Using information from electronic patient records for clinical, epidemiological and health services research

Submitted for the degree of Doctor of Philosophy

Angela Gibson-White

2015

Imperial College London
School of Public Health
Declaration of Originality

I (Angela Gibson-White) declare that the contents of this thesis are my own work. Where the work of others has been used, this has been indicated and appropriately referenced.
Copyright Declaration

‘The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work’
Acknowledgements

Completing this PhD would have been an impossible task without first and foremost, the help of God. I thank him for seeing me through from start to finish.

Many people have also played a significant role in making it possible, and I would like to first thank my supervisor Professor Azeem Majeed, for his sound supervision, guidance and constant push to ensure that the work was completed well; without all your contributions I would have been lost. I acknowledge the help of four great medical students who supported the data entry work, Usama Syed, Fahmida Mannan, Zain Chaudhry and Shirin Ahmed; thank you for your help with pulling data from all the databases used. I would also like to thank Dr. Antonis A. Kousoulis at Clinical Practice Research Datalink (CPRD) for all the help provided with obtaining up to date CPRD information and for proofreading some of the CPRD work done.

I must also recognise the contribution from my family and friends, their constant prayers and encouragement were invaluable. I would like to thank Shola Solanke for pushing me to apply to Imperial College; I was not thinking that big, but am glad I listened. Thank you Lanre Iroche for proofreading all those pages for me, I am sure it was such a chore and I really appreciate it! Pastor Shola Adeaga thank you for your prayers and for advising me not to change any deadline set, because it was time to finish. Thank you Pastors Agu and Sola Irukwu for your prayers and for calling me Doctor from day one, practically using this to ensure I did not quit.

To my wonderful siblings, Lizzy, Arinola and Jumoke, thank you for being so proud of me and pushing me to keep going. Thank you Dr Ife Nnatuanya for all your help and advice, your shared experience was invaluable. To my awesome mother, thank you for telling me all the time that getting this PhD was a possibility. You just firmly believed that I would make it to the end, and yes I did, even though it took seven years!

To my two beautiful children, Ore and Oreofe, thank you for asking me every day when exactly I planned to finish because it was taking too long; it made me to want to see the end of it too. Last but not least, to my wonderful and most supportive husband, Wale, for thinking about this PhD sometimes even more than I did. Thank you for believing in me, pushing me and not letting me give up when the thought crossed my mind.
Abstract

**Background:** Improving current and future healthcare is heavily reliant on continuous research and the secondary use of data from patients' medical records, particularly from electronic records. Considerable amounts of data are collected during the care and treatment of a patient, and this data can offer many opportunities, not only for supporting and improving individual patient care or making important contributions to research, but also for investigating causes of diseases, establishing the prevalence of risk factors, and identifying populations at risk of adverse outcomes. However, the management of such data poses challenges, which many believe can be mitigated by storing it electronically. The traditional method of storing medical information in a paper-based format has severe limitations, especially concerning the amount of effort needed to extract information. In contrast, data from electronic patient records (EPRs) is much easier to extract and allows healthcare professionals access to the information needed in a timely manner to provide appropriate care to patients and improve the public’s health. The UK still faces the hurdle of balancing public interest with individual privacy. There is clearly a benefit regarding the use of EPRs but there is an increasing need for public education in order to be able to reap the maximum benefits they offer. This thesis examines the benefits and impact of EPRs in the contexts of clinical care and epidemiological and health services research.

**Methods:** The methods used for this research project involved reviewing published materials available through electronic searching, grey literature and websites of bodies such as the Department of Health, and the Health and Social Care Information Centre. The use of the main national primary care databases and secondary care databases and their growth over time was also examined.

**Results:** EPRs are extremely beneficial to research and have a significant potential to improve patient overall care. The use of EPRs is growing as technology advances and health systems move from paper to electronic records.

**Conclusions:** The use of EPRs will only be successful when both the public, researchers and healthcare providers agree on their benefits. The use of EPRs will take healthcare to another level, where the accuracy of data entered is of very high quality and standardised, data security is well-controlled, and there is acceptance by the public concerning the use of their data both for providing clinical care and for other secondary uses.
Table of Contents

DECLARATION OF ORIGINALITY .................................................................................. 2
COPYRIGHT DECLARATION ............................................................................................ 3
ACKNOWLEDGEMENTS .................................................................................................... 4
ABSTRACT .............................................................................................................................. 5
TABLE OF CONTENTS ........................................................................................................ 6
LIST OF FIGURES ............................................................................................................... 10
LIST OF TABLES ................................................................................................................. 10
LIST OF ABBREVIATIONS ............................................................................................... 11
CHAPTER 1 - GENERAL BACKGROUND ...................................................................... 13
  1.1 Identified Research Gap ................................................................................................. 13
  1.2 Research Plan ................................................................................................................. 14
  1.3 Aims and Objectives ...................................................................................................... 14
  1.4 Outline of the Thesis ...................................................................................................... 15
CHAPTER 2 - INTRODUCTION ........................................................................................ 17
  2.1 Definition of Electronic Patient Records ....................................................................... 19
  2.2 Secondary Uses of EPRs and Associated Benefits ........................................................ 19
  2.3 NHS National Programme for Information Technology ................................................ 20
  2.4 Commissioned Simulations ............................................................................................ 21
  2.4.1 Surveillance ............................................................................................................. 21
  2.4.2 Interventional Clinical Trials .................................................................................. 22
  2.4.3 Prospective Tracking of Identified Cohorts of Patients .......................................... 22
  2.4.4 Observational Epidemiological Research ............................................................... 23
  2.4.5 Outcomes of the Simulations ................................................................................... 24
  2.5 Attempt to Implement the NHS Connecting for Health Plan ................................................ 26
  2.5.1 Personal Demographics Service ............................................................................. 26
  2.5.2 Summary Care Record ............................................................................................ 26
  2.5.3 Choose and Book ..................................................................................................... 26
  2.5.4 Electronic Prescription Service ............................................................................... 27
  2.5.5 Secondary Uses Service ........................................................................................... 27
  2.6 Reasons for the NHS Connecting for Health Failure ..................................................... 27
  2.6.1 Confidentiality Issues .............................................................................................. 27
  2.6.2 Procurement and Cost Issues .................................................................................. 28
  2.6.3 Political Reasons ..................................................................................................... 30
  2.7 care.data - Sharing Information .................................................................................... 30
  2.7.1 Benefits and the Risks Associated with care.data ................................................... 31
  2.7.2 Abandonment of the Project .................................................................................. 32
  2.8 General Practice Extraction Services ............................................................................. 33
2.8.1 Maintaining Patient Confidentiality ................................................................. 34
2.9 Chapter Summary ..................................................................................................... 35

CHAPTER 3 - LITERATURE REVIEW ........................................................................... 36

3.1 Primary and Secondary Care UK Databases.......................................................... 36
3.1.1 Primary Care Databases ................................................................................... 37
3.1.2 Clinical Governance and Quality Improvement ............................................... 37
3.1.3 Health Serving Planning, Need Assessments and Commissioning .................... 37
3.1.4 Regulation, Accountability and Performance Management ............................... 37
3.1.5 Monitoring Healthcare Use ............................................................................. 38
3.1.6 Monitoring Death Rates ................................................................................. 38
3.1.7 Monitoring Health Inequalities ....................................................................... 38
3.1.8 Pharmaco-epidemiology ............................................................................... 38
3.1.9 Public Health and Health Services Research .................................................. 38
3.1.10 Financial Flows and Payment by Result ......................................................... 39
3.1.11 Resource Allocation, Risk Adjustment and Case Mix ....................................... 39
3.1.12 Clinical Practice Research Datalink ............................................................... 39
3.1.13 QResearch ..................................................................................................... 40
3.1.14 The Health Improvement Network ................................................................. 40
3.1.15 ResearchOne .................................................................................................. 41
3.1.16 Prescribing Databases .................................................................................. 41
3.1.17 Quality and Outcomes Framework ................................................................. 42
3.1.18 Secondary Care Database: Hospital Episode Statistics .................................... 44

3.2 Secondary Uses Services ......................................................................................... 46

3.3 Benefits of Healthcare Databases ........................................................................... 47

3.4 Chapter Summary ..................................................................................................... 48

CHAPTER 4 - CHALLENGES ASSOCIATED WITH THE MAXIMISATION OF
CURRENT EPRS FOR RESEARCH .............................................................................. 49

4.1 Introduction ............................................................................................................. 49
4.2 The Patient and Public View of EPRs and Confidentiality ...................................... 49
4.3 Privacy Laws and their Impact on Research ............................................................ 52
4.3.1 Data Protection Act 1998 ............................................................................... 53
4.3.2 Human Rights Law 1998 ............................................................................... 55
4.3.3 Common Law Confidentiality ........................................................................ 56
4.3.4 Administrative Law ....................................................................................... 57
4.3.5 Section 60 of the Health and Social Care Act 2001 ........................................ 57
4.3.6 Health and Social Care Act 2012 ................................................................... 58
4.3.7 Caldicott Principles 1997 ............................................................................. 59
4.4 Data Security .......................................................................................................... 59
4.5 Chapter Summary .................................................................................................... 61

CHAPTER 5 - OVERVIEW OF EPRS IN EUROPE, COMPARING WITH THE UK 63

5.1 Introduction .......................................................................................................... 63
8.3 Contribution to Science .................................................................................................. 91
8.4 Recommendations .......................................................................................................... 92
8.5 Conclusions .................................................................................................................... 93

REFERENCES ....................................................................................................................... 95

APPENDIX 1 – PUBLISHED ARTICLE: THE WELLCOM TRUST REPORT:
MOVING FORWARDS THE USE OF GENERAL PRACTICE ELECTRONIC
PATIENT RECORDS FOR RESEARCH ........................................................................ 105

