup>, jointly with collaborators from Surrey University and the RCGP RSC. An additional manuscript from the peer-review conference proceedings of Medical Informatics Europe Conference 2019 (MIE 2019) was published in *Studies in Health Technology and Informatics*<sup>137</sup> jointly with collaborators from Surrey University and the RCGP RSC. A paper based on the results of Figure 4.8 has been submitted as a conference paper to 43rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society.

##### 4.4.2 Presentations and Engagement

These results have been presented at a number of internal meetings such as;

- STAC review for Health and Life Science (Teddington, November 2019)
- PEG review for Health and Life Science (Teddington, February 2020)
- Data Science Strategy in Healthcare (Cambridge, November 2019)
- NPL Transfer Learning Workshop (Teddington, October 2019)

Additionally, this content has been presented to external visitors including:

- Angela Douglas Deputy Chief Scientific Officer NHS England; Catherine Ross Clinical Leadership Lead (Office of the Chief Scientific Officer); Mary Abodunrin Team Coordinator (Office of the Chief Scientific Officer) (October 2019)



- Zoe Webster Director, AI and Data Economy, Innovate UK (November 2019)

#### 4.4.3 Other Highlights

Related to this project members of the team are supervising a PhD student jointly with University of Surrey to forecast outbreaks of flu and hay fever based on the RGCP RSC data to inform resource allocation ahead of peak outbreaks. In particular, the use of deep learning and ability to analyse the data in this case study will be leveraged in the PhD studentship, enabling further engagement with the RGCP, University of Surrey, University of Oxford and the healthcare community more broadly.

#### 4.5 FUTURE DIRECTION OF CASE STUDY

For the future, if continued we will continue development of the methodology and testing of the generality of the DAE analysis. This will focus on the interpretation of the results to improve the uptake of the method. Interpretation of the ISIC and the NLST data provided within this study, will help improve the generalisation of how to interpret these data. Following this, we will aim to implement this methodology to the same platform as the RGCP RSC database. Once completed we will translate this to the RGCP RSC system to enable access to larger datasets without requiring data extractions. Additionally, we could develop this as a tool for users of the RGCP RSC database to identify patterns in various datasets and provide insight into the data within the database such as cancer, diabetes and respiratory conditions.

Further work on developing the methodology for combining imaging and non-imaging data in medical classification tasks will continue using the ISIC dataset. Particularly the focus will be on the generalisation of application to deep learning algorithms and datasets, as well as methods for identifying the reasons for the improvements in terms of the input features. Application to other datasets in the ISIC archive (here we focused on the HAM10000 datasets) and to other datasets such as the NLST datasets through the CRUK sandpit project. We will also look to linking with other projects where possible, such as the digital pathology NMS project.

Some initial follow up on the prediction of Episode type work is planned whereby the task will be to first predict if the episode is missing and then attempt to correct it. Given the small size of the dataset this will be piloted and then reviewed early in the project to see if it is feasible to continue.

## 5 CASE STUDY 4: SYNTHETIC DATA

**Full title:** GENERATION OF SYNTHETIC DATA FOR MEDICAL PURPOSES

**Technical lead:** Ignacio X. Partarrieu

**Web:** <https://www.npl.co.uk/research/data-science/digital-health/case-study/synthetic-data>

### 5.1 CASE STUDY PROPOSAL

For this case study, the interest in synthetic data lies in the realm of medical imaging data (see Section 4.3.2 for generation of synthetic CMR data). Synthetic data was defined to be data not physically acquired. Healthcare data is a valuable resource when properly cleaned and analysed, as has been evidenced during the current health crisis<sup>138</sup>. Such data can be used to provide insights into disease diagnosis and patient prognosis, and to provide novel understandings that aid in patient treatment. This means that there is demand from governments and the public alike that new technologies be developed and deployed to healthcare services to assist clinicians reliably<sup>139</sup>. However, real data is often unstructured, complex, and difficult to access due to legal anonymization requirements such as the general data protection regulations (GDPR)<sup>140</sup>. It is thought that synthetic data, i.e. data engineered to maintain the essential characteristics of an original dataset, but that is procedurally generated, so lacking privacy issues, might be able to help researchers extract some of the insights available from a dataset without some of the aforementioned drawbacks.

Some additional advantages of synthetic data are that due to its engineered nature it can be produced as a well-structured dataset from the outset, limiting some of the common issues of interoperability related to unclear metadata sometimes associated with clinically acquired data. The generative process can be employed to produce datasets both larger than the original and containing more data points of rarer conditions providing balanced classes by modifying the distribution. This means synthetic datasets could also prove beneficial for machine learning (ML) algorithms<sup>141</sup>.

The scope of this case study was to survey the various methods of creating synthetic datasets from small training datasets. One such method was then to be implemented. There are many possible approaches, which often depend substantially on the types of data used. For medical imaging data convolutional neural networks and kernel density estimation algorithms have shown to be promising<sup>142</sup>, producing simulated results based upon an example dataset. Such work has also been undertaken for unstructured healthcare records using generative adversarial networks<sup>143</sup>.

Alternatively, a simpler approach known as data augmentation may be used where images are slightly modified by the application of filters and other operations such as cropping, translation, flipping, scaling, jittering, shearing, etc<sup>144</sup>. Appropriateness of these is dependent on the domain, for example when there is asymmetry in the data that may occur when flipping an image of lungs, it might be less appropriate. The most rigorous approach is typically considered to be a computation of the image in a virtual scanner using a virtual test object, which relies on a precise understanding of the mechanics which create an image and an accurate implementation of these. The advantage to such an approach is improved control over uncertainties and artefacts in the dataset, but it generally comes at the cost of increased computational complexity.

The care pathway for patients often includes the use of various forms of imaging, such as computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET) and ultrasound (US) amongst others. It is thought that these imaging



modalities often contain complementary information, and if appropriately processed this could be used to improve the patient experience and patient outcomes. In Section 5.2 the implementation of a generalisable method that could potentially be applied to varying datasets by capturing their essential characteristics is given.

Other examples of synthetic data are situations where images that have been acquired are enhanced or modified to have the characteristics of a different acquisition modality, such as the conversion of MRI data to PET data or vice versa, or even from MRI sequence to another<sup>145</sup>. This is useful in situations where it is not plausible to run all the necessary scans due to time and cost constraints, or in situations where scans of one type provide information which can be enhanced by combination with properties of another modality.

## 5.2 CASE STUDY TECHNICAL WORK

The aim of this case study was to establish a method that could overcome several key trials in order to produce data of value. In particular, the method implemented needed to:

- generalise well, i.e. be able to produce models for different scenarios with minimal tweaking;
- have code that ran within reasonable timeframe for iterative purposes;
  - for our purposes, a reasonable time frame could be considered to be less than 3.6 hours to produce initial results on a high spec workstation;
- be tested qualitatively and quantitatively on a per image and batch basis;
- have a gold standard against which to compare outputs.

To test the developed method, it was essential to use a large dataset of paired images that measure a quantitative physical value in a patient. This would allow for writing a method to convert from one dataset to another and for a quantitative examining of the resulting images. One dataset with the desired characteristics was the VoxTox data described in Section 2.

### 5.2.1 Specific study goals

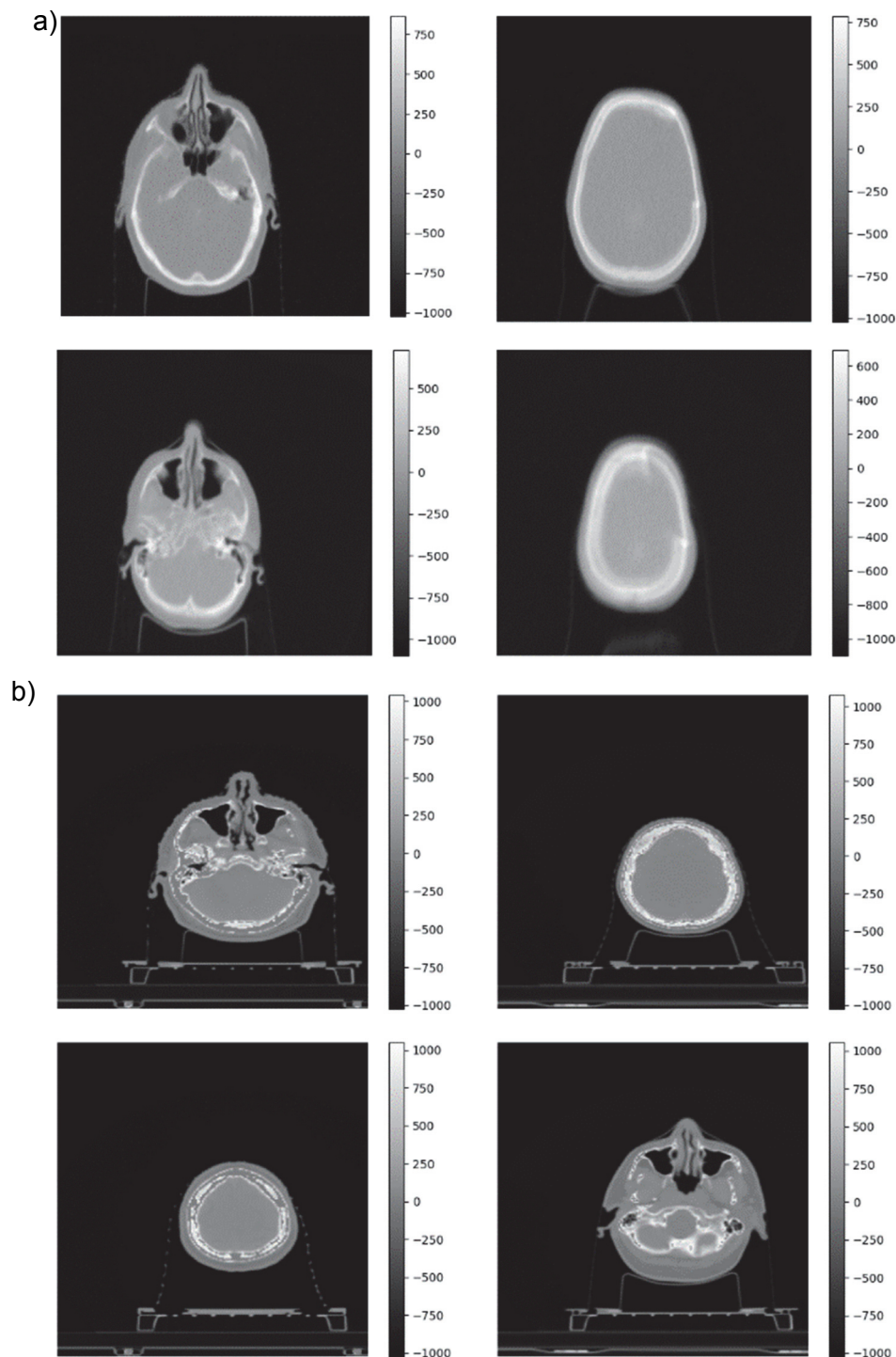
The aim of this study was to create synthetic CT data from the MVCT data, and vice versa. This synthetic data could then be assessed through comparison of the resulting synthetic key characteristics to the original datasets.

### 5.2.2 Dataset

The VoxTox dataset consists of radiotherapy (RT) treatment planning data. For RT, typically one, high quality CT scan is acquired for contouring the tumour region of interest (ROI). Before each treatment is applied, a low resolution and contrast scan is acquired on a megavoltage CT (MVCT). This results in many MVCT scans for any particular CT acquisition, typically 20 or more<sup>146</sup>.

Although the CT is typically well calibrated and has a drift <1%, this is not the case for the MVCT. In fact, components of the MVCT tend to wear out at a faster rate than for CT and so the data acquired can become noisier over time, even though there are no true underlying changes<sup>147,148,149</sup>. In this dataset, two different MVCT scanners are used and patients are typically scanned on both. These characteristics mean MVCT images tend to be both noisier than CT images, and less accurate as they struggle to discern the correct density at more extreme values (e.g. Figure 5.).

An additional characteristic of this dataset is that the CT scans are reconstructed with a larger field of view (FOV) than the MVCT scans, meaning that additional object such as the scanner bed were visible, but also that more image slices were available for CT than for MVCT images.



**Figure 5.1 Example result of scanning a patient with a) an MVCT scanner b) a CT scanner. Values in the colour bars are in Hounsfield Units. We can observe that the range of MVCT values is smaller than that of CT values, due to the intrinsic properties of the MVCT pipeline.**

### 5.2.3 Methods

#### Data pre-processing

Although minimal data pre-processing was desirable, in order to improve the training of the generative adversarial network in use, it was necessary to crop the images so that parts of the scanner were not in the FOV. There was no attempt made to register the resulting images since the purpose was to develop a generalisable method, and registration often relies on key features of the acquisition modalities to be implemented.