APPENDIX 2 – SUBMITTED FOR PUBLICATION: RESEARCH OUTPUTS OF
PRIMARY CARE DATABASES IN THE UNITED KINGDOM: BIBLIOMETRIC
ANALYSIS ........................................................................................................................... 108

APPENDIX 3 – SUBMITTED FOR PUBLICATION: RESEARCH OUTPUTS OF
ENGLAND’S HOSPITAL EPISODE STATISTICS (HES) DATABASE:
BIBLIOMETRIC ANALYSIS ............................................................................................ 118
List of Figures

Figure 6.1: Overview of bibliometric analysis ................................................................. 72
Figure 7.1: Cumulative growth of primary care database publications from 2004-2013 ....... 77
Figure 7.2: Cumulative growth of publications using HES data from 1996-2014 ............... 78

List of Tables

Table 3.1: The three components of QOF ........................................................................... 43
Table 6.1: Search Criteria used in Web of Science ............................................................. 71
Table 7.1: Publications utilising CPRD data between 1993 and 2013 .................................. 74
Table 7.2: Publications utilising QResearch data between 2004-2013 ............................... 75
Table 7.3: Publications utilising THIN data between 2004-2013 ....................................... 75
Table 7.4: Annual number of publications published per database between 2004 and 2013 .. 76
Table 7.5: Top three journals most often published in across all three databases from 2004- 2013 ..................................................................................................................... 77
Table 7.6: Number of publications using HES data between 1996 and 2014 ...................... 78
Table 7.7: Speciality areas of publications using HES data between 1996 and 2014 .......... 79
Table 7.8: Journals with highest number of HES data publication ....................................... 80
Abbreviations

List of Abbreviations

ADRs  Adverse Drug Reactions
AOMRC  Academy of Medical Royal Colleges
BMA  British Medical Association
CCG  Clinical Commissioning Group
CiH  Connecting For Health
CPRD  Clinical Practice Research Datalink
CQRS  Calculated Quality Reporting Service
CRC  Clinical Research Collaboration
CRS  Care Record Service
DMP  Dossier Medical Person
EPR  Electronic Patients Record
EMR  Electronic Medical Record
ES  Enhanced Services
GMC  General Medical Council
GMS  General Medical Services
GP  General Practitioner
GPES  General Practice Extraction Service
GPRD  General Practice Research Datalink
HES  Hospital Episode Statistics
HISS  Hospital Information Support System
HRA  Health Research Authority
HSCIC  Health and Social Care Information Centre
IAG  Independent Advisory Group
IIGOP  Independent Information Governance Oversight Panel
ISAC  Independent Scientific Advisory Committee
IT  Information Technology
KIS  Key Information Summary
LSOA  Lower Layer Super Output Area
MHRA  Medicine and Healthcare products Regulatory Agency
MINAP  Myocardial Ischemia National Audit Project
NASP  National Application Service Provider
NIGB  National Information Governance Board
NIHR  National Institute of Health Research
NICE  National Institute for Health and Care Excellence
NCIN  National Cancer Intelligence Network
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NHS</td>
<td>National Health Service</td>
</tr>
<tr>
<td>NISP</td>
<td>National Infrastructure Services Service Providers</td>
</tr>
<tr>
<td>NPfIT</td>
<td>National Programme for IT</td>
</tr>
<tr>
<td>ONS</td>
<td>Office of National Statistics</td>
</tr>
<tr>
<td>PACT</td>
<td>Prescribing Analysis and Cost</td>
</tr>
<tr>
<td>PALS</td>
<td>Patient Advice and Liaison Service</td>
</tr>
<tr>
<td>PAS</td>
<td>Patients Administrative System</td>
</tr>
<tr>
<td>PCMD</td>
<td>Primary Care Mortality Database</td>
</tr>
<tr>
<td>PCOs</td>
<td>Primary Care Organisations</td>
</tr>
<tr>
<td>PCTs</td>
<td>Primary Care Trusts</td>
</tr>
<tr>
<td>PDS</td>
<td>Personal Demographics Service</td>
</tr>
<tr>
<td>QMAS</td>
<td>Quality Management Analysis System</td>
</tr>
<tr>
<td>QOF</td>
<td>Quality and Outcomes Framework</td>
</tr>
<tr>
<td>QPID</td>
<td>Quality Prevalence and Indicator Database</td>
</tr>
<tr>
<td>RBAC</td>
<td>Role-Based Access Control</td>
</tr>
<tr>
<td>RCGP</td>
<td>Royal College of General Practitioners</td>
</tr>
<tr>
<td>RCP</td>
<td>Royal College of Physicians</td>
</tr>
<tr>
<td>SCD</td>
<td>Sickle Cell Disease</td>
</tr>
<tr>
<td>SCI</td>
<td>Science Citation Index</td>
</tr>
<tr>
<td>SCR</td>
<td>Summary Care Record</td>
</tr>
<tr>
<td>SUS</td>
<td>Secondary Uses Services</td>
</tr>
<tr>
<td>StHA</td>
<td>Strategic Health Authority</td>
</tr>
<tr>
<td>THIN</td>
<td>The Health Improvement Network</td>
</tr>
<tr>
<td>TPP</td>
<td>A healthcare software company</td>
</tr>
<tr>
<td>UCL</td>
<td>University College London</td>
</tr>
<tr>
<td>VAMP</td>
<td>Value Added Medical Products</td>
</tr>
</tbody>
</table>
CHAPTER 1 - GENERAL BACKGROUND

Electronic patient records (EPRs) have been the subject of debate for many years, and although they have many benefits, they also present some challenges. Researchers have tried to highlight the benefits that EPRs bring to research and generally to healthcare, but the limitations of the existing databases that hold patient records have not allowed the full benefits of EPRs to be realised. National Health Service (NHS) England has been trying for some time to address the limitations of individual databases by moving EPRs into a central national database, which would make access to patient data easier for secondary use. Meanwhile, regardless of the limitations that exist in single databases, such as Clinical Practice Research Datalink (CPRD), researchers continue to use data stored in such databases for research and are constantly reviewing their impact.

Examples include:

- Chen et al.\(^{96}\) looked at the academic impact of a public electronic health database, as described in Section 8.1.2 of this thesis, and performed a bibliometric analysis of studies that had used the General Practice Research database (GPRD, now CPRD), to assess the influence of a single electronic health database in research. They found that a public electronic health database, such as CPRD, could promote scientific production in many ways.

- Research was also carried out by Kousoulis et al.\(^{98}\) on the CPRD and the database of the Royal College of General Practitioners (RCGP), and again is described in Section 0. The authors reported that data from CPRD had been used to produce close to 2000 research reports which had been published in peer-reviewed journals.

1.1 Identified Research Gap

Research studies, like those mentioned above conducted by Chen et al.\(^{96}\) and Kousoulis et al.\(^{98}\) show that data held in primary care databases has advantages and is benefitting research. The fact that the Hospital Episode Statistics (HES) database, similarly detailed in Section 8.1.2 also has the capability to be linked to other databases, is allowing more robust data to be extracted and used for research purposes.

Work continues to be carried out using individual databases, such as CPRD and HES, in order to assess their impact on healthcare and research; however there is a gap in the existing
available research, as no work has yet been done on the combined impact of large English healthcare databases over time. In addition, there has been no comparison made between these large databases.

1.2 Research Plan
Based on the research gap identified, publications from four large English databases were evaluated. These databases were as follows:

**Primary Care Databases:**
- Clinical Practice Research Datalink (CPRD)
- The Health Improvement Network (THIN)
- QResearch

**Secondary Care Databases:**
- Hospital episode statistics (HES)

Publications that have used data from CPRD, THIN and QResearch for research purposes over a 9-year period (2004 to 2013) were evaluated, whilst HES was evaluated for its use in published research over a 19-year period (1996-2013). The justification for looking at a 9-year period for the primary care databases was to ensure that all data, from as early as possible, were captured for all databases, and to be able to capture all data entry that had occurred in later years. However, CPRD was also evaluated individually, looking at publications that used CPRD data from 1993-2013. All publications found using HES data up to 2013, except for reports, were evaluated.

A bibliometric analysis was undertaken to help answer the research question, which is stated below.

1.3 Aims and Objectives
The main research question that this project aims to address is whether the use of EPRs can advance clinical, epidemiological and health services research.

In order to achieve this aim, the study has the following objectives:
To survey the existing literature on the benefits of using patient electronic data to support research, in order to understand the progress made so far.

To investigate the reasons why there are barriers to advancing the use of electronic databases in the UK, even though they have considerable potential.

To conduct a bibliometric review to analyse the combined research outputs and longitudinal growth in the number of publications using data from the three primary care databases (CPRD, QResearch and THIN) and a secondary care database (HES), in order to demonstrate the benefits of EPRs in research and the progress that has been made over time.

1.4 Outline of the Thesis

This thesis consists of eight chapters, with Chapter 1 providing some background information to the thesis and its outline.

Chapter 2 presents the background to what has taken place in the UK so far in terms of EPRs, what initiatives have been explored in the past, and the present initiatives or plans to move the UK to a single national database for the improvement of research and healthcare in general.

Chapter 3 provides a summary of a literature review on the current data sources in the UK that presently store patient information, how they are used, and the benefits that they have brought to research. Although these systems are not interoperable, have limitations and have had to be linked to another database in order to obtain more robust data, researchers are using them to progress their work.

Chapter 4 highlights the challenges that are associated with the successful implementation of EPRs and current views held by the public and patients on how confidentiality is protected. It also summarises the challenges that researchers face in interpreting the data protection laws when it comes to conducting research.