#### Implementation of a generative adversarial network

A generative adversarial network (GAN)<sup>150</sup> was chosen as the best way to proceed, as they have had several successes in producing synthetic data in varied scenarios<sup>151,152,153</sup>. Such networks work by training two models simultaneously. A *generative* model  $G$  is used to capture information from the target data distribution, and a *discriminator* model  $D$  is used to estimate the probability that a sample came from the training data rather than  $G$ . It has been shown that the result of such a network is a unique solution in the space of the arbitrary functions  $G$  and  $D$ , where  $G$  recovers the distribution of the training data and  $D$  is equal to a half everywhere. This is equivalent to solving the following min-max problem with value function  $V(G, D)$ <sup>154</sup>

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log(D(x))] + E_{z \sim p_z(z)} [f(D, G)] \quad (5.1)$$

where  $p_{data}(x)$  represents the generator's distribution over data  $x$ ,  $p_z(z)$  is a prior on input 'noise' variables,  $E[\cdot]$  is the expectation value and  $f(D, G)$  is the function that  $G$  is being trained to minimize<sup>155</sup>.

It was thought that a cycleGAN<sup>156</sup> program could be used to analyse CT images to produce MVCT images and vice versa. This is a result of several specific characteristics of such a network:

- CycleGANs were specifically developed to train on unpaired images, so require little pre-processing of the data
- CycleGANs train two GANs
  - One for the conversion from dataset  $A$  to dataset  $B$ , with generator  $G_1$
  - One for the conversion from dataset  $B$  to dataset  $A$ , with generator  $G_2$
- Discriminator models are then used to determine how plausible the generated images are and to update them accordingly
- Importantly, such networks enforce the idea of *cycle consistency*, that attempts to enforce  $G_1(G_2(A)) \approx A$  and vice versa

The structure of the generators and discriminators used may be seen in Table 5.3, along with the overall structure of the network<sup>157,158,159</sup>.

**Table 5.3 CycleGAN steps. A Generator and Discriminator are run on both the MVCT and CT data (where original image in the table can be either).**

Pre-processing	Cropping of CT images	
	Mild rotation for augmentation	2° in any direction
	Normalisation of intensity values	Between 0 and 1
Generator	Encoder layer 1 ( $e_g1$ )	Input = original image Filters = 32
	Encoder layer 2 ( $e_g2$ )	Input = $e_g1$

	Filters = 64	
	Encoder layer 3 ( $e_g3$ )	Input = $e_g2$ Filters = 128
	Encoder layer 4 ( $e_g4$ )	Input = $e_g3$ Filters = 256
	Decoder layer 1 (d1)	Inputs = $e_g4$ & $e_g3$ Filters = 128
	Decoder layer 2 (d2)	Inputs = d1 & $e_g2$ Filters = 64
	Decoder layer 3 (d3)	Inputs = d2 & $e_g1$ Filters = 32
	Output ( $G_{out}$ )	Input = d3
Discriminator	Encoder layer 1 ( $e_d1$ )	Input = original image or $G_{out}$ Filters = 32
	Encoder layer 2 ( $e_d2$ )	Input = $e_d1$ Filters = 64
	Encoder layer 3 ( $e_d3$ )	Input = $e_d2$ Filters = 128
	Encoder layer 4 ( $e_d4$ )	Input = $e_d3$ Filters = 256
	Output ( $D_{out}$ )	Input = $e_d4$
Cycle consistency	Test $G_1(G_2(A)) \approx A$	Input = $G_{out}$ and $D_{out}$
	Test $G_2(G_1(B)) \approx B$	

The network also included options to train on ‘patches’ of the images for speed purposes, as convolutions of full images take longer to compute. In the same vein, the convolution ‘stride’ could be changed to convolve only every  $n^{\text{th}}$  row/column.

One additional complexity of GANs is that there are many possible optimizers that can be used for searching the problem space<sup>160</sup>, loss functions for assessing improvements<sup>161</sup>, metrics for quantifying overall performance<sup>162</sup>, amongst other parameters. This fact, combined with the time it can take to run the convolutions over medical images, result in an exponentially complex (due to the many possible parameters) and large (due to the range these parameters can take) search space. As such, in order to trial the network, it was deemed necessary to use parameters from an established network to build upon<sup>163</sup>.

### Assessment of output

Due to the case-by-case manner in which cycleGANs have been applied thus far, no clear guidelines exist with regards to the assessment of their performance. Often, such networks are compared to previous iterations on similar datasets and improvements are considered qualitatively. Initial assessment of output images in this study was therefore also done visually, through inspection of qualitative properties such as image sharpness and contrast. The progression of network loss values to assess when the network stopped progressing was analysed. One advantage of the dataset under study was that clear, quantitative values were being assessed for both generator scenarios. This meant that it was possible to establish quantitative metrics of performance. This was performed by comparing images which have gone through both generators to the original images. Such an assessment is valuable as it provides insight into the information lost during the transformations. We assume that the generated density values after both transforms could be modelled as:

$$f_{i,r}(\theta) = I_{i,r} + n_{i,r}(\theta) \quad (5.2)$$

where  $I_{i,r}$  is an unprocessed image i.e. the original data,  $f_{i,r}(\theta)$  was the generated value at pixel  $i$  for patient  $r$  and cycleGAN parameter space  $\theta$  and  $n_{i,r,\theta}$  was the associated noise term. Dataset ensemble characteristics were also compared. Established metrics<sup>164,165</sup> could then be used to assess noise characteristics and signal retrieval of the original and generated images.

### Noise properties in the background

Four regions of interest (ROIs) were defined, of  $10 \times 10 \times 3$  voxels, one in each corner of the sampled images. The mean of the pixels in the  $k^{\text{th}}$  ROI is then:

$$m_{r,k}(\theta) = \frac{1}{P} \sum_{i \in ROI_k} f_{i,r}(\theta) \quad (5.3)$$

Where  $P$  is the number of pixels in the ROI. We can then define the following noise metrics:

- *Image roughness (IR)* which measures pixel to pixel variability in the image (i.e. the noise that an individual might perceive when viewing an individual image).

$$IR_{r,k}(\theta) = \frac{\left( \sqrt{\frac{1}{P-1} \sum_{i \in ROI_k} (f_{i,r}(\theta) - m_{r,k}(\theta))^2} \right)}{m_{r,k}(\theta)} \quad (5.4)$$

- *Ensemble noise (EN)* which measures the noise across the data space and allows us to determine whether the noise characteristics of the generated image space resemble the noise characteristics of the original data space.

$$EN_k(\theta) = \frac{\left( \sqrt{\frac{1}{R-1} \sum_{r=1}^R (m_{r,k}(\theta) - \overline{m_k(\theta)})^2} \right)}{\overline{m_k(\theta)}} \quad (5.5)$$

where  $R$  represents the total number of patients and  $\overline{m_k(\theta)}$  represents the average of ROI means over the number of patients. Both of these metrics are variants of the coefficient of variations (COV), a commonly used, normalized, measure of dispersion that is the ratio of the standard deviation over the mean. Insights could also be gleaned from the datasets without the normalization term.

Image roughness, as defined in equation (4), is calculated on a per image and ROI basis. An average over the patient dataset can be obtained as:

$$IR(\theta) = \frac{1}{K \times R} \sum_{r=1}^R \sum_{k=1}^K IR_{r,k}(\theta) \quad (5.6)$$

Similarly, ensemble noise is still defined over ROIs. In order to obtain a statistic representative of the dataset, averaging over the ROIs gives:

$$EN(\theta) = \frac{1}{K} \sum_{k=1}^K EN_k(\theta) \quad (5.7)$$

## Signal properties

Due to the varying nature of the data, measuring signal properties was considered more challenging as there was no consistent feature in the images which could be used reliably. Thus, the maximum pixel value within an original patient scan can be found as:

$$h_{r,0} = \max(A_{i,r}) \in k_{patient} \quad (5.8)$$

where  $k_{patient}$  represents all pixels within the patient topology. The corresponding pixel in the generated dataset (the pixel at the exact same location) can then be labelled  $h_r(\theta)$ . Using this value, a *contrast recovery of maximum value*  $CRC_r^{max}$  can be defined as:

$$CRC_r^{max} = \frac{h_r(\theta)}{h_{r,0}} \quad (5.9)$$

In order to determine whether the images produced going through a full cycle of the network are accurate in reconstructing the original densities, the value of  $CRC_r^{max}$  should lie as close to one as possible.

Calculating the signal properties across the ensemble of all patient images real and generated is a more complex problem, as this conventionally requires a known, consistent signal source in all images at an identifiable location for comparative purposes. Such a signal source is not present for these images. Due to this, rather than a summary statistic we propose that comparing signal properties between the real and generated datasets could be done by comparing box plots of the  $CRC_r^{max}$  distribution across each.

## Other network properties

In order to visualise network progression and track events that might cause the network to not train optimally, code to visualise the progression of the loss functions for  $G_1$ ,  $G_2$  and adversarial loss over each epoch was implemented.

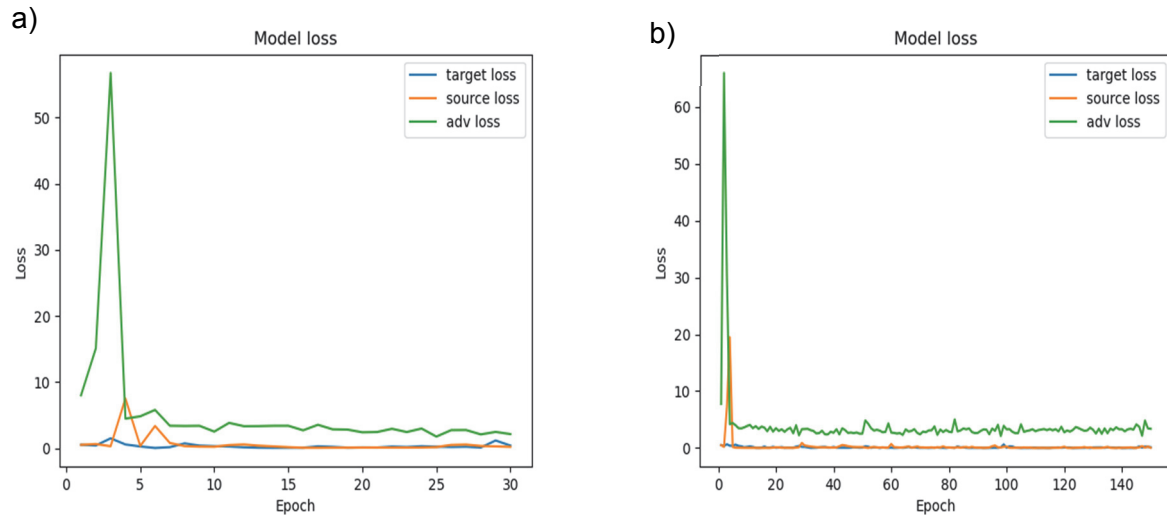
### 5.2.4 Results

## Network artefacts

The following artefacts can occur with this network:

### a) Mode collapse

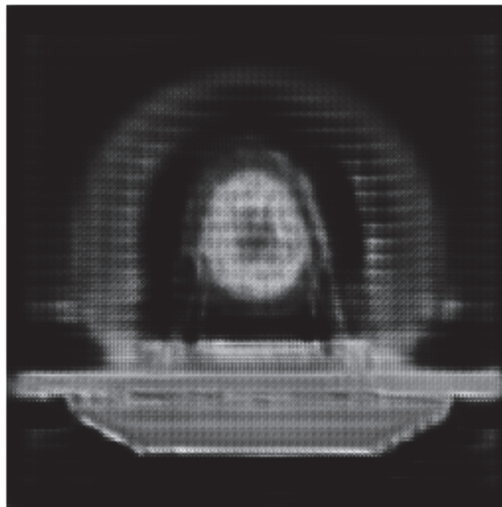
Mode collapse is an artefact that happens when the generator performs poorly in a manner which allows the discriminator to consistently determine which images are fake<sup>166</sup>. This results in all loss functions taking on consistently low values early on in training. It was frequently observed during the setting up and trialling of the network and served as an indicator of when to stop time-consuming computations, with examples shown in Figure .



**Figure 5.2 Examples of mode collapse for networks trained with varying parameters  $\theta$ .**

b) Erroneous assignment

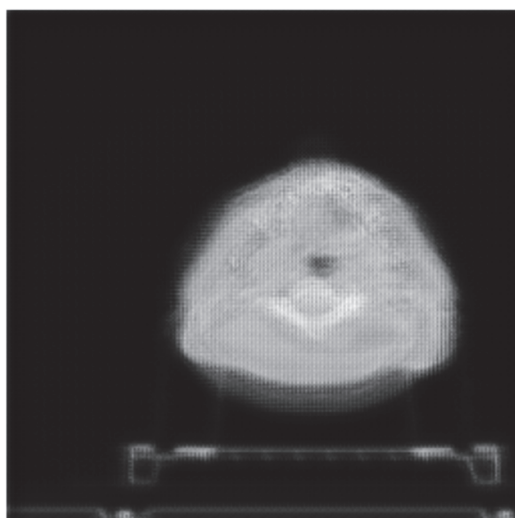
When multiple objects of roughly similar geometry exist in the problem space a GAN may mis-assign the attributes of one to the other. The artefact seen in Figure 5. is an example of this, where the network has mistakenly assigned the characteristics of the head to the scanner, thus emphasising the wrong features.



**Figure 5.3 Example of an erroneous assignment. In this image, the network has incorrectly assumed that the scanner, visible around the patient, was the target. This resulted in the network emphasizing the scanner features and faint 'ears' being created.**

c) Checkerboard artefact

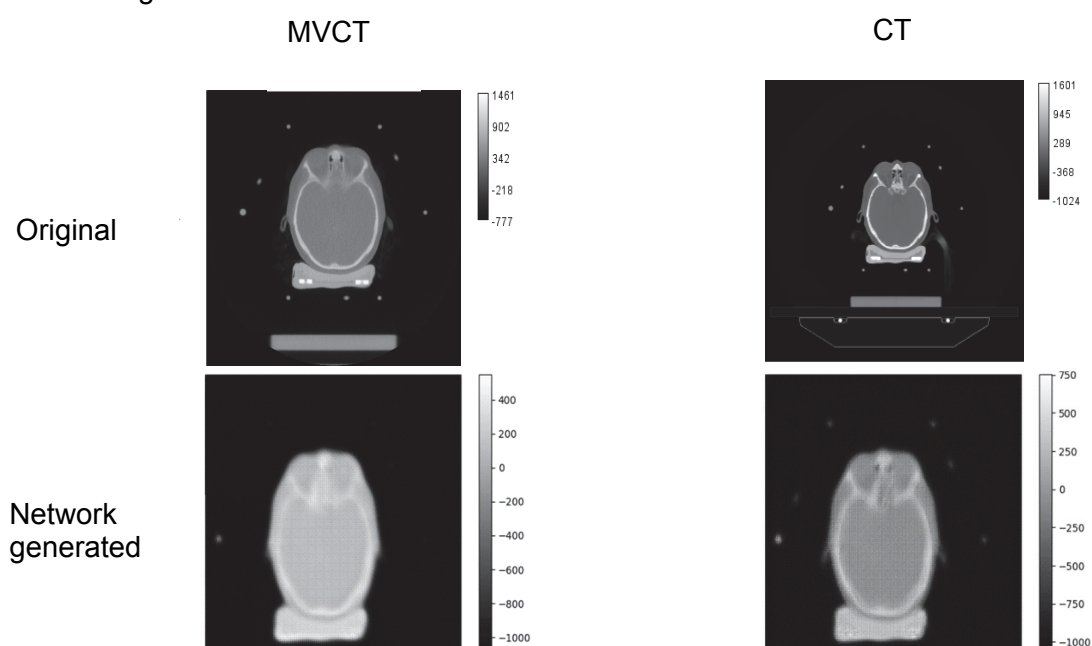
Due to the convolutional nature of GANs, checkerboard artefact readily occurs<sup>167</sup>. An example is seen in Figure . This can be due to overlaps in filter kernels from image size parameters, convolutional stride or simply not training a network long enough<sup>168</sup>. Several methods have been proposed for their removal<sup>169</sup>, however, these can lead to high frequency artefacts in the image instead. Due to the likelihood of the latter being confused for underlying health issues the former issue was considered a necessary compromise.



**Figure 5.4** Example of the checkerboard artefact. Striations are clearly visible in the image.

### Current results

Currently, the generators appear to train well enough to reproduce the underlying geometry of images (and even to remove non relevant information surrounding the patient), however, they still suffer from the checkerboard artefact. They also undershoot the maximal values in the original datasets by ~800 HU, leading to a lack of contrast between different tissues, as shown in Figure 5.5.



**Figure 5.5** Example of resulting images, columns and rows are as labelled. The network generated images are the result of a single generator step.



### 5.2.5 Discussion

This aim of this study was to develop an algorithm that could produce synthetic data in a manner that was as generalisable as possible, and to minimise pre-processing requirements. Methods of assessing the synthetic data generated were also to be investigated. A dataset that consisted of patients scanned on two different modalities, but which measured an identical physical characteristic was the original dataset of choice. A cycleGAN approach was chosen for this as it lent itself naturally to the problem at hand, and have been used for transforming medical images in the literature (such as for conversion from T1-weighted magnetic resonance images to functional magnetic resonance images)<sup>170</sup>. Additionally, the metrics proposed here are useful as long as there is a quantitative physical underlying value being measured. However, in the case of certain modalities where a more qualitative assessment is usual, such as MRI, these metrics may no longer hold. For qualitative images, a potential way of overcoming arbitrary biases might be to normalise them to have a mean of zero and a standard deviation of one before analysis. This could potentially remove acquisitions effects from subsequent analysis.

A key point to note is that individual image quality metrics proposed here only quantitatively assess results over a full cycle i.e.  $G_1(G_2(A))$  or  $G_2(G_1(B))$ , due to the difficulties of comparing images of acquired or generated from different scanner types. This difficulty comes from the fact that a quality metric will give different values for different scanner types due to differences in the physics of the data acquisition. This means that only qualitative observations when assessing an image that has been generated at a half cycle can be made. This issue is overcome by comparing ensemble or batch statistics, which use the distribution of the data as a quality indicator. A point to note is that the  $CRC_r^{max}$  might be particularly sensitive to noise. In order to combat this, one could potentially take the volume (e.g. a 1 ml sphere) with the maximum value, which would reduce this effect. This would be analogous to the  $SUV_{peak}$  used in positron emission tomography<sup>171</sup>. The Fréchet Inception Distance (FID)<sup>172</sup> was recommended as a method of calculating the difference between generated batch results and the original dataset. This has not been implemented but could be of value in future research.

This study has been highlighted some of the foibles of CycleGANs. Such networks can learn and produce realistic images rapidly and without excessive pre-processing of the data (e.g. Figure 5.5, a creation of the network based upon the examples it was fed). They require no atlas, no registration, no segmentation and hardly any interaction before being trained, as long as the dimensions of the objects being compared in the images are roughly comparable (e.g. a rule of thumb is that the network might struggle if an object is twice the size in pixels in one of the images, though this is network dependant and worse performances are recorded<sup>173</sup>). This can save a lot of time for setting the process up. The lack of clear tissue assignments through segmentation also means that GANs can achieve a higher granularity of detail that qualitatively appears more representative than arbitrarily defining several tissue classes and ignoring the fact that in reality they mix.

However, such networks are not without their drawbacks. They require large datasets for training. The exponential parameter space makes it difficult to identify which functions will improve network performance. It is only recently that the details of why certain methods perform better than others are being delved into<sup>174</sup>. This lack of understanding means that ML is often seen as a black-box method at best, little more than alchemy at worst. The ability to quantify accuracy of generated images, rather than of classification, should be pursued if synthetic images are to be used in the clinic. An additional important drawback is the need for high-performance computing resources. Running a cycleGAN can take days on a regular PC before an assessment can be made as to whether it is worth stopping training in order to change the parameter space. A temporal debt can thus be accrued quite quickly if changes

do not work out as planned. Even if they do, it can be unclear when to stop training. As such it is good practice to keep track of the networks being trained at regular intervals to ensure smooth operation.

In this case study the generation of synthetic data for clinical purposes has been investigated. To start various methods of producing synthetic data were discussed, before settling on a ML approach due to the promising results these methods currently produce in the literature. These methods appear to be the most general purpose, as they often require the least labelling and are able to provide more granular predictions. As shown above, this comes from the flexibility of the parameter space which is explored by such methods. In order to accurately determine the best set of parameters  $\theta$ , an exhaustive search of such space needs to be performed. It would be of interest to identify specifically how each parameter affects the final images output, and how interactions between variations of such parameters affect the result.

A potential drawback of using ML methods to produce synthetic data is that it is possible to create data which appear convincing, but which are inaccurate in key aspects. This means that a careful assessment is necessary of the situations in which synthetic data may be of use is necessary. The tool created above should thus be used for specific clinical problems of interest<sup>175</sup>, once trained on a dataset exhibiting similar characteristics, and so should be assessed on a case by case basis. Additionally, the tool cannot create information where there is none, but it will still produce an output. Therefore, care must be taken to assess whether there is a potential for information to be extracted or not. A more general difficulty is the fact that for appropriate training of CNNs and other AI-style algorithms a suitable dataset is required<sup>176</sup>. Finding such a dataset can be an issue. However, the advantage of GANs here is that minimal pre-processing of the data will be required once such a dataset has been identified.

### 5.3 CASE STUDY OUTCOMES/HIGHLIGHTS

The main highlight of this case study is the implementation of a network capable of dealing with both dynamic and static images, which can be trained on a small sample to produce reasonable outputs. This code has been implemented both on a regular PC and on HPC facilities. Quality metrics of use for assessing the accuracy of the network in preserving key image features and batch characteristics have been identified.

This work has been presented at several of the LS&H PEG meetings throughout the duration of the project, as well at a STAC meeting where it received a certain amount of interest from attendees who had had to undergo several scans and could attest to the unpleasantness of the procedure.

### 5.4 FUTURE DIRECTION OF CASE STUDY

The metrics proposed here for assessing the value of individual output synthetic data rely on the network having a cyclical structure. The one-way quality issue is somewhat more nuanced, as there is no reference value to compare against. It is for this reason that suggest batch statistics as a proxy for data quality are suggested. Although these lose the granularity necessary to make image-based insights, they provide valuable information about the overall performance. Further work could be done in this area by investigating other batch statistics such as the previously mentioned Fréchet Inception Distance.

The CycleGAN implemented could also do with further improving. Reasons for the undershoot of intensity values need to be established and corrected, and a systematic search of the parameter space carried out to detect which cycleGAN factors most affect the

quality of the output. Once such a search has been carried out and the factors established, it will be possible to further investigate why these factors cause the increase or decrease in the quality of the output. With a more thorough understanding of this it should become possible to provide guidelines for parameter selection, which are sorely lacking in the field. These could then be used to adapt the network for other uses such as a PET to SPECT transform and vice versa.

## 6 CASE STUDY 5: MENTAL HEALTH AND COVID-19

**Full title:** Study into the Covid-19 crisis using primary care mental health consultations and prescriptions data

**Technical lead:** Agnieszka Lemanska

**Web:** <https://www.npl.co.uk/research/data-science/digital-health/case-study/mental-health>

### 6.1 CASE STUDY PROPOSAL

#### 6.1.1 Background

In September 2020, Public Health England (PHE) published a report that presented real time intelligence on the mental health and wellbeing of the population in England during the Covid-19 pandemic<sup>177</sup>. It compiles routinely updated indicators from multiple sources and summarises important findings on the health of the population and on the resilience of NHS services. The report was needed to inform policy, planning and commissioning in mental health and social care and to assist stakeholders at both national and local level on the ability to respond to a changing environment where decisions are needed to be made quickly. The report concluded that certainty in existing evidence is limited, and that this is posing a barrier to decision-making<sup>178</sup>. This research contributes to the emerging new research that is aimed at informing future policy decisions.

#### 6.1.2 Short project summary

The effect of the 2020 pandemic and of the national measures introduced to control it is not yet fully understood. The aim of this study was to investigate how different types of primary care data can help quantify the effect of the coronavirus disease (Covid-19) crisis on mental health. A retrospective cohort study investigated changes in weekly counts of mental health consultations and prescriptions. The data were extracted from one of the UK's largest primary care databases between January 1<sup>st</sup> 2015 and October 31<sup>st</sup> 2020 (end of follow-up). The 2020 trends were compared to the 2015-19 average with 95% confidence intervals using longitudinal plots and analysis of covariance (ANCOVA).

A total of 504 practices (7,057,447 patients) contributed data. During the period of national restrictions there were 31% ( $3957 \pm 269$ ,  $p < 0.001$ ) fewer consultations and 6% ( $4878 \pm 1108$ ,  $p < 0.001$ ) more prescriptions per week compared to the 2015-19 average. By the end of the study, the number of events was recovering, increasing by 75 ( $\pm 29$ ,  $p = 0.012$ ) per week. Prescriptions had returned to the 2015-19 levels by the end of the study ( $p = 0.854$ ). The significant reduction in the number of consultations represents part of the crisis. Future service planning and quality improvements are needed to reduce the negative effect on health and healthcare.