Chapter 5 briefly looks at EPR initiatives undertaken in other parts of Europe and how England compares in its advances towards EPRs. The chapter aims simply to provide a brief overview of the situation regarding EPRs in Europe, as this thesis is focused on EPRs in England.
Chapter 6 outlines the rationale and methods employed for the investigation into the use of EPRs for research, and describes the bibliometric analysis that was conducted on research outputs from four main databases (CPRD, THIN, QResearch & HES).

Chapter 7 details the findings and results from the bibliometric analysis, revealing research growth over time and the speciality areas that have made advances through the use of outputs from the four databases.

Chapter 8 presents discussion, recommendations and conclusions of the study findings from the literature review and the bibliometric analysis performed. It also describes the limitations of the research and potential improvements when undertaking future work.
CHAPTER 2 - INTRODUCTION

The implementation of EPRs has met many challenges, not because they are not useful or beneficial but because there remain some uncertainties regarding privacy breaches and confidentiality issues, which are a concern for patients and the general public. There are also some limitations to the EPRs which currently exist in the NHS.

The need to meet the challenges in improving the quality of healthcare, the ability to continue to find new and improved treatments, to prevent disease, and provide information on risk factors affecting health is a continuous challenge for the NHS and for health authorities globally. In recent years, there have been considerable advances in the computerisation of medical records helping to support clinical trials and safety monitoring, thus providing the NHS with greater understanding of the nature of diseases, and the effectiveness of drugs and medical devices.

Advancing EPRs and trying to ensure that they become the only way to store patients’ health records has become inevitable, due to the limitations associated with paper-based records, which are viewed as a poor tool for optimising patient care delivery and clinical activity. There are various reasons why paper records can no longer support advances in healthcare, including the fact that such records are often very difficult to read because the spaces available on medical forms to write in are often limited. Illegibility leads to errors and sometimes rework, for example, repeat tests. It is also easy for hand-written notes to get lost in transit, and they are not easily accessible: for paper records to be shared between physicians they must be retrieved from where they are stored, and then sent via the post, email or scanning. This is a time-consuming process that can often impact on a patient’s health. As people become more mobile and are treated by more than one physician, allied health professional or organisation, it becomes increasingly difficult to move paper medical records to the various points of care, therefore limiting the level of care that can be provided. The public is also keen to know that all their information will be available in one place should it be required during an emergency, so that in a situation where they are incapacitated, the treating physician has all the information needed to make the best decisions regarding their care. According to a 2006 Harris Interactive Poll for the Wall Street Journal Online, 55% of adults thought an electronic health record would decrease medical errors, 60%
thought an EPR would reduce healthcare costs, and 54% thought that the use of an EPR would influence their decision about selecting a personal physician.¹

There are also financial implications concerning the maintenance of paper records; these may appear to be a cheaper option compared to electronic records, but hidden costs, like manpower needed to retrieve records and ensure that they are accurately stored and filed, repeat testing due to illegibility or missing records, secure sending of records to other physicians, all add up. Implementing electronic records may initially be expensive, but the long-term cost savings will be significant. The confidentiality of EPRs may be seen as highly risky because of computer hackers, yet there is also a high chance of an unknown breach of confidentiality with paper records. Paper records are supposed to be held in securely locked cabinets, which only authorised personnel can access, but it is not impossible for charts or records to be left in an insecure location and become accessible to unauthorised personnel, thus breaching confidentiality.

If data continues to be stored in the paper format there will be a huge impact on research. As the UK is a major site for biomedical research and attracts a lot of interest it has become increasingly expensive and slow to set up clinical research studies in the UK. Improving and enhancing the healthcare system through the use of electronic records might allow a breakthrough in research. In order to make evidence based decisions, clinicians need high quality data that should be derived from multiple sources: inpatient and outpatient care, acute and chronic care settings, urban and rural care, and populations at risk, which can only be accomplished through electronic health records and discrete, structured data.¹ Moving healthcare to the next level cannot be sustained using paper records, and although many electronic health record systems now exist in the UK, there is still some ground to cover in order to completely realise the full benefits of electronic records, to enhance patient care and to move research to the next level in the UK.

The shortfalls of paper based health records have led to the growing recognition of the potential benefits of electronic health records for improving healthcare quality and efficiency. Electronic records can also facilitate research; for example, the use of patient records in electronic format for pre-screening and better feasibility assessment could increase the success of a trial, save costs and potentially make the UK a more attractive place to carry out research, thus providing both healthcare and economic benefits. Consequently, the NHS is
keen to adopt the use of EPRs to take its healthcare services to the next level through the use of new technology.

2.1 Definition of Electronic Patient Records

‘Wherever (and whenever) a patient is treated, there is a record of that treatment. Using information and communication technologies, these records can be made safer and available for other health professionals. These organisational records will become the Electronic Patient Records, and a subset of them will contribute to a lifelong record of a patient’s health and healthcare - the Electronic Health Record.’

One of the most important features of an EPR is that health information can be created, managed and shared by authorised healthcare professionals or providers in a digital format. EPRs could improve quality and convenience of patient healthcare, allow for quick and reliable prescribing, and allow for greater accuracy in diagnosis, potentially saving costs and improving healthcare efficiency.

In addition to healthcare professionals benefitting from EPRs, patients will also benefit by being able to see or access some details of their health records, such as test results. Patients will also be able to benefit from the use of data from EPRs in producing NHS performance statistics.

2.2 Secondary Uses of EPRs and Associated Benefits

Most data collected during the provision of care is collected primarily to support and improve individual patient care, but such data is also extremely valuable for other purposes that support healthcare delivery. In recent years there have been considerable advances in the computerisation of patient records and access to data contained in these records for research purposes. Patients also benefit by receiving closer medical attention, follow-up and continuity of care. Early stage research is also known to benefit the NHS overall, because it maintains the UK at the forefront of research. The 2014 Wellcome Trust report entitled ‘Medical Research: What’s it worth?’ makes it clear that the rate of return derived from the UK public and charity-funded biomedical and healthcare research is substantial. The conclusion of the report was that there is a sizeable return, especially in three disease areas (cancer, cardiovascular disease and mental health) where there is high morbidity. However, meeting any research target requires careful consideration of how the data available should be handled.
Having patient data stored electronically allows easy transfer, which can significantly reduce errors provided that the information entered is accurate. Access to EPRs provides professionals with information rapidly and makes it easier for them to treat a patient. In addition, electronic records enable a better understanding of the medical history of patients and present the opportunity to provide the best possible care.

The use of secondary data from EPRs to enhance research will undoubtedly offer great opportunities. However, it is ethical to also suggest that the benefits of using such data for secondary purposes must not outweigh the risk of records being used in a manner that does not respect the rights of the data owner and protect their well-being. As patient information is highly sensitive, patient groups and the general public must be able to trust that this data is handled appropriately and be willing to provide consent for their data to be used for research. Although medical research is well-established in the UK, patients remain very nervous about the threat to their privacy, and do not trust that their data is always well protected or that anonymisation really does exist. Patients are concerned that researchers will be given access to datasets containing identifiers which are not necessarily required for their work, especially when done without their consent. However, rapid changes in the law, technology, and society are reshaping the way identifiable information about patients is handled. In Britain, doctors are obliged to maintain confidentiality regarding their patients, and standards for transparency and confidentiality have been set for researchers by the Department of Health (DoH), but these appear unclear and are open to interpretation by researchers.

2.3 NHS National Programme for Information Technology

In 2005 a new commitment was announced by the Chancellor of the Exchequer to develop the capability of the NHS national information technology (IT) system via the NHS Connecting for Health (CfH) Agency, which replaced the former NHS Information Authority. The aim of the NHS CfH Agency was to maintain and develop the NHS national IT infrastructure, and it was commissioned to deliver the NHS National Programme for IT (NPfIT), an initiative by the DoH in England to move the NHS towards a single, centrally-mandated electronic care record system for patients and to ensure that data could be exchanged between General Practitioners (GPs) and hospitals. The data provided was to be secure and accessed only by healthcare professionals who were authorised to do so. According to Trisha Greenhalgh, the project was one of the largest civil computing projects
in the world and aimed to connect thousands of healthcare professionals at any NHS location through an integrated IT infrastructure for all NHS organisations. It also recognised the potential benefits of research in developing the NHS Care Record Service (NHS CRS) to assist with research on health and disease. The NHS CfH aimed to create electronic records for 50 million patients, as well as creating summary care records (SCRs) to facilitate safer and more integrated care for patients, whether seen in general practice, hospitals or other community settings. As part of the programme, simulation exercises were commissioned by the UK Clinical Research Collaboration (CRC) advisory group to the NHS CfH, in order to investigate the feasibility of using the NHS CRS as a platform for healthcare research.

2.4 Commissioned Simulations

2.4.1 Surveillance

The team’s vision for surveillance was one that could actively track patients’ responses to medical interventions, as well as devices and diseases, or any patient safety incidents that require reporting. The team concluded that the NHS CRS could be used to support an advanced pharmaco-vigilance and disease surveillance system; however, there remain many improvements that could be made to the already existing data within the Secondary Uses Service (SUS) and the NHS CRS. The data as it stood did not meet all the requirements for surveillance.

- Complete historical data will be needed for current drug safety research to help with the recognition of adverse events and to limit new ones.
- Data used for research should be effectively anonymised but there must be a way to link the data back to a patient via their physician, or an honest broker, for validation of the data or follow up of adverse events.
- Data from hospitals and day care needs to be more available, and an NHS-wide standard incident reporting system will be required for adverse events.
- Adequate codes are needed to distinguish the types of incidents.