### 6.2 CASE STUDY TECHNICAL WORK

#### 6.2.1 Introduction

The Covid-19 pandemic, and the national and worldwide efforts to contain and manage the infections, have challenged all facets of life and placed an unprecedented strain on people, patients and healthcare systems<sup>179,180,181,182,183</sup>. Continued social distancing and isolation

measures, coupled with limited access to healthcare and face-to-face consultations are challenging for everyone, but people with mental health conditions are especially vulnerable.

In mid-March 2020 (week 11), the UK government introduced national restrictions. The aim was to control the rate of infections, prevent an overburden on the National Health Service (NHS), and reduce the number of deaths associated with the Covid-19. However, the emerging research indicates that the pandemic and the lockdown affected the mental health of the population. For example, increased anxiety and depression in adults experiencing isolation has been reported<sup>184</sup>.

In this work, the 2020 trends in primary care events (consultations) and prescriptions data for mental health conditions such as anxiety, depression, and sleep disorders were compared to the 2015-19 average to understand the effect of the pandemic and lockdown on mental health and on access to healthcare. This is important because as restrictions ease it is necessary to plan future workloads and consider how to improve healthcare services provided to people in the UK and worldwide during pandemics. This work has applications beyond mental health, as the emerging research indicates that both mental and physical health are affected, for example by delayed diagnosis of long-term conditions such as cancer or diabetes<sup>185</sup>.

## 6.2.2 Methods

### Study design and data extraction

This was a population-based, longitudinal retrospective cohort study using the Oxford Royal College of General Practitioners (RCGP) Clinical Informatics Digital Hub (ORCHID) database. ORCHID extracts electronic healthcare records (EHRs) from the RCGP Research and Surveillance Centre (RSC) network of primary care practices. ORCHID is representative of the English population<sup>186</sup> and comprises over 500 practices with nearly 10 million patient records (>10% of the English population).

Two types of data were used: 1) consultations and 2) prescriptions data. Events were extracted from EHRs using clinical codes in the SNOMED-CT system. They included consultations, diagnosis and symptoms for common mental health conditions such as anxiety, depression and sleep disorders. The prescriptions data were extracted with lists of medications for pharmacological groups such as antidepressants, anxiolytics and hypnotics (sleep medications). These lists were generated based on the British National Formulary (BNF78, 2019-2020, bnf.org). The lists of SNOMED-CT codes and medications can be obtained from the corresponding author. The study was from January 1<sup>st</sup> 2015 to October 31<sup>st</sup> 2020, when the follow-up ended.

### Statistical analysis

Weekly counts of events and prescriptions were calculated and plotted for the year 2020 against the 2015 to 2019 average with 95% confidence intervals. Analysis of covariance (ANCOVA) was used to estimate the difference in the weekly counts between 2020 and the 2015-19 average for the period of week 12 (beginning of lockdown) to week 43 (end of follow-up) and its statistical significance adjusted for the weekly trend. The binary grouping variable for ANCOVA was year 2020 vs year 2015-19 average. The week variable was a covariate. Database extraction was in Structured Query Language (SQL) Server Management Studio version 18.3.1 and statistical analyses in R version 3.5.1.

## Approvals

This study was approved by the RCGP board (data request number RSC\_1620). Ethical approval was not required. Patients consented to the research on an opt-out basis.

### 6.2.3 Results

#### The cohort, events and prescription counts

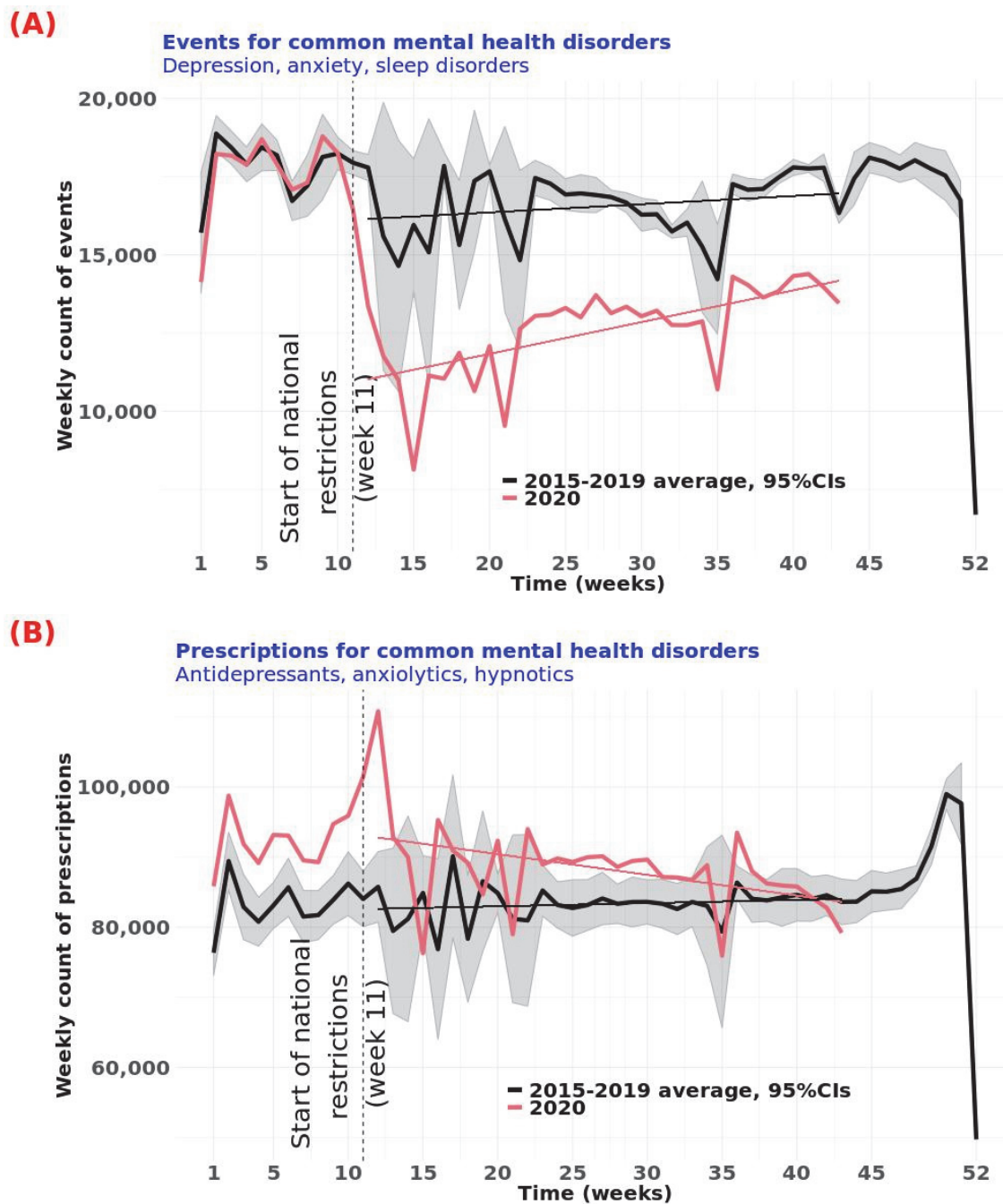
In total, 504 practices and 7,057,447 participants contributed data (over 5 million participants each year due to registering and de-registering). There were over 5 million events recorded for 1,205,825 (17%) participants and nearly 26 million prescriptions for 1,459,223 (21%) participants. Table 6.1 presents the yearly split and rates per 100,000 patients.

**Table 6.1 Weekly counts of events (primarily consultations) and prescriptions**

	Participants	Events	Rate per 100,000	Prescriptions	Rate per 100,000
Total, 2015- 2020	7,057,447	5,000,003		25,755,647	
2015	5,080,326	877,569 (18%)	17,274	4,142,322 (16%)	81,537
2016	5,112,815	888,096 (18%)	17,370	4,323,328 (17%)	84,559
2017	5,173,919	883,931 (18%)	17,084	4,390,547 (17%)	84,859
2018	5,250,640	864,950 (17%)	16,473	4,414,077 (17%)	84,067
2019	5,345,376	862,485 (17%)	16,135	4,526,997 (18%)	84,690
2020 (Jan – Oct)	5,202,588	622,972 (12%)		3,958,376 (15%)	
Whole 2020 (estimate)		747,566 (15%)	14,369	4,750,051 (18%)	91,301

#### Trends in 2020 compared to previous years (2015 to 2019)

From week 11 onwards, when the national restrictions started, the 2020 count of events dropped sharply below the 2015-19 average (Figure 6.1A). Although the numbers were gradually recovering, events remained below until the end of the study. The weekly count of prescriptions (Figure 6.1B) followed a different trend. It was above the 2015-19 average before the lockdown. We observed a peak increase in week 12 when the lockdown started and then the numbers returned to the 2015-19 values.



**Figure 6.1 Trends in weekly counts of (A) events and (B) medications for common mental health disorders**

### The ANCOVA analysis

In both ANCOVA models the grouping variable (Year), the covariate (Week), and the interaction term were determined to be significant (Table 6.2). We therefore report that in the period of week 12 to 43 of 2020 there were on average 12,599 events per week which was 31% ( $3,957 \pm 269$ ,  $p < 0.001$ ) less than the 2015-19 average for the same period. The 2020 weekly counts were slowly recovering towards the 2015-19 average, increasing by around 1% ( $75 \pm 29$ ,  $p = 0.012$ ) events each week. Prescriptions followed a different trend. There were on average 6% ( $4,878 \pm 1108$ ,  $p < 0.001$ ) more prescriptions per week in weeks 12 to 43 of 2020 than 2015-19. The 2020 prescription counts returned to the 2015-19 levels with no statistically significant difference in week 43 ( $p = 0.854$ ).

Immediately following the start of lockdown (week 12), a significant drop by 46% ( $5,123 \pm 526$ ,  $p < 0.001$ ) in consultations was observed and a peak increase in prescriptions with 11% ( $10,154 \pm 2,164$ ,  $p < 0.001$ ) more prescriptions issued that week than the 2015-19 average.

**Table 6.2 ANCOVA analysis of weekly counts of (A) events and (B) prescriptions**

(A) Events	Covariate value	Coefficient	Standard Error	p-value
Year 2020	Week 12 (start)	11,024	372	<0.001
2015-19 average		5,123	526	<0.001
Year 2020	Week 27.5 (mean)	12,599	190	<0.001
2015-19 average		3,957	269	<0.001
Year 2020	Week 43 (end of	14,173	372	<0.001
2015-19 average	follow-up)	2,790	526	<0.001
Week		102	21	<0.001
Week * 2015-19 average		-75	29	0.012

(B) Prescriptions	Covariate value	Coefficient	Standard Error	p-value
Year 2020	Week 12 (start)	92,784	1530	<0.001
2015-19 average		-10,154	2,164	<0.001
Year 2020	Week 27.5 (mean)	88,208	783	<0.001
2015-19 average		-4,878	1108	<0.001
Year 2020	Week 43 (end of	83,633	1530	<0.001
2015-19 average	follow-up)	399	2,164	0.854
Week		-295	85	<0.001
Week * 2015-19 average		340	120	0.006

## 6.2.4 Discussion

By using two different types of data, we were able to investigate the impact of the coronavirus crisis on primary care. A significant drop in consultations was observed following lockdown. Prescriptions followed an opposite trend with more prescriptions issued during the national restrictions period than in the years from 2015 to 2019. These findings and the learning from this project are important. As we recover from the pandemic, we need to review the strengths and weaknesses of the healthcare services and put crisis-related plans in place for future national and global emergencies.

This research adds to the emerging evidence that the Covid-19 pandemic, and the measures to control it, may have indirectly affected mental health<sup>187</sup> as well as physical health<sup>188,189</sup> in the population and the quality, quantity and type of services available to people. Although a drop in consultations might be expected in a lockdown when people are self-isolating and shielding, there is the potential that they may also neglect to seek help. The increase in prescriptions indicates the reliance on medications (potentially without sufficient follow-ups), rather than considering other therapeutic options.

It is important to note that the results indicate a drop in consultations but not in rates of mental health disorders. This therefore supports the need for different mechanisms of



delivering consultations, including remote appointments via phone or internet. Research is needed to investigate the effectiveness and availability of remote consultations.

With the use of two different types of primary care data, we were able to improve the understanding of the strengths and limitations of healthcare services in the UK. This learning can help improve the resilience of healthcare systems for a future crisis.

### 6.3 CASE STUDY OUTCOMES/HIGHLIGHTS

- MIE2021 conference abstract (<https://mie2021.org/en/>) 29-31<sup>st</sup> May 2021.
- A joint publication with PHE on the effect of the lockdown on mental health and services.
- A publication on the effect of the pandemic and the measures to contain it on mental health and primary care services.

### 6.4 FUTURE DIRECTION OF CASE STUDY

So far, the trends in consultation and prescribing data during 2020 as compared to previous years have been analysed. This analysis allows us to establish the effect of the pandemic and lockdown on mental health of the population and to study the response and resilience of healthcare services.

As future work, we will undertake analysis on an individual patient level. We will study characteristics and risk factors for mental health decline during the pandemic including obesity, comorbidities, and socioeconomic status. We will also investigate the effect of Covid-19 on mental health and whether people who were ill report worsening of mental health. There is emerging research that points to Covid-19 as a neurodegenerative disease and shows mental health issues among the chronic Covid-19 symptoms (long-Covid)<sup>190,191,192,193</sup>.

We will also undertake a data-driven approach to analyse the factors considered in this report and to potentially identify new trends in data, and compare the findings to the standard statistical approach as presented in this report. This will allow comparison and confirmation of findings and will serve as a proof of concept for the data driven approach using this dataset.

## 7 MAIN OUTPUTS AND LEARNINGS

### 7.1 LIST OF PEER-REVIEWED PUBLICATIONS

- Analysis of primary care Computerized Medical Records (CMR) data with deep autoencoders (DAE). *Front. Appl. Math. Stat.*; 2019; 5. Thomas SA, Smith N, Livina V, Yonova I, Webb R, de Lusignan S.
- Building confidence in digital health through metrology, *British Journal of Radiology*, 2020; 93. N. Smith, D. Sinden, S. Thomas, M. Romanchikova, J. Talbott, M. Adeogun.
- Linkage of the CHHiP randomised controlled trial with primary care data: a study investigating ways of supplementing cancer trials and improving evidence-based practice, *BMC Medical Research Methodology*; 2020; 20. Agnieszka Lemanska, Rachel C. Byford, Clare Cruickshank, David P. Dearnaley, Filipa Ferreira, Clare Griffin, Emma Hall, William Hinton, Simon de Lusignan, Julian Sherlock, Sara Faithfull.
- Primary care prostate cancer case ascertainment, *Digital Personalized Health and Medicine*; 2020; 270. Agnieszka Lemanska, Sara Faithfull, Harshana Liyanage, Sophie Otter, Marina Romanchikova, Julian Sherlock, Nadia A.S. Smith, Spencer A. Thomas, Simon de Lusignan.
- Study into Covid-19 crisis using primary care mental health consultations and prescriptions data. *Accepted for publication in "Studies in Health Technology and Informatics"* 2021. Agnieszka Lemanska, Uy Hoang, Nathan Jeffreys, Clare Bankhead, Kam Bhui, Filipa Ferreira, Sally Harcourt, Anthony James, Harshana Liyanage, Brian D Nicholson, Julian Sherlock, Gillian Smith, Nadia Smith, Spencer A Thomas, John Williams and Simon de Lusignan.
- Combining Image Features and Patient Metadata to Enhanced Transfer Learning, *Submitted to IEEE Engineering in Medicine and Biology Conference 2021*. Spencer A Thomas.

### 7.2 WHAT HAS BEEN ACHIEVED OVER THE THREE YEARS

The first two years went to plan, but the third-year budget was halved due to Covid-19 response<sup>194</sup> projects, so deliverables were slightly rescope and some of the original case studies were not continued as planned at the end of the second year.

However, the overall goals of the project were still achieved, namely:

- to build an area of expertise for NPL and increase NPL's profile as an active NMI working in the area of digital health, evidenced by:
  - the website, network of contacts, peer-reviewed publications, conference presentations, and final report;
  - the team's readiness and capability to help with the Covid-19 response;
- to understand the role metrology can play in the curation and analysis of healthcare data
  - evidenced through work developed in the five case studies discussed in this report;

- the findings have helped shape future work under the NMS formulation programme, as well as other funding streams.

### 7.3 LEARNING POINTS FOR FUTURE WORK

Clinical data is difficult to access. It is of varying quality, often siloed within trusts and centres and as a result, not used to its full potential. There are many current initiatives working towards making better use of data in healthcare as shown in this report.

Large amounts of data are collected in the NHS, but a lack of standards in what attributes are included and how they are recorded, makes the utilisation of much of this data challenging. In addition, without standards for what and how to record data, combining multiple sources of data becomes extremely challenging due to incompatible formats, missing values, mismatch of measurement units and lack of metrological traceability.

Mining such heterogeneous data requires extensive 'data cleaning' to remove errors, input missing data or convert it into compatible types. For large datasets, such as population data, this becomes a daunting and impractical task. Furthermore, data mining and cleansing tools have to be re-developed every time any of the source data formats changes, adding to the strain on the clinical resources and increasing the likelihood of errors. Standardisation of data and metadata formats ensure that the data are collected and used in the best possible way, greatly increasing the 'data half-life' and long-term value.

Through on-going collaborations and discussions with stakeholders held during the duration of this project, the following needs have been identified, highlighted by the work performed:

- improved data and metadata capture and management systems;
- traceable and standardised software and data analysis toolkits, and for reference methods and datasets to validate them;
- new data standards and digital capabilities in healthcare in the UK;
- framework to validate and interpret AI methods deployed in the clinic;
- training for end-users in the healthcare workforce on how to evaluate and use data and new technologies.

NPL can play its part in helping to address these needs. Key to this is building a good reputation and trust in this area to be able to lead by example. It is of extreme importance to continue to work closely with healthcare partners, and to continue to build the networks in the clinical, academic and industrial domain.

## **8 ACKNOWLEDGMENTS**

This project was funded by the department of Business, Engineering and Industrial Strategy (BEIS) through the National Measurement Strategy under the Life Sciences & Healthcare theme (Digital Health – Data Curation) at the National Physical Laboratory (NPL) in the United Kingdom.

The authors would like to thank David Sinden, Jessica Talbott and Maria Theodorou for their contributions to the project at different stages of its lifetime. The authors would like to thank Matt Hall for doing an independent technical review of this report.

The authors would like to thank their collaborators for allowing access to their data: the ORCHID-RCGP team and the Cambridge University Hospitals team that own the VoxTox data.

This work uses data provided by patients and collected by the NHS as part of their care and support.

## 9 REFERENCES

- <sup>1</sup> Nadia A. S. Smith et al., “Building Confidence in Digital Health through Metrology,” *The British Journal of Radiology* 93, no. 1109 (January 28, 2020): 20190574, <https://doi.org/10.1259/bjr.20190574>.
- <sup>2</sup> Mark D. Wilkinson et al., “The FAIR Guiding Principles for Scientific Data Management and Stewardship,” *Scientific Data* 3, no. 1 (March 15, 2016): 160018, <https://doi.org/10.1038/sdata.2016.18>.
- <sup>3</sup> Mark D. Wilkinson et al., “Addendum: The FAIR Guiding Principles for Scientific Data Management and Stewardship,” *Scientific Data* 6, no. 1 (March 19, 2019): 6, <https://doi.org/10.1038/s41597-019-0009-6>.
- <sup>4</sup> “Metrology Is Key to Reproducing Results: Nature News & Comment,” accessed March 2, 2021, <https://www.nature.com/news/metrology-is-key-to-reproducing-results-1.22348>.
- <sup>5</sup> Anne L. Plant et al., “How Measurement Science Can Improve Confidence in Research Results,” *PLoS Biology* 16, no. 4 (April 23, 2018), <https://doi.org/10.1371/journal.pbio.2004299>.
- <sup>6</sup> Leonard P. Freedman, Iain M. Cockburn, and Timothy S. Simcoe, “The Economics of Reproducibility in Preclinical Research,” *PLOS Biology* 13, no. 6 (June 9, 2015): e1002165, <https://doi.org/10.1371/journal.pbio.1002165>.
- <sup>7</sup> “JCGM 200: International vocabulary of metrology – Basic and general concepts and associated terms (VIM)”. Joint Committee for Guides in Metrology, 2012.
- <sup>8</sup> “JCGM 100: Evaluation of Measurement Data - Guide to the Expression of Uncertainty in Measurement”. Joint Committee for Guides in Metrology, 2008.
- <sup>9</sup> Nicholas I. Cole et al., “An Ontological Approach to Identifying Cases of Chronic Kidney Disease from Routine Primary Care Data: A Cross-Sectional Study,” *BMC Nephrology* 19, no. 1 (April 10, 2018): 85, <https://doi.org/10.1186/s12882-018-0882-9>.
- <sup>10</sup> Mark D. Wilkinson et al., “Addendum: The FAIR Guiding Principles for Scientific Data Management and Stewardship,” *Scientific Data* 6, no. 1 (March 19, 2019): 6, <https://doi.org/10.1038/s41597-019-0009-6>.
- <sup>11</sup> Chang Sun et al., “Analyzing Partitioned FAIR Health Data Responsibly,” n.d., 6.
- <sup>12</sup> “Overview - FHIR v4.0.1,” accessed March 2, 2021, <https://www.hl7.org/fhir/overview.html>.
- <sup>13</sup> “MNIST Handwritten Digit Database, Yann LeCun, Corinna Cortes and Chris Burges,” accessed February 23, 2021, <http://yann.lecun.com/exdb/mnist/>.
- <sup>14</sup> Babak Ehteshami Bejnordi et al., “Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer,” *JAMA* 318, no. 22 (December 12, 2017): 2199, <https://doi.org/10.1001/jama.2017.14585>.
- <sup>15</sup> “Home - Grand Challenge,” grand-challenge.org, accessed February 23, 2021, <https://camelyon16.grand-challenge.org/>.
- <sup>16</sup> “Camelyon Challenge - Are Today’s Algorithms Ready to Diagnose Cancer?,” microDimensions, accessed February 23, 2021, <https://micro-dimensions.com/blog/2016/7/28/camelyon-challenge-are-todays-algorithms-ready-to-diagnose-cancer>.
- <sup>17</sup> Geert Litjens et al., “1399 H&E-Stained Sentinel Lymph Node Sections of Breast Cancer Patients: The CAMELYON Dataset,” *GigaScience* 7, no. giy065 (June 1, 2018), <https://doi.org/10.1093/gigascience/giy065>.
- <sup>18</sup> “Home - Grand Challenge,” grand-challenge.org, accessed February 23, 2021, <https://camelyon17.grand-challenge.org/>.
- <sup>19</sup> “Excel: Why Using Microsoft’s Tool Caused Covid-19 Results to Be Lost,” *BBC News*, October 5, 2020, sec. Technology, <https://www.bbc.com/news/technology-54423988>.
- <sup>20</sup> David Oliver, *Making Our Health and Care Systems Fit for an Ageing Population* (London: The King’s Fund, 2014).
- <sup>21</sup> Raman Kumar and Ranabir Pal, “India Achieves WHO Recommended Doctor Population Ratio: A Call for Paradigm Shift in Public Health Discourse!,” *Journal of Family Medicine and Primary Care* 7, no. 5 (October 2018): 841–44, [https://doi.org/10.4103/jfmpc.jfmpc\\_218\\_18](https://doi.org/10.4103/jfmpc.jfmpc_218_18).
- <sup>22</sup> “Future of Patient Data,” Future Agenda, accessed March 1, 2021, <https://www.futureagenda.org/focus-on/the-future-of-patient-data/>.
- <sup>23</sup> Karim Abouelmehdi et al., “Big Data Security and Privacy in Healthcare: A Review,” *Procedia Computer Science*, The 8th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN 2017) / The 7th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH-2017) / Affiliated Workshops, 113 (January 1, 2017): 73–80, <https://doi.org/10.1016/j.procs.2017.08.292>.