The team believed that implementation of the above recommendations would lead to improved patient safety with appropriate level of risks and benefits due to earlier detection of adverse events, improved delivery of care due to more real-time feedback of potential risks, and more public confidence in the healthcare they receive. The team also believed that there
was an opportunity for the NHS CfH and the SUS to deliver a nationwide system that would move the necessary data from all parts of the NHS to expert surveillance systems.5

2.4.2 Interventional Clinical Trials

The purpose of this simulation was to explore the capability of the NHS CRS infrastructure to support clinical trials. The data used for this assessment was not from the NHS CRS, because it did not contain any clinical trial data, but instead was from the assessment of a clinical trial protocol, the ADOPT trial, a Diabetes Outcomes Progression Trial, a pragmatic, real-life, double-blind, randomised multicentre study which was evaluating specifically whether thiazolidinediones, with their different mode of action, can provide more durable glycaemia control than metformin or sulphonylureas, and a protocol assessment was used for this simulation. The trial was conducted in a primary care setting, with the assessment based upon data from the GPRD (now CPRD).5 The simulation suggested that comprehensive data covering a wide population in the UK was needed to allow a broad representation of the population.

- Data across all care settings needs to be accurate and up to date. There will also need to be a way to ensure that data is entered accurately and can be linked at the individual patient level.
- There needs to be strict guidance on how data can be accessed. Researchers should be able to access anonymised data but identifiable data will need a governance mechanism with tighter access controls.
- Established standards are needed when it comes to recording, processing and transferring data.
- Data would also need to be easily supplied from sources when required. The team were confident that the NHS CRS had the potential to deliver significant value for the future of clinical research in the UK, but there remained major hurdles which needed to be overcome. They recommended merging already existing datasets, such as the GPRD, with data from the HES database.

2.4.3 Prospective Tracking of Identified Cohorts of Patients

The UK Biobank is a large study of 500,000 people aged between 40-69 years, aiming to help tackle the diseases of the 21st century, such as cancer, heart disease, stroke, diabetes and dementia.5 The simulation team investigated the extent to which the SUS is providing extra
data that are relevant to the UK Biobank’s objectives, which are to study 21st century
diseases, identify markers of these diseases and their progression, and to identify lifestyle
factors that can influence the onset of a disease or if an individual is pre-disposed.

- The team concluded that for Biobank to be able to provide accurate benefits
  concerning the assessments performed, access to data will need to be at the patient
  level and must be coded as well as textual.
- There must be clear guidelines on how data will be used and these guidelines must be
  clearly defined between the provider and the receiver.
- The security arrangements need to sufficient to protect the data and confidentiality of
  the patients. Security guidelines or processes need to define exactly how security will
  be routinely audited.
- Cohort management will also need to be implemented or planned for in order to help
  maintain and update the permissions required to access patient record data.
- The NHS CfH and UK CRC also recommended that the UK Biobank must be able to
  access complete medical records of all consented participants. The information will
  primarily come from the NHS CRS but will include data from other sources,
  including ‘sealed envelopes’, which is a method whereby sensitive information about
  a patient is protected and only accessible with their consent.

The SUS are providing data that is relevant or required by UK Biobank but there remain
many limitations associated with this data.

2.4.4 Observational Epidemiological Research

The simulation team investigated how the NHS CRS and SUS are able to support
observational epidemiological research. The simulation showed that there is a need to be able
to combine data from different sources at the individual level using a unique identifier, such
as an NHS number. To allow high quality and valid studies to be conducted, there must be
the ability to link data sources at the individual level with access to patient data.

The team recommended that there must be ways to link anonymised and pseudonymised data
back to real data for checking and validation purposes. Information, such as primary care
details, laboratory results, and disease management, must be added to SUS data. Data on
private sector treatment and outcomes within the NHS also need to be added. The team
recommended that there must continuous processes to improve the quality and completeness
of the data, and that this might require an innovative approach. The creation and maintenance of disease registers needs to be explored, as these would need to capture key clinical diagnoses and outcomes of interest.

2.4.5 Outcomes of the Simulations

More extensive data is required to promote research than that currently available through the SUS. The NPfIT embarked on a 10-year project to end in 2014 that was supposed to provide comprehensive secure EPRs and eventually integrate several systems within the healthcare system, including SCRs, the electronic transmission of prescriptions, Choose & Book, and the Health Space web service. It was estimated that this would cost around £18 billion to implement; however, several components of the proposed system fell behind schedule and therefore faced criticisms.

One of the criticisms the programme faced was the safeguards promised by the NCfH, which the British Medical Association (BMA) did not feel were credible. The plans for protecting patient information, privacy or even allowing them to opt out were flawed, and differed from the initial proposal. The BMA’s review report published in 2006 showed that all a patient’s information uploaded via the national database would eventually become accessible to many NHS staff and therefore encouraged GPs to facilitate their patients opting out.5 Public concerns about the centralisation of health data increased, and in November 2006 a poll revealed that 53% of patients opposed a central medical records database with no right to opt out. At the same time, a report for the Information Commissioner described Government plans to widely share health information on children with other services, including social services, schoolteachers and the police, and concluded that the proposed measures were both unsafe and illegal.6

In September 2007 the House of Commons Health Committee called for more information to be published on the proposed design, and for data placed in ‘sealed envelopes’ to be withheld from the SUS, a suggestion that the DoH rejected. Another independent review of the intended system chaired by Dr. Glyn Hayes took place in August 2009 and this suggested that the cost of the programme was disproportionate to the intended benefits of the system.7 Professional organisations and activist groups also expressed concerns that the implementation of the proposed electronic health record system would threaten medical privacy and potentially other human rights.6 Criticisms of the programme included the fact that its central direction and political drive was bound to hinder any advanced progress.
Another criticism was that the scale and proposed content of an electronic health record would almost certainly threaten medical privacy, and potentially other human rights. Although these two critical strands are to some extent contradictory (if the National Programme is setting back progress in electronic health records, the privacy lobby should have little to fear), they have a common origin in a perception that the National Programme was generally pushed by people who lacked knowledge of the NHS itself and the needs of patients.

The NHS launched a major campaign in 2009 announcing changes to the way health records were stored and accessed, and the public was informed about SCRs, which would contain a patient’s name, address, date of birth and NHS number, details of medicines taken currently and in the past, allergies and any health conditions, such as asthma or heart problems, notes on diagnoses, treatments or operations, and plans or reminders for care in the future.

Patients and the general public had many negative reactions concerning the type of information being stored in a SCR, which led to the coalition government setting up two reviews of SCRs. These reviews led to the suspension of some parts of the patient information programme but SCRs went ahead; however, content was limited to core information and an opt-out form would be included in patient information packs. The review of the content of the records, led by NHS Medical Director Sir Bruce Keogh, concluded that a core record should only contain a patient’s demographic details, medications, allergies and adverse reactions, and that these should continue to be copied from the GP’s medical records. The review said that the DoH should only consider expanding the content of a SCR when sufficient trust had been built into the system and when patients requested that it should do so. They also recommended that there should be considerable patient involvement in deciding the evolution of patient records and the option to opt out of having a SCR should be made very clear to patients. Sir Bruce Keogh concluded that it was not unreasonable for citizens to expect that, in an advanced NHS, clinicians treating them in an emergency situation would have access to basic medical information that would allow them to make an appropriate decision. Implementation of SCRs would go ahead with a clearly defined minimal scope and any additional information would be added only after patients had consented to it.
2.5 Attempt to Implement the NHS Connecting for Health Plan

The NHS NPfIT decided to progress with the plan to implement a single national database for electronic care records for patients. The aim was to connect thousands of general practitioners and hundreds of hospitals, providing secure and audited access to these records by authorised health professionals. The NHS NPfIT was to be the largest public sector IT programme ever attempted in the UK. It was originally expected to cost £2.3 billion over three years, but in June 2006 the total cost was estimated by the National Audit Office to be £12.4bn over 10 years.9

The aim of the NPfIT was to introduce integrated EPR systems through the Spine system. The Spine is a collection of national applications, services and directories that support the NHS in the exchange of information across national and local NHS systems. The Spine connects clinicians, patients and local service providers throughout England to essential national services, including the electronic prescription service, SCR, Choose and Book, and demographics services.9

2.5.1 Personal Demographics Service

Personal demographics service (PDS) stores demographic information about each patient and their NHS number. Patients are not able to opt out of this part of the spine but can ask for their records to be marked as private, meaning that they cannot be viewed by NHS staff.9

2.5.2 Summary Care Record

The SCR is a summary of a patient's clinical information, such as major diagnoses, medication, allergies and adverse reactions to drugs.9

2.5.3 Choose and Book

This system enables patients needing an outpatient appointment to choose which hospital they are referred to by their GP, and to book a convenient date and time for their appointment.9
2.5.4 Electronic Prescription Service

This service enables prescribers, such as GPs and practice nurses, to send prescriptions electronically to a dispenser (such as a pharmacy) of the patient's choice. This makes the prescribing and dispensing process more efficient and convenient for patients and staff.9

2.5.5 Secondary Uses Service

The SUS uses data from patient records to provide anonymised and pseudonymised business reports and statistics for research, planning and public health delivery.10

Some parts of the National Programme were delivered successfully, but other parts encountered significant challenges, especially the deployment of the CRS. The costs of the project were also very uncertain and increased over time. Following three reports by the National Audit Office, the Government announced in September 2011 that it would not go ahead with the project which would be dismantled, although those components that had already been implemented would be kept, including SCRs, Choose and Book, and the electronic prescription service.

2.6 Reasons for the NHS Connecting for Health Failure

The failure of the NHS CfH can be attributed to many reasons, such as bad planning and a lack of buy-in by the public and healthcare professionals. The project was also said to be very politically driven, which may have contributed to its failure. Some of the major reasons given for the failure are discussed below.