- <sup>24</sup> Moshaddique Al Ameen, Jingwei Liu, and Kyungsup Kwak, "Security and Privacy Issues in Wireless Sensor Networks for Healthcare Applications," *Journal of Medical Systems* 36, no. 1 (February 1, 2012): 93–101, <https://doi.org/10.1007/s10916-010-9449-4>.
- <sup>25</sup> Yisroel Mirsky et al., "CT-GAN: Malicious Tampering of 3D Medical Imagery Using Deep Learning," *ArXiv:1901.03597 [Cs]*, June 6, 2019, <http://arxiv.org/abs/1901.03597>.
- <sup>26</sup> Jonathan M. Peake, Graham Kerr, and John P. Sullivan, "A Critical Review of Consumer Wearables, Mobile Applications, and Equipment for Providing Biofeedback, Monitoring Stress, and Sleep in Physically Active Populations," *Frontiers in Physiology* 9 (2018), <https://doi.org/10.3389/fphys.2018.00743>.
- <sup>27</sup> "The Topol Review — NHS Health Education England," The Topol Review — NHS Health Education England, accessed March 1, 2021, <https://topol.hee.nhs.uk/>.
- <sup>28</sup> "Nhs-Long-Term-Plan-Version-1.2.Pdf," accessed March 1, 2021, <https://www.longtermplan.nhs.uk/wp-content/uploads/2019/08/nhs-long-term-plan-version-1.2.pdf>.
- <sup>29</sup> "Home," NHSX, accessed March 1, 2021, <https://www.nhsx.nhs.uk/>.
- <sup>30</sup> "The Future of Healthcare: Our Vision for Digital, Data and Technology in Health and Care," GOV.UK, accessed March 1, 2021, <https://www.gov.uk/government/publications/the-future-of-healthcare-our-vision-for-digital-data-and-technology-in-health-and-care/the-future-of-healthcare-our-vision-for-digital-data-and-technology-in-health-and-care>.
- <sup>31</sup> "NHSX\_AI\_report.Pdf," accessed March 1, 2021, [https://www.nhsx.nhs.uk/media/documents/NHSX\\_AI\\_report.pdf](https://www.nhsx.nhs.uk/media/documents/NHSX_AI_report.pdf).
- <sup>32</sup> "The NHS AI Lab," NHSX, accessed March 1, 2021, <https://www.nhsx.nhs.uk/ai-lab/>.
- <sup>33</sup> "NHS Accelerated Access Collaborative," accessed March 1, 2021, <https://www.england.nhs.uk/aac/>.
- <sup>34</sup> "Home," NHS Digital, accessed March 1, 2021, <https://digital.nhs.uk/>.
- <sup>35</sup> "NHS England » Regional Teams," accessed March 1, 2021, <https://www.england.nhs.uk/about/regional-area-teams/>.
- <sup>36</sup> "Apo-Nid196221.Pdf," accessed March 1, 2021, <https://apo.org.au/sites/default/files/resource-files/2017-07/apo-nid196221.pdf>.
- <sup>37</sup> "£600K Wade-Gery Digital Review May Be Delayed," Digital Health, November 25, 2020, <https://www.digitalhealth.net/2020/11/600k-wade-gery-digital-review-may-be-delayed/>.
- <sup>38</sup> "An Alternative (Re)View of NHS Digital and NHSX," Digital Health, November 13, 2020, <https://www.digitalhealth.net/2020/11/an-alternative-review-of-nhs-digital-and-nhsx/>.
- <sup>39</sup> "£600K Wade-Gery Digital Review May Be Delayed," Digital Health, November 25, 2020, <https://www.digitalhealth.net/2020/11/600k-wade-gery-digital-review-may-be-delayed/>.
- <sup>40</sup> "Data and COVID-19," NHSX, accessed March 1, 2021, <https://www.nhsx.nhs.uk/covid-19-response/data-and-covid-19/>.
- <sup>41</sup> "NHS Digital Response to Coronavirus (COVID-19)," NHS Digital, accessed March 1, 2021, <https://digital.nhs.uk/coronavirus>.
- <sup>42</sup> "BETA - NHS Digital, Data and Technology Standards Framework," NHS Digital, accessed March 1, 2021, <https://digital.nhs.uk/about-nhs-digital/our-work/nhs-digital-data-and-technology-standards/framework>.
- <sup>43</sup> Alexander Mense and Bernd Blobel, "HL7 Standards and Components to Support Implementation of the European General Data Protection Regulation (GDPR)" 13, no. 1 (2017): 7.
- <sup>44</sup> "National Data Guardian," GOV.UK, accessed March 1, 2021, <https://www.gov.uk/government/organisations/national-data-guardian>.
- <sup>45</sup> "DCB0086: Data Security and Protection Toolkit," NHS Digital, accessed March 1, 2021, <https://digital.nhs.uk/data-and-information/information-standards/information-standards-and-data-collections-including-extractions/publications-and-notifications/standards-and-collections/dcb0086-data-security-and-protection-toolkit>.
- <sup>46</sup> "A Guide to Good Practice for Digital and Data-Driven Health Technologies," GOV.UK, accessed March 1, 2021, <https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology>.
- <sup>47</sup> "HDR-UK-One-Institute-Strategy-Compressed-1.Pdf," accessed March 1, 2021, <https://www.hdr.uk.ac.uk/wp-content/uploads/2019/04/HDR-UK-One-Institute-Strategy-compressed-1.pdf>.
- <sup>48</sup> "Home," HDR UK, accessed March 1, 2021, <https://www.hdr.uk.ac.uk/>.
- <sup>49</sup> "UKHDRA," accessed March 1, 2021, <https://ukhealthdata.org/>.

- <sup>50</sup> “Our Hubs,” *HDR UK* (blog), accessed March 1, 2021, <https://www.hdruk.ac.uk/helping-with-health-data/our-hubs-across-the-uk/>.
- <sup>51</sup> “HDRUK Innovation Gateway | Homepage,” accessed March 1, 2021, <https://www.healthdatagateway.org/>.
- <sup>52</sup> “About Academic Health Science Networks,” AHSN Network, accessed March 1, 2021, <https://www.ahsnnetwork.com/about-academic-health-science-networks>.
- <sup>53</sup> “Home,” DigitalHealth.London, accessed March 1, 2021, <https://digitalhealth.london/>.
- <sup>54</sup> “OneLondon | North London Partners in Health and Care,” accessed March 16, 2021, <https://www.northlondonpartners.org.uk/ourplan/Areas-of-work/Digital/one-london.htm>.
- <sup>55</sup> “Accelerating Cutting-Edge Innovation in London and the Greater South East of England,” MedCity, accessed March 1, 2021, <https://www.medcityhq.com/>.
- <sup>56</sup> “Artificial Intelligence to Help Save Lives at Five New Technology Centres,” GOV.UK, accessed March 1, 2021, <https://www.gov.uk/government/news/artificial-intelligence-to-help-save-lives-at-five-new-technology-centres>.
- <sup>57</sup> “London Medical Imaging & AI Centre for Value Based Healthcare,” accessed March 1, 2021, <https://www.aicentre.co.uk/>.
- <sup>58</sup> “Home,” *ICAIRD* (blog), accessed March 1, 2021, <https://icaird.com/>.
- <sup>59</sup> “National Consortium of Intelligent Medical Imaging (NCIMI) — University of Oxford, Medical Sciences Division,” accessed March 1, 2021, <https://www.medsci.ox.ac.uk/research/networks/national-consortium-of-intelligent-medical-imaging>.
- <sup>60</sup> “NPIC | Northern Pathology Imaging Co-Operative,” accessed March 1, 2021, <https://www.virtualpathology.leeds.ac.uk/npic/>.
- <sup>61</sup> “Home,” PathLAKE, accessed March 1, 2021, <https://www.pathlake.org/>.
- <sup>62</sup> “Digital Health & Care Innovation Centre,” accessed March 1, 2021, <https://www.dhi-scotland.comhttps://www.dhi-scotland.com>.
- <sup>63</sup> “Health and Medical Sciences,” The Alan Turing Institute, accessed March 1, 2021, <https://www.turing.ac.uk/research/research-programmes/health-and-medical-sciences>.
- <sup>64</sup> “About,” accessed March 1, 2021, <https://www.adalovelaceinstitute.org/about/>.
- <sup>65</sup> “COVID-19 Response,” NHSX, accessed March 1, 2021, <https://www.nhsx.nhs.uk/covid-19-response/>.
- <sup>66</sup> “NHS Digital Response to Coronavirus (COVID-19),” 19.
- <sup>67</sup> “COVID-19,” *HDR UK* (blog), accessed March 1, 2021, <https://www.hdruk.ac.uk/covid-19/>.
- <sup>68</sup> “Covid Corpus,” accessed March 1, 2021, <https://www.covidcorpus.org/Home>.
- <sup>69</sup> Nadia A. S. Smith et al., “Building Confidence in Digital Health through Metrology,” *The British Journal of Radiology* 93, no. 1109 (January 28, 2020): 20190574.
- <sup>70</sup> “Data Analysis, Modelling and Management,” NPLWebsite, accessed March 1, 2021, <https://www.npl.co.uk/covid-response/data-science-data-management>.
- <sup>71</sup> “Supporting the NHS during the Covid-19 Pandemic,” NPLWebsite, accessed March 1, 2021, <https://www.npl.co.uk/our-work/case-studies/supporting-nhs-during-covid19-pandemic>.
- <sup>72</sup> “NPL Digital Health Web,” n.d., <https://www.npl.co.uk/research/data-science/digital-health>.
- <sup>73</sup> Li Sze Chow and Raveendran Paramesran, “Review of Medical Image Quality Assessment,” *Biomedical Signal Processing and Control* 27 (May 1, 2016): 145–54, <https://doi.org/10.1016/j.bspc.2016.02.006>.
- <sup>74</sup> K. J. Ruchala et al., “Megavoltage CT on a Tomotherapy System,” *Physics in Medicine and Biology* 44, no. 10 (September 1999): 2597–2621, <https://doi.org/10.1088/0031-9155/44/10/316>.
- <sup>75</sup> Jennifer B. Smilowitz et al., “Long-Term Dosimetric Stability of Multiple TomoTherapy Delivery Systems,” *Journal of Applied Clinical Medical Physics* 18, no. 3 (2017): 137–43, <https://doi.org/10.1002/acm2.12085>.
- <sup>76</sup> Poonam Yadav et al., “The Effect and Stability of MVCT Images on Adaptive TomoTherapy,” *Journal of Applied Clinical Medical Physics* 11, no. 4 (2010): 4–14, <https://doi.org/10.1120/jacmp.v11i4.3229>.
- <sup>77</sup> Wilfried Schneider, Thomas Bortfeld, and Wolfgang Schlegel, “Correlation between CT Numbers and Tissue Parameters Needed for Monte Carlo Simulations of Clinical Dose Distributions,” *Physics in Medicine and Biology* 45, no. 2 (February 1, 2000): 459–78, <https://doi.org/10.1088/0031-9155/45/2/314>.
- <sup>78</sup> Katja M. Langen et al., “QA for Helical Tomotherapy: Report of the AAPM Task Group 148a), TG-148,” *Medical Physics* 37, no. 9 (August 20, 2010): 4817–53, <https://doi.org/10.1118/1.3462971>.

- <sup>79</sup> Simon J Thomas et al., "Recalculation of Dose for Each Fraction of Treatment on TomoTherapy," *The British Journal of Radiology* 89, no. 1059 (March 2016), <https://doi.org/10.1259/bjr.20150770>.
- <sup>80</sup> F. Crop, A. Bernard, and N. Reynaert, "Improving Dose Calculations on Tomotherapy MVCT Images," *Journal of Applied Clinical Medical Physics* 13, no. 6 (2012): 241–53, <https://doi.org/10.1120/jacmp.v13i6.3986>.
- <sup>81</sup> Neil Gunn Burnet et al., "Applying Physical Science Techniques and CERN Technology to an Unsolved Problem in Radiation Treatment for Cancer: The Multidisciplinary 'VoxTox' Research Programme," *CERN IdeaSquare Journal of Experimental Innovation* 1, no. 1 (June 30, 2017): 3, <https://doi.org/10.23726/cij.2017.457>.
- <sup>82</sup> Langen et al., "QA for Helical Tomotherapy."
- <sup>83</sup> "NEMA PS3 / ISO 12052, Digital Imaging and Communications in Medicine (DICOM) Standard, National Electrical Manufacturers Association, Rosslyn, VA, USA," DICOM, accessed March 30, 2021, <https://www.dicomstandard.org/current>.
- <sup>84</sup> Sonja Dieterich et al., "Chapter 11 - Information Technology in Radiation Oncology," in *Practical Radiation Oncology Physics*, ed. Sonja Dieterich et al. (Philadelphia: Elsevier, 2016), 156–64, <https://doi.org/10.1016/B978-0-323-26209-5.00011-0>.
- <sup>85</sup> Jean-Pierre Bissonnette et al., "Quality Assurance for Image-Guided Radiation Therapy Utilizing CT-Based Technologies: A Report of the AAPM TG-179," *Medical Physics* 39, no. 4 (2012): 1946–63, <https://doi.org/10.1118/1.3690466>.
- <sup>86</sup> S J Thomas, "Relative Electron Density Calibration of CT Scanners for Radiotherapy Treatment Planning.," *The British Journal of Radiology* 72, no. 860 (August 1, 1999): 781–86, <https://doi.org/10.1259/bjr.72.860.10624344>.
- <sup>87</sup> George Box and Friends, *Improving Almost Anything*, Revised Edition (Wiley, 2014).
- <sup>88</sup> E. S. Page, "CONTINUOUS INSPECTION SCHEMES," *Biometrika* 41, no. 1–2 (June 1, 1954): 100–115, <https://doi.org/10.1093/biomet/41.1-2.100>.
- <sup>89</sup> "Cancer Survivorship," The Cancer Atlas, accessed February 19, 2020, <http://canceratlas.cancer.org/rky>.
- <sup>90</sup> J. Maddams, M. Utley, and H. Møller, "Projections of Cancer Prevalence in the United Kingdom, 2010–2040," *British Journal of Cancer* 107, no. 7 (September 25, 2012): 1195–1202, <https://doi.org/10.1038/bjc.2012.366>.
- <sup>91</sup> Ana Correa et al., "Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC) Sentinel Network: A Cohort Profile," *BMJ Open* 6, no. 4 (April 1, 2016), <https://doi.org/10.1136/bmjopen-2016-011092>.
- <sup>92</sup> "Medical Informatics Europe 2020 - Geneva," accessed March 16, 2021, <http://www.mie2020.org/>.
- <sup>93</sup> Agnieszka Lemanska et al., "Primary Care Prostate Cancer Case Ascertainment," n.d., 2.
- <sup>94</sup> Correa et al., "Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC) Sentinel Network," April 1, 2016.
- <sup>95</sup> Simon de Lusignan et al., "RCGP Research and Surveillance Centre: 50 Years' Surveillance of Influenza, Infections, and Respiratory Conditions," *British Journal of General Practice* 67, no. 663 (October 1, 2017): 440–41, <https://doi.org/10.3399/bjgp17X692645>.
- <sup>96</sup> Zayd Tippu et al., "Ethnicity Recording in Primary Care Computerised Medical Record Systems: An Ontological Approach," *Journal of Innovation in Health Informatics* 23, no. 4 (14 2017): 920, <https://doi.org/10.14236/jhi.v23i4.920>.
- <sup>97</sup> C. J. Hirst, C. Cabrera, and M. Kirby, "Epidemiology of Castration Resistant Prostate Cancer: A Longitudinal Analysis Using a UK Primary Care Database," *Cancer Epidemiology* 36, no. 6 (December 2012): e349–353, <https://doi.org/10.1016/j.canep.2012.07.012>.
- <sup>98</sup> N. Smith et al., *Automated Differentiation of Incident and Prevalent Cases in Primary Care Computerised Medical Records (CMR)*, vol. 247 (IOS Press, 2019), <https://ora.ox.ac.uk/objects/uuid:37dc69b5-fbd6-442b-8fd3-d04831a9eff0>.
- <sup>99</sup> Paolo Campanella et al., "The Impact of Electronic Health Records on Healthcare Quality: A Systematic Review and Meta-Analysis," *European Journal of Public Health* 26, no. 1 (February 1, 2016): 60–64, <https://doi.org/10.1093/eurpub/ckv122>.
- <sup>100</sup> Christopher Martin Pearce et al., "The Computerized Medical Record as a Tool for Clinical Governance in Australian Primary Care," *Interactive Journal of Medical Research* 2, no. 2 (August 12, 2013): e26, <https://doi.org/10.2196/ijmr.2700>.
- <sup>101</sup> Ruth Farmer et al., "Promises and Pitfalls of Electronic Health Record Analysis," *Diabetologia* 61, no. 6 (June 1, 2018): 1241–48, <https://doi.org/10.1007/s00125-017-4518-6>.



- <sup>102</sup> N. Smith et al., *Automated Differentiation of Incident and Prevalent Cases in Primary Care Computerised Medical Records (CMR)*, vol. 247 (IOS Press, 2019), <https://ora.ox.ac.uk/objects/uuid:37dc69b5-fbd6-442b-8fd3-d04831a9eff0>.
- <sup>103</sup> Smith et al.; Alvaro Ulloa et al., "An Unsupervised Homogenization Pipeline for Clustering Similar Patients Using Electronic Health Record Data," *ArXiv:1801.00065 [q-Bio]*, March 21, 2018, <http://arxiv.org/abs/1801.00065>.
- <sup>104</sup> D. M. Fleming, F. G. Schellevis, and W. J. Paget, "Health Monitoring in Sentinel Practice Networks: The Contribution of Primary Care," *European Journal of Public Health* 13, no. suppl\_3 (September 1, 2003): 80–84, [https://doi.org/10.1093/eurpub/13.suppl\\_3.80](https://doi.org/10.1093/eurpub/13.suppl_3.80).
- <sup>105</sup> Cécile Souty et al., "Improving Disease Incidence Estimates in Primary Care Surveillance Systems," *Population Health Metrics* 12, no. 1 (July 26, 2014): 19, <https://doi.org/10.1186/s12963-014-0019-8>.
- <sup>106</sup> "RCGP Research and Surveillance Centre: 50 Years' Surveillance of Influenza, Infections, and Respiratory Conditions | British Journal of General Practice," accessed February 18, 2020, <https://bjgp.org/content/67/663/440>.
- <sup>107</sup> "RCGP Research and Surveillance Centre: 50 Years' Surveillance of Influenza, Infections, and Respiratory Conditions | British Journal of General Practice."
- <sup>108</sup> N. Smith et al., *Automated Differentiation of Incident and Prevalent Cases in Primary Care Computerised Medical Records (CMR)*, vol. 247 (IOS Press, 2019), <https://ora.ox.ac.uk/objects/uuid:37dc69b5-fbd6-442b-8fd3-d04831a9eff0>.
- <sup>109</sup> Campanella et al., "The Impact of Electronic Health Records on Healthcare Quality."
- <sup>110</sup> Freda Mold et al., "The Impact of Patient Online Access to Computerized Medical Records and Services on Type 2 Diabetes: Systematic Review," *Journal of Medical Internet Research* 20, no. 7 (06 2018): e235, <https://doi.org/10.2196/jmir.7858>.
- <sup>111</sup> Anand Avati et al., "Improving Palliative Care with Deep Learning," *BMC Medical Informatics and Decision Making* 18, no. Suppl 4 (12 2018): 122, <https://doi.org/10.1186/s12911-018-0677-8>.
- <sup>112</sup> Kasper Jensen et al., "Analysis of Free Text in Electronic Health Records for Identification of Cancer Patient Trajectories," *Scientific Reports* 7, no. 1 (April 7, 2017): 1–12, <https://doi.org/10.1038/srep46226>.
- <sup>113</sup> Teeradache Viangteeravat, Oguz Akbilgic, and Robert Lowell Davis, "Analyzing Electronic Medical Records to Predict Risk of DIT (Death, Intubation, or Transfer to ICU) in Pediatric Respiratory Failure or Related Conditions," *AMIA Joint Summits on Translational Science Proceedings. AMIA Joint Summits on Translational Science 2017* (2017): 287–94.
- <sup>114</sup> Campanella et al., "The Impact of Electronic Health Records on Healthcare Quality."
- <sup>115</sup> Anand Avati et al., "Improving Palliative Care with Deep Learning," *BMC Medical Informatics and Decision Making* 18, no. Suppl 4 (12 2018): 122, <https://doi.org/10.1186/s12911-018-0677-8>.
- <sup>116</sup> Samuel J. Wang et al., "A Cost-Benefit Analysis of Electronic Medical Records in Primary Care," *The American Journal of Medicine* 114, no. 5 (April 1, 2003): 397–403, [https://doi.org/10.1016/s0002-9343\(03\)00057-3](https://doi.org/10.1016/s0002-9343(03)00057-3).
- <sup>117</sup> Wang et al.
- <sup>118</sup> Correa et al., "Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC) Sentinel Network," April 1, 2016.
- <sup>119</sup> Ana Correa et al., "Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC) Sentinel Network: A Cohort Profile," *BMJ Open* 6, no. 4 (April 1, 2016), <https://doi.org/10.1136/bmjopen-2016-011092>; Simon de Lusignan et al., "RCGP Research and Surveillance Centre: 50 Years' Surveillance of Influenza, Infections, and Respiratory Conditions," *British Journal of General Practice* 67, no. 663 (October 1, 2017): 440–41, <https://doi.org/10.3399/bjgp17X692645>.
- <sup>120</sup> "RCGP Research and Surveillance Centre: 50 Years' Surveillance of Influenza, Infections, and Respiratory Conditions | British Journal of General Practice."
- <sup>121</sup> Avati et al., "Improving Palliative Care with Deep Learning"; Spencer A. Thomas et al., "Analysis of Primary Care Computerized Medical Records (CMR) Data With Deep Autoencoders (DAE)," *Frontiers in Applied Mathematics and Statistics* 5 (2019), <https://doi.org/10.3389/fams.2019.00042>.
- <sup>122</sup> Thomas et al., "Analysis of Primary Care Computerized Medical Records (CMR) Data With Deep Autoencoders (DAE)."
- <sup>123</sup> C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal* 27, no. 3 (n.d.): 379–423.

- <sup>124</sup> Spencer A. Thomas et al., “Enhancing Classification of Mass Spectrometry Imaging Data with Deep Neural Networks,” in *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2017, 1–8, <https://doi.org/10.1109/SSCI.2017.8285223>.
- <sup>125</sup> Thomas et al.
- <sup>126</sup> Thomas et al., “Analysis of Primary Care Computerized Medical Records (CMR) Data With Deep Autoencoders (DAE).”
- <sup>127</sup> Thomas et al., “Enhancing Classification of Mass Spectrometry Imaging Data with Deep Neural Networks.”
- <sup>128</sup> Correa et al., “Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC) Sentinel Network,” April 1, 2016; Lusignan et al., “RCGP Research and Surveillance Centre,” October 1, 2017.
- <sup>129</sup> Beata Nowok, Gillian M. Raab, and Chris Dibben, “Synthpop: Bespoke Creation of Synthetic Data in R,” *Journal of Statistical Software* 74, no. 1 (October 28, 2016): 1–26, <https://doi.org/10.18637/jss.v074.i11>.
- <sup>130</sup> “ISIC Archive,” accessed February 20, 2020, <https://www.isic-archive.com/#!/topWithHeader/tightContentTop/challenges>.
- <sup>131</sup> “ISIC Archive.”
- <sup>132</sup> “National Lung Screening Trial (NLST),” *cgvArticle*, National Cancer Institute, September 8, 2014, <https://www.cancer.gov/types/lung/research/nlst>.
- <sup>133</sup> “ISIC Archive.”
- <sup>134</sup> T. Rai et al., “Can ImageNet Feature Maps Be Applied to Small Histopathological Datasets for the Classification of Breast Cancer Metastatic Tissue in Whole Slide Images?,” in *Medical Imaging 2019: Digital Pathology*, vol. 10956 (Medical Imaging 2019: Digital Pathology, International Society for Optics and Photonics, 2019), 109560V, <https://doi.org/10.1117/12.2512853>. T. Rai et al., “Can ImageNet Feature Maps Be Applied to Small Histopathological Datasets for the Classification of Breast Cancer Metastatic Tissue in Whole Slide Images?,” in *Medical Imaging 2019: Digital Pathology*, vol. 10956 (Medical Imaging 2019: Digital Pathology, International Society for Optics and Photonics, 2019), 109560V, <https://doi.org/10.1117/12.2512853>.
- <sup>135</sup> Rai et al., “Can ImageNet Feature Maps Be Applied to Small Histopathological Datasets for the Classification of Breast Cancer Metastatic Tissue in Whole Slide Images?,” 2019. Rai et al., “Can ImageNet Feature Maps Be Applied to Small Histopathological Datasets for the Classification of Breast Cancer Metastatic Tissue in Whole Slide Images?,” 2019.
- <sup>136</sup> Thomas et al., “Analysis of Primary Care Computerized Medical Records (CMR) Data With Deep Autoencoders (DAE).”
- <sup>137</sup> Simon de Lusignan et al., “Analysis of Primary Care Computerised Medical Records with Deep Learning,” *Studies in Health Technology and Informatics* 258 (January 1, 2019): 249–50.
- <sup>138</sup> Matthew, Gould, Ming Tang, and Indra Joshi, “The Power of Data in a Pandemic - Technology in the NHS,” accessed February 16, 2021, <https://healthtech.blog.gov.uk/2020/03/28/the-power-of-data-in-a-pandemic/>.
- <sup>139</sup> Allan Tucker et al., “Generating High-Fidelity Synthetic Patient Data for Assessing Machine Learning Healthcare Software,” *Npj Digital Medicine* 3, no. 1 (November 9, 2020): 1–13, <https://doi.org/10.1038/s41746-020-00353-9>.
- <sup>140</sup> “General Data Protection Regulation (GDPR) – Official Legal Text,” General Data Protection Regulation (GDPR), accessed February 19, 2020, <https://gdpr-info.eu/>.
- <sup>141</sup> Maayan Frid-Adar et al., “Synthetic Data Augmentation Using GAN for Improved Liver Lesion Classification,” in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 2018, 289–93, <https://doi.org/10.1109/ISBI.2018.8363576>.
- <sup>142</sup> Frank Preiswerk et al., “Synthesizing Dynamic MRI Using Long-Term Recurrent Convolutional Networks,” *ArXiv:1807.09305 [Eess]*, July 24, 2018, <http://arxiv.org/abs/1807.09305>; E. A. Nadaraya, “On Estimating Regression,” *Theory of Probability & Its Applications* 9, no. 1 (January 1, 1964): 141–42, <https://doi.org/10.1137/1109020>.
- <sup>143</sup> Ian Goodfellow et al., “Generative Adversarial Nets,” n.d., 9.
- <sup>144</sup> Berkman Sahiner et al., “Deep Learning in Medical Imaging and Radiation Therapy,” *Medical Physics* 46, no. 1 (2019): e1–36, <https://doi.org/10.1002/mp.13264>.
- <sup>145</sup> Yongsik Jung et al., “The Feasibility of Synthetic MRI in Breast Cancer Patients: Comparison of T2 Relaxation Time with Multiecho Spin Echo T2 Mapping Method,” *The British Journal of Radiology* 92, no. 1093 (September 21, 2018): 20180479, <https://doi.org/10.1259/bjr.20180479>.