2.6.1 Confidentiality Issues

The public has always been sceptical about the centralisation of health data, and there were concerns about the lack of communication regarding the main purpose of sharing information. GPs themselves did not seem to believe that confidentiality would not be an issue and wanted consent procedures to be put in place to allow patients to decide whether their details should be uploaded to a central database. The consent procedures for SCRs were found to be completely unsatisfactory, and the proposed plan would mean that patients might be unaware that certain records were intended for sharing. Patients could ideally only opt out of their information being shared if they were aware of its existence, but expecting patients to
be aware and to opt out every time they interacted with the NHS was unrealistic. Anderson et al. explored the question of just how consent would be obtained from an intoxicated teenager who has presented requesting emergency contraception. In addition, the option to opt out was not extended to children. The Government informed the general public that all households would receive leaflets informing them of the opt-out options, but it was reported that many did not receive the leaflet, and the opt-out campaign group reported that patients were not being informed of their rights.

2.6.2 Procurement and Cost Issues

The central funding for the National Programme announced in 2003 that £400m would be available in financial year 2003/04, £700m in 2004/05 and £1.2bn in 2005/06. Service providers that were engaged to carry out the requirements of the NHS IT programme were split into three groups:

- Local service providers, who were expected to provide a full range of IT services within a specified locality or a single Strategic Health Authority (StHA).
- National Infrastructure Service Providers (NISP) were contracted to facilitate the infrastructure required across the NHS to support improved broadband capacity and connectivity for healthcare professionals and support staff.
- National Application Service Providers (NASP) were contracted to deliver discrete applications services. This included key national applications for electronic appointment bookings and the electronic transfer and processing of prescriptions. It also included discrete application services around the integrated CRS, including a data spine or summary record, and access services for authorised users.

The three different types of providers were expected to work together towards the common goal of delivering the integrated services. However, there were many criticisms concerning the methods used to procure services. The Head of the Programme, Richard Granger, set in place principles that stressed speed, completion and payment by results, and the contracts created were very large and were put into place quickly, with negotiations not narrowed down to preferred suppliers but instead leaving bidders going head-to-head against each other until the very last minute. Alistair Maughan detailed six reasons why the NHS NPfIT failed, including the use of heavy handed tactics which were used to ram through contract terms which were considerably harsher than had ever before been seen within a government or even private sector context. The fact that everything was done quickly meant that some appropriate
contract terms were agreed retrospectively because appropriate due diligence was not put in place. Many vendors also withdrew and declined to continue with contracts because of the complexities involved. The hasty procurement method, amongst other reasons, meant that fine details were ignored, vendors who could actually deliver were not awarded contracts, and healthcare professionals were not engaged as well as they could have been. Right from the onset, the programme lacked clinical engagement, with the focus being on technology and not on health service change.

The complex issues with vendors and procurement had a direct effect on cost, and it appears that the actual cost of the project was never known and kept changing over time. The DoH initially set out with a negotiated cost of around £6.4bn to deliver care records systems to 220 trusts in the north and eastern England and also the Midlands, but the final contract for the project likely cost the DoH £9.8bn and covered only 22 trusts. By the time the decision was made to dismantle the project, costs had escalated to nearly £10bn, and the cancellation of vendor contracts, such as that with Fujitsu, cost the DoH a further £3.1m in legal fees. Many factors contributed to the failure of the NHS NPfIT, but the lack of clinical engagement was key. Many healthcare professionals did not buy into the implementation of the system and thought it was over-ambitious, whilst many trusts believed that a centralised system would not suit their needs.

Sir David Nicholson, Chief Executive of the NHS from 2011-2014 said that:

“A modernised NHS needs information systems that are driven by what patients and clinicians want. The NPfIT has provided us with a foundation but we now need to move on if we are going to achieve the efficiency and effectiveness required in today’s health service. Restoring local control over decision-making and enabling greater choice for NHS organisations is very crucial as we continue to use the secure exchange of information to drive up quality and safety.”

Many lessons have been learnt by the DoH concerning the failure of the NHS IT programme, and in the future, a clear plan, sound cost estimates, realistic timelines, better procurement strategies, and above all public and healthcare professional engagement, will need to be worked on to deliver a paperless NHS and reform the way that the NHS in England uses information.
2.6.3 Political Reasons

CfH was created because of strong political support within the Government and direct leadership from the then Prime Minister. The Prime Minister pushed hard for a severely curtailed implementation timetable for NPfIT in order to have something tangible to show voters by the next general election, which was due in 2005.18 As the agenda was politically driven, there was little or no consultation with key stakeholders concerning the direction decision and failure to facilitate clinical engagement was the most significant major cause for its downfall.

2.7 care.data - Sharing Information

The Government continued to aim to modernise health services in order to improve health outcomes and make appropriate decisions for patients. This goal, without doubt, requires high quality information that is easily accessible by NHS organisations and researchers. The Government hoped to achieve this through the care.data programme, which was to build on existing data services and expand them to provide linked data. It was planned to eventually cover all care settings, both in and outside of hospital. Changes made to the Health and Social Care Act 2012 empowered the Health and Social Care Information Centre (HSCIC) to collect and share confidential information from medical records.19 HSCIC can, under certain circumstances, acquire personal and confidential data from GP practices without seeking patient consent.19

The care.data service was to be used by HSCIC to extract data from medical records to support better research and provide timely information within the NHS to improve patient healthcare. The RCGP and the BMA issued a joint statement with the HSCIC that clearly outlined the benefits of care.data and the need for greater transparency to improve the quality of patient care.20 The NHS began to create awareness of this initiative at the start of 2014, and all households across the UK should have received a leaflet entitled ‘Better information means better care’. The NHS planned the first extraction from GP practices to commence in March 2014 and GP practices were encouraged to hand out and display posters about the initiative. Patients were to have a minimum of four weeks from the time they received the leaflet to register any objections before the first extraction commenced. However, many critics believed that there was not sufficient public consultation over the implementation of care.data and the failure to include an opt-out form with the ‘Better information means better care’ leaflet was very disappointing.
The distribution of the generic leaflets was reported to have failed to reach many households and the BBC Health Correspondent, Nick Triggle, reported that households who had received the leaflet often found it buried in their junk mail. The project suffered from a significant amount of negative press even before its go-live date, and the benefits that the project could offer were dampened by ineffective communication. In addition to the concerns of little or no public consultation, privacy was also a primary concern, and many members of the public believed that there would be a breach of privacy laws and their human rights would ultimately be abused. The Secretary of State for Health stated that he was committed to respecting individuals’ objections in relation to the disclosure of their records, except in cases such as in the event of a civil emergency. A patient could inform their GP of their wishes to object to their records been used without reason, and the right to object was implemented as a constitutional right, rather than a legal right. GPs were concerned that the implementation of care.data would harm patient-doctor confidentiality, as patients might start to withhold information from their doctors. However, researchers believe that it is time to move forward and really embrace the benefits of having health data readily available in one place, which will improve health outcomes. Sheather and Brennan believe that the potential locked in GP medical records is extensive and needs to be exploited.

2.7.1 Benefits and the Risks Associated with care.data

The NHS continues to believe that the benefits of care.data are invaluable and that it will no doubt improve healthcare and provide guidance on how the NHS is managed. The NHS has listed the benefits of care.data which are described below.

2.7.1.1 Summary of predicted benefits

- Improved monitoring of outcomes: with the linkage of primary care and secondary care data, diagnosis will be made easier and will help identify the number of patients recorded on the quality and outcomes framework (QOF) disease register.

- Improved monitoring of trends and data quality: it will be easier to understand the trends in A&E attendance, unplanned admissions, and reasons for readmissions.

- NHS payments will be accurate because the services provided will be easily traceable.

In addition, some researchers have expressed belief in its full potential and stated that there is no doubt that the successful implementation of care.data will move healthcare forward. In an article on patient confidentiality, Sheather and Brennan note that, despite the concerns
raised, as real as they are, there are enormous potential benefits for individuals, the proper functioning of the health service, and the development of future healthcare and treatments contained in the data.

2.7.1.2 Summary of predicted risks

Whilst there are great benefits associated with the implementation of care.data, there are also several risks that make patients and the general public nervous.

- Impact on patient-physician confidentiality: patients may no longer want to share personal details with their doctors because they would not trust that the data would not be shared and be used for what it was not originally intended for.
- There are also concerns that the database may be too difficult to manage.
- Loss of data from one large database: the risk of significant loss of data from one large database is greater than from many databases/GP surgery computers.

The potential benefits of implementing a database such as care.data are great, and can indeed benefit the public in the ways described by the NHS, but public education about its benefits, the guaranteed strategies on how confidentiality will be maintained, as well as full transparency on the part of the NHS about how the data will be used, remains a significant concern to the public and has led to strong protests. In addition to public concerns, there are privacy campaign groups that oppose care.data; the main such group is MedConfidential, which campaigns for confidentiality and consent in health and social care. MedConfidential was established by organisations that include Privacy International, Big Brother Watch, No2ID, the Foundation for Information Policy, and the Big Opt Out. It is an independent, non-partisan organisation which works with patients and health professionals, drawing advice from a network of experts in the fields of health informatics, computer security, law/ethics and privacy. MedConfidential believes there must be no conflict between good research, good ethics and good medical care.

2.7.2 Abandonment of the Project

The plan to start data extraction in March 2014 was abandoned on 18th February 2014, when the NHS with the support of the BMA announced that it would postpone care.data. The NHS believed that a delay would allow time to build understanding of the benefits of care.data and the advantages it can bring to healthcare at large, something that it failed to achieve before informing the public that it would go live. The NHS plans to spend time explaining exactly
what safeguards are in place and how people can opt out if they wish. In addition to the concerns of the privacy campaigners and the general public, the NHS also heard concerns from various bodies, such as the RCGP, the BMA and the patient watchdog Health Watch England, and has since put together an independent care.data advisory board to improve governance surrounding the care.data programme. This advisory group has commenced work on ensuring that communication on care.data is effective, as well as ensuring that the concerns of the general public are addressed. HSCIC recently published a ‘six-month extension work stream’ outlining the steps that must be taken and the associated dependencies before data can be extracted and/or disseminated.26,27

This work stream will look into the provision of additional guidance and materials for general practice, and will look to increase awareness amongst GPs of the programme, working towards the publication by the HSCIC of a code of practice for confidentiality and legislative changes set out by the DoH. HSCIC also stated that the collection of data will be accomplished in three stages and a project had been established to deliver a technology platform that will process primary care data in such a way that it ensures the public and HSCIC have confidence in the security of the data. This enhanced technical platform will act as a foundation for the development of other services that the HSCIC will provide in addition to care.data.