- 
- <sup>146</sup> Burnet et al., “Applying Physical Science Techniques and CERN Technology to an Unsolved Problem in Radiation Treatment for Cancer.”
- <sup>147</sup> Weronika Jackowiak et al., “Influence of the Type of Imaging on the Delineation Process during the Treatment Planning,” *Reports of Practical Oncology and Radiotherapy* 20, no. 5 (2015): 351–57, <https://doi.org/10.1016/j.rpor.2015.05.004>.
- <sup>148</sup> Thomas Rockwell Mackie et al., “Image Guidance for Precise Conformal Radiotherapy,” *International Journal of Radiation Oncology, Biology, Physics* 56, no. 1 (May 1, 2003): 89–105, [https://doi.org/10.1016/s0360-3016\(03\)00090-7](https://doi.org/10.1016/s0360-3016(03)00090-7).
- <sup>149</sup> Spencer Martin and Slav Yartsev, “KVCT, MVCT, and Hybrid CT Image Studies—Treatment Planning and Dose Delivery Equivalence on Helical Tomotherapy,” *Medical Physics* 37, no. 6 (June 2010): 2847–54, <https://doi.org/10.1118/1.3432566>.
- <sup>150</sup> Goodfellow et al., “Generative Adversarial Nets.”
- <sup>151</sup> David Abramian and Anders Eklund, “Generating FMRI Volumes from T1-Weighted Volumes Using 3D CycleGAN,” *ArXiv:1907.08533 [Cs, Eess]*, October 18, 2019, <http://arxiv.org/abs/1907.08533>.
- <sup>152</sup> Weining Hu, Meng Li, and Xiaomeng Ju, “Improved CycleGAN for Image-to-Image Translation,” n.d., 9.
- <sup>153</sup> Phillip Isola et al., “Image-to-Image Translation with Conditional Adversarial Networks,” *ArXiv:1611.07004 [Cs]*, November 26, 2018, <http://arxiv.org/abs/1611.07004>.
- <sup>154</sup> Ian Goodfellow, “NIPS 2016 Tutorial: Generative Adversarial Networks,” *ArXiv:1701.00160 [Cs]*, April 3, 2017, <http://arxiv.org/abs/1701.00160>.
- <sup>155</sup> Goodfellow.
- <sup>156</sup> Jun-Yan Zhu et al., “Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks,” *ArXiv:1703.10593 [Cs]*, August 24, 2020, <http://arxiv.org/abs/1703.10593>.
- <sup>157</sup> Abramian and Eklund, “Generating FMRI Volumes from T1-Weighted Volumes Using 3D CycleGAN.”
- <sup>158</sup> Zhu et al., “Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks.”
- <sup>159</sup> Rowel Atienza, *Advanced Deep Learning with TensorFlow 2 and Keras: Apply DL, GANs, VAEs, Deep RL, Unsupervised Learning, Object Detection and Segmentation, and More, 2nd Edition* (Packt Publishing Ltd, 2020).
- <sup>160</sup> “Module: Tf.Keras.Optimizers | TensorFlow Core v2.4.1,” TensorFlow, accessed February 17, 2021, [https://www.tensorflow.org/api\\_docs/python/tf/keras/optimizers](https://www.tensorflow.org/api_docs/python/tf/keras/optimizers).
- <sup>161</sup> “Module: Tf.Keras.Losses | TensorFlow Core v2.4.1,” TensorFlow, accessed February 17, 2021, [https://www.tensorflow.org/api\\_docs/python/tf/keras/losses](https://www.tensorflow.org/api_docs/python/tf/keras/losses).
- <sup>162</sup> “Module: Tf.Keras.Metrics | TensorFlow Core v2.4.1,” TensorFlow, accessed February 17, 2021, [https://www.tensorflow.org/api\\_docs/python/tf/keras/metrics](https://www.tensorflow.org/api_docs/python/tf/keras/metrics).
- <sup>163</sup> Atienza, *Advanced Deep Learning with TensorFlow 2 and Keras*.
- <sup>164</sup> S Tong, A M Alessio, and P E Kinahan, “Noise and Signal Properties in PSF-Based Fully 3D PET Image Reconstruction: An Experimental Evaluation,” *Physics in Medicine and Biology* 55, no. 5 (March 7, 2010): 1453–73, <https://doi.org/10.1088/0031-9155/55/5/013>.
- <sup>165</sup> Harrison H. Barrett and Kyle J. Myers, *Foundations of Image Science* (Wiley, 2004).
- <sup>166</sup> Goodfellow, “NIPS 2016 Tutorial.”
- <sup>167</sup> Augustus Odena, Vincent Dumoulin, and Chris Olah, “Deconvolution and Checkerboard Artifacts,” *Distill* 1, no. 10 (October 17, 2016): e3, <https://doi.org/10.23915/distill.00003>.
- <sup>168</sup> Odena, Dumoulin, and Olah.
- <sup>169</sup> Odena, Dumoulin, and Olah.
- <sup>170</sup> Tucker et al., “Generating High-Fidelity Synthetic Patient Data for Assessing Machine Learning Healthcare Software”; Abramian and Eklund, “Generating FMRI Volumes from T1-Weighted Volumes Using 3D CycleGAN.”
- <sup>171</sup> P. J. Julyan et al., “SUVpeak: A New Parameter for Quantification of Uptake in FDG PET,” *Nuclear Medicine Communications* 25, no. 4 (April 2004): 407.
- <sup>172</sup> Martin Heusel et al., “GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium,” *ArXiv:1706.08500 [Cs, Stat]*, January 12, 2018, <http://arxiv.org/abs/1706.08500>.
- <sup>173</sup> Aharon Azulay and Yair Weiss, “Why Do Deep Convolutional Networks Generalize so Poorly to Small Image Transformations?,” *ArXiv:1805.12177 [Cs]*, December 31, 2019, <http://arxiv.org/abs/1805.12177>.
- <sup>174</sup> Andrew Brock et al., “High-Performance Large-Scale Image Recognition Without Normalization,” *ArXiv:2102.06171 [Cs, Stat]*, February 11, 2021, <http://arxiv.org/abs/2102.06171>.

- <sup>175</sup> Christopher J. Kelly et al., “Key Challenges for Delivering Clinical Impact with Artificial Intelligence,” *BMC Medicine* 17, no. 1 (December 2019): 195, <https://doi.org/10.1186/s12916-019-1426-2>.
- <sup>176</sup> Martin J. Willemink et al., “Preparing Medical Imaging Data for Machine Learning,” *Radiology*, February 18, 2020, 192224, <https://doi.org/10.1148/radiol.2020192224>.
- <sup>177</sup> “PHE COVID-19: Mental Health and Wellbeing Surveillance Report,” GOV.UK, accessed September 18, 2020, <https://www.gov.uk/government/publications/covid-19-mental-health-and-wellbeing-surveillance-report/1-about-this-report>.
- <sup>178</sup> “Monitoring the Mental Health Impact of COVID-19 to Build a Roadmap for Recovery,” *HDR UK* (blog), accessed September 17, 2020, <https://www.hdruk.ac.uk/news/monitoring-the-mental-health-impact-of-covid-19-to-build-a-roadmap-for-recovery/>.
- <sup>179</sup> “Survivors of Covid-19 Show Increased Rate of Psychiatric Disorders, Study Finds,” the Guardian, August 3, 2020, 19, <http://www.theguardian.com/world/2020/aug/03/survivors-of-covid-19-show-increased-rate-of-psychiatric-disorders-study-finds>.
- <sup>180</sup> Catherine E. Robb et al., “Associations of Social Isolation with Anxiety and Depression During the Early COVID-19 Pandemic: A Survey of Older Adults in London, UK,” *Frontiers in Psychiatry* 11 (September 17, 2020), <https://doi.org/10.3389/fpsy.2020.591120>.
- <sup>181</sup> Nicola Davis, “‘Sheer Fear’: Mental Health Impacts of Covid-19 Come to Fore,” *The Guardian*, August 14, 2020, sec. World news, <https://www.theguardian.com/world/2020/aug/14/sheer-fear-mental-health-impacts-of-covid-19-come-to-fore>.
- <sup>182</sup> Amy Dawel et al., “The Effect of COVID-19 on Mental Health and Wellbeing in a Representative Sample of Australian Adults,” *Frontiers in Psychiatry* 11 (2020), <https://doi.org/10.3389/fpsy.2020.579985>.
- <sup>183</sup> Pallab K. Maulik, Graham Thornicroft, and Shekhar Saxena, “Roadmap to Strengthen Global Mental Health Systems to Tackle the Impact of the COVID-19 Pandemic,” *International Journal of Mental Health Systems* 14, no. 1 (December 2020): 57, <https://doi.org/10.1186/s13033-020-00393-4>.
- <sup>184</sup> Robb et al., “Associations of Social Isolation with Anxiety and Depression During the Early COVID-19 Pandemic.”
- <sup>185</sup> Richard Williams et al., “Diagnosis of Physical and Mental Health Conditions in Primary Care during the COVID-19 Pandemic: A Retrospective Cohort Study,” *The Lancet Public Health* 5, no. 10 (October 2020): e543–50, [https://doi.org/10.1016/S2468-2667\(20\)30201-2](https://doi.org/10.1016/S2468-2667(20)30201-2).
- <sup>186</sup> Correa et al., “Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC) Sentinel Network,” April 1, 2016.
- <sup>187</sup> Williams et al., “Diagnosis of Physical and Mental Health Conditions in Primary Care during the COVID-19 Pandemic.”
- <sup>188</sup> Jules Mesnier et al., “Hospital Admissions for Acute Myocardial Infarction before and after Lockdown According to Regional Prevalence of COVID-19 and Patient Profile in France: A Registry Study,” *The Lancet Public Health* 5, no. 10 (October 2020): e536–42, [https://doi.org/10.1016/S2468-2667\(20\)30188-2](https://doi.org/10.1016/S2468-2667(20)30188-2).
- <sup>189</sup> C. H. Earnshaw et al., “Reduction in Skin Cancer Diagnosis, and Overall Cancer Referrals, during the COVID-19 Pandemic,” *British Journal of Dermatology* 183, no. 4 (2020): 792–94, <https://doi.org/10.1111/bjd.19267>.
- <sup>190</sup> Dawel et al., “The Effect of COVID-19 on Mental Health and Wellbeing in a Representative Sample of Australian Adults.”
- <sup>191</sup> “Survivors of Covid-19 Show Increased Rate of Psychiatric Disorders, Study Finds.”
- <sup>192</sup> Vijay Pattni, Jeffrey Phillips, and Rajnish Saha, “Psychiatric and Neuropsychiatric Syndromes and COVID-19,” *The Lancet Psychiatry* 7, no. 8 (August 2020): 664, [https://doi.org/10.1016/S2215-0366\(20\)30299-6](https://doi.org/10.1016/S2215-0366(20)30299-6).
- <sup>193</sup> Jonathan P. Rogers et al., “Psychiatric and Neuropsychiatric Presentations Associated with Severe Coronavirus Infections: A Systematic Review and Meta-Analysis with Comparison to the COVID-19 Pandemic,” *The Lancet. Psychiatry* 7, no. 7 (2020): 611–27, [https://doi.org/10.1016/S2215-0366\(20\)30203-0](https://doi.org/10.1016/S2215-0366(20)30203-0).
- <sup>194</sup> “Data Analysis, Modelling and Management,” NPLWebsite, accessed March 1, 2021, <https://www.npl.co.uk/covid-response/data-science-data-management>.