2.8 General Practice Extraction Services

The NHS extracts data through the General Practice Extraction Service (GPES), a centrally managed service that will extract information from general clinical IT systems for a wide range of purposes.31 This is the only national service for extracting data from patient records held in general practice systems across England and is owned and managed by HSCIC. HSCIC requires this data in order to link information provided by GP practices to personal and confidential data obtained from other settings, such as hospitals, in order to analyse it effectively.

GP practices do not have the choice to opt out from these data extractions, and HSCIC is empowered by NHS England under the Health and Social Care Act 2012 to collect data from GPs, who are legally obliged to comply with the requirement. Alternative methods of providing the data may be considered by HSCIC if GPs specifically object to GPES as the method of extraction. However, patients can request that their data is not transferred by either
writing to their GP to state this or through completing an opt-out form. Read-only codes are be added to patients’ records to stop any uploading or transferring of data out of GP practices.

The only time data is extracted, even if a patient opts out, is when the law requires it. Extracted data is personal and confidential, and includes details of referrals, NHS prescriptions, and personal identifiers, such as date of birth, gender, NHS number and postcode. The data extracted is currently passed to third parties, such as NHS organisations, and in the future this will be extended to research bodies and commercial companies. Data extracted by GPES may not be anonymised but does have to be completely anonymised before it reaches any third party. All data extracted by GPES is stored, processed and transmitted securely, while access to data is role-dependent and there is an audit trail of who has accessed the system. GPES holds the data for a short period of time in order to respond to any queries from patients who have consented to have their data transferred and who wish to have a copy of the information being held about them. GPES aims to delete data as soon as it has been released to and accepted by a customer. The recipients of the data need to sign a data-sharing agreement with NHS HSCIC, stipulating that the data will be stored securely, accessed and used only for the agreed purposes and in line with the non-disclosure policies.

2.8.1 Maintaining Patient Confidentiality

As with any information system that holds confidential data, it is essential that the rights and privacy of patients are protected. The GP community has also clearly expressed that the trust between them and their patients must be maintained, consequently, governance principles were instigated in conjunction with the National Information Governance Board (NIGB), the BMA and the RCGP. Organisations requesting data need to sign up to the GPES standards and usage principles, and are subject to audits to ensure that these standards are adhered to. If an organisation is found not to adhere, their rights to access data will be withdrawn. An Independent Advisory Group (IAG), made up of GP and patient representatives, considers requests for information from third parties who wish to use GPES and recommends an appropriate course of action to the HSCIC. The IAG assesses what data is to be extracted and why the data is needed, restricting use only to the original purpose for which it was requested. In order to fulfil this role, the IAG receives and considers ‘IAG packs’ provided by HSCIC, which contains a customer requirement summary, an information governance assessment carried out by HSCIC, and a benefits plan. Patients can also choose not to participate in
GPES and the system is able to handle the exemption of a patient. No confidential data will be made public at any time.

The safeguards put in place by GPES are also subject to constant review to ensure that it continues to adopt best practices that assure GPs and the general public that GPES is secure and able to protect patient confidentiality.

## 2.9 Chapter Summary

The NHS in the UK is keen to explore the benefits of EPRs and has made a few attempts to move towards a central electronic system that will help drive research and thus improve healthcare. These initiatives (CfH and care.data) have encountered various challenges in their implementation, despite the large investments made and the associated predicted benefits. The NHS has managed to maintain some elements of the NHS CfH system that still function today, such as the SCRs.

Failure to achieve a central system occurred for various reasons, including bad planning, increased costs, bad publicity, and concerns around the risk that systems would encourage breaches in privacy.\textsuperscript{14,15,16} The NHS has learnt from these experiences and is piloting care.data in pathfinder sites, with the hope that the project will be successful in the future.
CHAPTER 3 - LITERATURE REVIEW

The previous chapter described the various initiatives that have taken place within the NHS to move towards a central national system. This has been on the agenda of the NHS for year now, and it is believed that a central system will provide benefits that will not only improve healthcare services, but will also allow research to progress. However, various challenges encountered have caused both of the big projects initiated by the NHS to be abandoned.

In this chapter a summary is provided of the various databases that currently hold patient electronic records, some of their uses, and the current limitations that exist with their use. Limitations include inaccurate data, a lack of available standards to help guide those entering data, and the fact that the current systems need to be linked in order for more robust data to be obtained. This highlights why the NHS is looking to move towards a single national central system, as it is hoped that a single system will help standardise data entry and help improve the accuracy of the data being captured.

Although the systems detailed below have their limitations, the EPRs that they contain are being used for research purposes. This chapter will look at these systems and how they are currently being used.

3.1 Primary and Secondary Care UK Databases

Attempts to implement some form of NHS healthcare database or to have an electronic healthcare record system are not new, and there are many systems already in existence that are not interoperable. Various systems, such as patient administration systems, lab and radiology systems, and hospital information support system (HISS) resource management have existed since the 1970s within the NHS. In the 1990s the NHS IT strategy began and the idea of EPRs started to emerge, and ever since then the NHS has been looking to see if every patient-based encounter within a single hospital could be linked via computer. Among the various systems already existing within primary and secondary care, there are databases that exist mainly for the secondary uses of patient data.

The largest database in secondary care is the HES database, which is a data warehouse containing details of all admissions, outpatient appointments, and A&E attendances at NHS hospitals in England, and enables the DoH to monitor activities in hospitals.
3.1.1 Primary Care Databases

In primary care there are several databases that exist to help progress research and also support clinical governance. The largest of these are CPRD, THIN, QResearch and ResearchOne.

There are several secondary uses of clinical data collected in primary care. Gnani and Majeed\textsuperscript{34} listed some of these in a report for the Association of Public Health Observatories.

3.1.2 Clinical Governance and Quality Improvement

The use of the electronic data collected has helped with clinical audits, and many primary care organisations (PCOs) have also found that this function is a useful way to drive practices to ensure that the data collected is accurate and complete. However, challenges remain concerning the way data is collected from diagnostic investigations, such as consultation times, which are normally very limited and the lack of standardised data collection.

3.1.3 Health Serving Planning, Need Assessments and Commissioning

The systems available in general practice are mostly set up to record clinical activity and to perform routine administrative tasks. There are limitations to the amount of information that these systems can collect, and information on risk factors, diseases and population sub-groups, which could be used to estimate disease burden within populations, is not being collected. This reduces the health planning that can be carried out in primary care; however, the introduction of the primary care QOF, quality management analysis systems (QMASs) and the quality prevalence and indicator database (QPID) has improved the potential for primary care data to help examine disease prevalence.

3.1.4 Regulation, Accountability and Performance Management

The NHS monitors the performance of general practices and primary care using financial administrative sources, such as general practice contracts and prescribing analysis and cost (PACT) data. Indicators are based on population health measures that are applied to general practice, but these have been found to be insufficient because they do not truly reflect the variability of data collected in primary care, especially in situations where independent information is used within practices. The use of QOF data will allow the use of more robust indicators that will help monitor the quality of care provided by general practices and PCOs.
3.1.5 Monitoring Healthcare Use

Monitoring the use of health service has historically been achieved using hospital data (hospital admissions) because general practices did not have methods or systems in place to monitor referral rates: these can help to predict the demand for specialist services, changes over time, and cost of patient care, which rises on referral to hospital. PCOs are now required to put methods in place for monitoring general practice referral rates.

3.1.6 Monitoring Death Rates

The primary care mortality database (PCMD) is an online database which has been developed to enable mortality data available to the NHS that is linked to GPs and GP practices where dead patients were registered. However, challenges may occur depending on whether official population estimates are correct, for example if a GP does not hold accurate or up-to-date information on deceased patients; details such as address and any name change must be up to date.

3.1.7 Monitoring Health Inequalities

Primary care data is a valuable source when it comes to identifying differences in health inequalities. To understand how primary care is used by different ages, ethnic or socio-economic groups, data would have to be captured by GPs. Considerable work has been put into the systematic allocation of socio-economic information, and this has been entered into medical records to allow general practices to provide robust data on health inequalities. For example, studies using primary care data have been able to show that the prevalence of coronary heart disease was highest in deprived areas and lowest in affluent areas.34

3.1.8 Pharmaco-epidemiology

Studying the use and effects of drugs on large groups of people has important implications for public health and patient safety. The data available in primary care and large databases, such as CRPD, can provide information that will allow the proactive study of any untoward effect of a particular drug on a set of patients. Long-term follow-up can easily be monitored through available information, and can help to steer safety clinical trials in the right direction.

3.1.9 Public Health and Health Services Research

Primary care data is vital to research and can be used to improve healthcare. Studies using primary care data have examined disease prevalence, treatment effectiveness, time trends,
and area and socio-economic variations, to help inform public health priorities. To conduct better control studies, rigorous monitoring of selected patients will need to take place, but this will require significant improvements to data quality, access to primary care data, range of data collected and linkage of data with other sources, as well as developing how the recording of socio-economic status at patient level occurs within primary care.

3.1.10 Financial Flows and Payment by Result

Changes in NHS financial processes mean that PCOs determine or commission the activities required. Primary care data would therefore have the potential to help develop accurate pricing for healthcare. In addition, such data would enable details to be easily accessed on services provided by different healthcare vendors for primary and hospital care.

3.1.11 Resource Allocation, Risk Adjustment and Case Mix

The NHS is working to reduce variation between the way medicine is practised in both primary and secondary care. Using risk adjustment methods can help to limit these variations when case mix data is collected, which can provide a thorough understanding of the causes of these variations. Budgets can then be adjusted for healthcare providers who manage patients with more complex medical problems.

3.1.12 Clinical Practice Research Datalink

CPRD is an observational data and intervention research service jointly funded by the NHS, the National Institute of Health Research (NIHR), and the Medicine and Healthcare Products Regulatory Agency (MHRA) and is one of the largest databases of longitudinal medical records from primary care in the world. The collection of information began in 1987 under the name GPRD until 2012, when the name was changed to CPRD. GPRD was initially part of Value Added Medical Products (VAMP), a commercial company that pioneered the design and marketing of a general practice office computer system, allowing the recording of information on individual patients.

CPRD (formerly GPRD) therefore contains nearly 30 years of longitudinal data. As of December 2014, the database contained data for over 13.5 million patients, of which approximately 5.7 million were currently active. GPs play a gatekeeper role in the healthcare system because they are responsible for primary care and for referring patients to specialists. As well as primary care data, CPRD now links to a number of other data sets,
such as HES. The database has clinical and prescription data, and can provide information to support pharmaco-vigilance, including information on demographics, medical symptoms, therapy (medicines, vaccines, devices), and treatment outcomes.\textsuperscript{38} It can be used to enhance clinical trials efficiency (protocol optimisation, feasibility and recruitment) and can provide data for both clinical and academic researchers.

Access to the data is subject to protocol approval by the MHRA Independent Scientific Advisory Committee (ISAC). Over 1,500 research reports published in peer-reviewed journals have used data from the CPRD and it has had a direct impact on public health and disease speciality areas.\textsuperscript{35}

### 3.1.13 QResearch

This is another large primary care database derived from the anonymised health records of over 12 million patients.\textsuperscript{36} The data currently comes from over 750 general practices using the EMIS clinical computer system which is widely used in the UK.\textsuperscript{36} Although the data contains socio-economic details of patients based on their postcode, it does not hold any identifiable data and access to the database is only open to academic researchers who have ethical approval to receive data. The data they receive is only what they require in order to answer their research question; they do not receive the whole dataset. QResearch has also led to many other projects, such as QFlu which was used for monitoring and tracking the prevalence of the swine flu outbreak in 2009, reporting to the Health Protection Agency.\textsuperscript{36} One of the limitations of QResearch is that, although it has links to external databases such as HES, the anonymisation process in compiling the database means that there is no way to identify patients (unlike CPRD, where patients can be contacted via their general practice).

### 3.1.14 The Health Improvement Network

The THIN database represents the collaboration between two companies; In Practice Systems Ltd (INPS), which developed the Vision software used by GPs in the UK to manage patient data, and CSD Medical Research UK (formerly known as EPIC) which provides access to the data for use in medical research.\textsuperscript{37}

THIN data collection commenced in 2003 and over 500 Vision practices have joined the scheme. THIN currently contains the electronic medical records of 11.1 million patients, including 3.7 million active patients, which represents 6.2\% of the UK population.\textsuperscript{37} In addition to the main consultation being recorded, most patients in THIN are linked to
postcode-level area-based socioeconomic, ethnicity and environmental indices. Data are based on a patient's postcode so that variables at ward level are available. This information is collected for THIN without compromising patient confidentiality in any way. A patient is identified only by a code allocated by the GP system and cannot be identified outside a practice. Research studies for publication conducted using THIN data are approved by a nationally accredited ethics committee, which has also approved the data collection scheme.

3.1.15 ResearchOne

ResearchOne is a health and care research database developed by TPP, a UK healthcare software company, in partnership with the University of Leeds and the UK Government’s Technology Strategy Board. ResearchOne consists of de-identified clinical and administrative data drawn from the EPRs currently held on the TPP system, called the ResearchOne clinical system. This has the potential to be one of the largest healthcare research databases in the world, as health and social care organisations using ResearchOne can contribute non-identifiable information for research. The data contained in ResearchOne is linked from primary and secondary settings, such as general practice, child health, community health, palliative hospitals, out-of-hours, A&E, and acute hospitals. Data from all these settings contributes to the research data set and represents a good source of geographic and demographic data across England. Practices are given the option to opt their records into ResearchOne and access is provided to researchers who have national research ethics committee approval on a not-for-profit basis. Individual patients are also able to opt out.

3.1.16 Prescribing Databases

PACT data is extensively used for monitoring general practice prescribing and setting prescribing budgets. This is a national dataset which analyses prescribing data in terms of cost, as well as a number of other parameters (volume), and at the general practice level it is also used as an educational and audit tool. Although it provides an accurate and complete record of the cost and volume of GP prescribing, it is not linked to patient or any diagnostic data and therefore cannot provide information on prescribing by age, sex, or specific disorders.

The strength of the data available in primary care is not reflected in the way that the data is used to influence policies, due to the weaknesses of the databases available and the lack of
standards to which data is recorded in primary care. Many gaps still remain in the way the databases function, with the most significant being that there are hardly any interactions between them. Primary care data is collected and stored for a great number of other reasons besides providing clinical care, but all are focused towards better healthcare.

3.1.17 Quality and Outcomes Framework

The QOF is part of the general medical services (GMS) contract for general practices and was introduced on 1 April 2004. The QOF rewards practices for the provision of ‘quality care’ and helps to fund further improvements in the delivery of clinical care.41

General practices have the option to participate, as this is voluntary, but most practices on GMS contracts, as well as many on personal medical services (PMS) contracts, take part in the QOF.

When QOF was first introduced as part of the GMS contract in 2004, the following principles were agreed concerning the QOF standards that should apply:

- Where responsibility for the ongoing management of a patient rests primarily with a GP and primary care team.
- Where there is evidence of health benefits resulting from improved primary care.
- Where the disease is a priority in a number of the four nations (England, Wales, Scotland and Northern Island)

3.1.17.1 The three components of QOF

The QOF contains three main components, known as domains:41 clinical, public health, and public health additional services. Each domain consists of a set of achievement measures, known as indicators, and practices are scored against the indicators with points being awarded based on the level of achievement. The 2013/14 QOF measured achievement against 121 indicators, and practices scored points based on achievement against each indicator, up to a maximum of 900 points, although this was recently modified by the retirement of two domains (quality and productivity, and patient experience). The 2014/15 QOF achievements are now measured against just 81 indicators, with a possible maximum score of 559 (see Table 3.1); the higher the score, the higher the financial reward for a surgery.
Financial payments are adjusted taking into account a surgery’s list size and the prevalence of chronic conditions. The financial incentive is hoped to encourage GPs to treat more patients in the community, rather than referring them to hospital for treatment.

Table 3.1: The three components of QOF

<table>
<thead>
<tr>
<th>Domain</th>
<th>Indicators</th>
<th>Points</th>
<th>Areas covered</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical Care</td>
<td>69 across 19 clinical areas</td>
<td>435</td>
<td>Coronary heart disease, heart failure, and hypertension</td>
</tr>
<tr>
<td>Public Health</td>
<td>7</td>
<td>97</td>
<td>Blood pressure, cardiovascular disease-primary prevention, obesity, and smoking</td>
</tr>
<tr>
<td>Public Health</td>
<td>5</td>
<td>27</td>
<td>Cervical screening, child health surveillance, maternity services, and contraceptive services</td>
</tr>
</tbody>
</table>

The QOF is supported using the calculated quality reporting service (CQRS) as an extraction tool. HSCIC collects QOF data and processes information about GP services in England via CQRS. The collection process involves gathering information in the form of data and copying it from one system to another. Data is extracted over a period of time and remains anonymised, and collected data is used for planning research and generating payments.

The data collected is used by

- NHS England, for measuring GP achievements under the QOF and delivery of enhanced services (ES) and other services, including vaccination programmes (collectively known as ‘quality services’).
- The DoH and other government departments, for data on GP activity related to certain conditions, such as dementia.
- Universities and academic researchers carrying out research.

HSCIC lists the use of the data for the following reasons:

- Improving screening and treatment times for serious conditions such as cancer.
- Planning services for major public health issues, like flu epidemics.
- Helping prevent people from developing diabetes and losing their sight.
- Improving services for people with learning disabilities.
- Calculating and generating payments to GPs.
3.1.18 Secondary Care Database: Hospital Episode Statistics

Hospital administrative data, such as that provided by the HES database in England, is increasingly being used for research and quality improvement. The HES database was developed in 1987 and was designed to collect a detailed record of each episode of care received by a patient, either from NHS inpatient providers or independent sector providers commissioned by the NHS to deliver care.\(^{33,42}\)

HES data is compiled from information sent by over 400 NHS trusts and foundation trusts, including acute hospital trusts, primary care trusts (PCTs) and mental health trusts in England.\(^{42}\) Independent sector organisations also send data to HES for activities commissioned by NHS England. Data is stored as a large collection of separate records, one for each episode of care, within a secure data warehouse.\(^{42}\) HSCIC liaises closely with NHS organisations to ensure that the data submitted is complete and accurate.

The HES database houses data containing patient details, such as age, sex, NHS number and location of residence, episode of care (hospital name, GP referral, emergency/elective, admission details, and clinical information, e.g. diagnosis, operative procedures).\(^{43}\) Records are collected and then submitted to SUS, which copies the information into a database and also makes it available to commissioners. SUS takes an extract from the database and sends it to HES at set times during the year. Data transferred to SUS will continue to change, but HES data remains fixed, because the data remains exactly as it was when a particular extract was taken; consequently, there are likely to be differences between analyses from SUS and from HES. The HES team validates and cleans the data extract, before deriving new items and making the information available in the data warehouse.\(^{44}\)

HES data is largely designed for non-clinical purposes, including hospital remuneration following delivery of care to patients. Some other secondary uses of the data available in HES include:

- *Monitoring trends and patterns in hospital activity.* An example of where HES was used to monitor hospital trends is a study by Tai-Yin Wu et al.\(^{45}\) who looked at ten-year trends in hospital admissions for adverse drug reactions (ADRs) in England between 1999 and 2009. The study concluded that the number of ADR admissions increased at a greater rate than the increase in total hospital admissions, and the authors suggested that some of this finding might be due to improved diagnostic coding. However, in-hospital mortality
due to ADR admissions also increased during the period. The authors believe that results from their study should help to prompt policymakers to implement further measures to reduce ADR incidents and their associated in-hospital mortality, together with methods to improve the recording of ADRs.

- **General medical research and statistical functions.** An example where data from HES was used to provide useful information to researchers was a study on colorectal resection in elderly patients, which showed that half of English patients aged over 80 years undergoing non-elective colorectal resection died within a year of surgery.46

- **Developing and monitoring government policies.** An example where HES data provided information that has helped to improve government policies was a study conducted by Hacker et al.47 looking at equity in waiting times for two surgical specialties, via a case study at a hospital in the north-west of England. The study investigated the extent to which equitable access is achieved in one routinely administered hospital waiting list system and the results showed that routine waiting list systems were not always delivered equitably. For one specialty area, female, older and deprived patients were significantly more likely to experience longer than average waits. This study provides information that supports policy makers’ work on hospital waiting times.

Another example that provided the government with important information that could influence policy was a retrospective study conducted by Milagros Ruiz48 on the impact of the doctors' strike in England on 21st June 2012. The study looked at patients admitted to hospital, patients with outpatient appointments, and A&E attendances over a three-week period (from 11th to 29th June 2012), excluding weekends and spanning the strike day. The authors were able to deduce that the 24-hour doctors’ strike in England significantly affected the provision of healthcare by NHS hospitals.

- **Providing the basis for national health indicators for clinical quality such as hospital readmission rates to examine health outcomes and to improve patient experience.** An example where HES data helped to provide national indicators on readmission rates was an observational study conducted by Aljuburi et al.49 looking at socio-economic deprivation and the risk of emergency readmission and inpatient mortality in people with sickle cell disease (SCD) in England. The study looked at SCD patients admitted between 2005/06 and the results showed that SCD patients from the most socio-
economically deprived areas and with comorbidities were at highest risk of both SCD re-
readmissions and in-hospital mortality, suggesting that there were inequalities in healthcare
access and health outcomes amongst people with SCD.

- There is so much value in the data contained in HES that it has also been used to
  influence healthcare internationally, as reported by Mohammed and Stevens.50 In 2001
  Dr Foster used HES data to produce standardised mortality ratios adjusted for case mix
  using the methodology of Jarman et al. in the study Explaining differences in English
  hospital death rates using routinely collected data which was published in the British
  Medical Journal (1999; 318 (7197):1515-1520). The methodology was later adopted by
  the Institute for Healthcare Improvement in the US in its drive to reduce hospital
  mortality.

As HES is not a live system, it usually takes between 9 and 12 months for a full financial year
 to become available through the on-line HES or safe haven extract; for example, complete
  2014/15 data would not be available until around December 2015.33,51 Similar databases to
  HES exist in Wales, Northern Ireland and Scotland. HES is increasingly being used by
  researchers due to its comprehensive coverage of NHS-funded hospital activity in England.

3.2 Secondary Uses Services

The SUS is the single comprehensive repository for healthcare data in England, which
enables a range of reporting and analyses to support the NHS in the delivery of healthcare
services.52 SUS is a data warehouse and the data contained in it is patient-level information
that can be patient identifiable, anonymised or pseudonymised, as required for the user's
needs. NHS providers and commissioners can use this data for 'secondary uses' purposes.
SUS provides a range of services that can be used to analyse, report and present the data it
contains. SUS data is held in a secure environment with patient confidentiality maintained to
the highest standards.

SUS can be accessed by any healthcare provider that submits patient data to SUS (NHS and
the independent sector), organisations that commission data from SUS, and organisations that
check healthcare compliance and consistency with national standards, such as area teams
(previously SHA). Access to SUS is restricted to three individuals per organisation and is
managed using role-based access controls (RBAC).52 Freely available aggregated data
containing details of all admissions to NHS hospitals in England are provided by the HES service.

3.3 Benefits of Healthcare Databases

The wealth of information available in healthcare databases from primary and secondary care is invaluable, and its use can play a key role in healthcare improvement in the UK. Many studies that have provided useful information in many disease areas have used data available from both primary and secondary databases. The use of the data available in these databases has increased significantly over time and has attracted much international interest; this is evident in the authorship of such studies. These studies have provided researchers with the tools to help make the most appropriate decisions regarding the care of patients and have provided data for monitoring bodies to monitor the progress of the NHS.

A summary of some of the benefits, according to Gnani and Majeed, for primary care and those according to the HSIC are listed below.33,52

- Primary care databases help to improve care during consultations
- Identification of patient groups that may benefit from health promotion and preventative medicine.
- Improved communication and patient follow-up, especially across healthcare organisations, and help with the management of chronic diseases.
- Help practices through the use of high quality data, e.g. cervical cytology recall, letters of referral etc. within their organisations.
- Support for clinical governance to improve quality of care and promote effective healthcare planning by PCOs.

Secondary care databases have similar benefits to primary care databases, and both offer all the benefits of better healthcare. Secondary health databases help to monitor and reveal health trends and patterns in NHS activity over time, and they can also support local service planning, assess effective delivery of care, inform the NHS of patient choices, and help determine that there is fair access to healthcare. In addition, they provide the basis for national indicators of clinical quality, and help to develop, monitor and evaluate government policy, as well as supporting the NHS and providing parliamentary accountability.

There is no doubt that these benefits can improve patient care, but success can only be achieved by ensuring that data entered into these databases (primary and secondary) is
accurate and variation is minimised. Healthcare providers entering data must also have the benefits of its secondary uses in mind and be enthused about entering data accurately. More education of healthcare providers on the potential of the data and the provision of national standards for data entry will need to be provided by the NHS.

3.4 Chapter Summary

Although the aim of the NHS to have a central system has not been successful, there remain several systems within the NHS that are being used to store patient data electronically within both primary and secondary care. Although these systems are limited in what they can capture and how they can be used to improve healthcare, nevertheless they enable researchers to use the data contained in these systems to conduct studies. However, the strength of the data contained in these databases is dependent upon the accuracy of the data entered and the willingness of healthcare professionals themselves to enter the data.
4.1 Introduction

The previous chapter described the various systems already in place in the NHS, such as CPRD, THIN, and QResearch in primary care and the secondary healthcare system HES. Some of the databases, like CPRD and HES, have existed for a long time and have been used to capture useful data.

This chapter presents a summary of the challenges associated with the maximisation of EPRs for research, such as the understanding of the general public and the lack of trust which exists concerning data security and the interpretation of the data protection laws. EPRs will help researchers to maximise the benefits of the use of patient data for research, but patients and the general public need to be confident that their privacy is protected and that adequate security systems are in place to protect their data.

4.2 The Patient and Public View of EPRs and Confidentiality

Using patient data to advance research is beneficial but also raises social, legal and ethical questions. Patients are keen to understand what their data is being used for and the security of the databases derived from clinical records, and patients and the public generally remain curious regarding the limitations in place where the use of their data is concerned. Although various safeguard methods, such as anonymisation, can be adopted, if patients do not understand or have the wrong perceptions then there will always be resistance despite research advances made through using their data. Growing interest in this issue led to the DoH conducting a consultation (DoH, 2009) to establish the opinions of the public, patients and other interested parties on the issues around patient consent and confidentiality in relation to additional uses of patient data for purposes other than care and treatment. The additional uses included healthcare research, clinical audits, measurement of quality and safety of care, improving public health, and planning improvements within the NHS. The primary aim of the consultation was to help ensure that future activities concerning additional uses of data should be communicated and conducted in a manner that would take into account public concerns, thus ensuring public confidence in the security and confidentiality of data.
The consultation was carried out with members of the general public, patients and healthcare professionals, and coverage included some traditionally hard-to-reach groups, such as the elderly, young people, black and minority ethnic groups, people with disabilities, recent immigrants, people with learning disabilities, people with mental health issues, rough sleepers, and lesbian and gay groups. Several consultation methods were used to encourage the widest possible responses from stakeholders, including a survey containing closed and open-ended questions, made available online on NHS CfH and DoH websites, as well as on paper allowing off-line completion, and regional stakeholder events were conducted at twelve locations across England. These stakeholder events served as open forums where participants could express their views, as an alternative to completing the survey. In addition, the consultation encompassed 16 small group meetings with the above-mentioned traditionally hard-to-reach groups, supplemented with six face-to-face interviews. Consultation materials were also sent out to 250 stakeholders, including patient groups and professional organisations.

Respondents were asked 23 questions, some of which focused on the use of anonymised patient data and linked, anonymised data for secondary or additional uses. Respondents were asked whether data in sealed envelopes should be used for additional purposes, and whether patients’ consent should be sought if the information in the sealed envelope were anonymised. The responses revealed a clear difference in views between researchers and other stakeholders: whilst 76% of researchers considered it acceptable to use anonymised sealed envelope information without the need for patient consent, only 26% of patients and 30% of the general public shared this view. Regardless of their differences, many respondents expressed the need for further information to be provided about sealed envelopes. Some suggested that the term ‘sealed envelope’ connotes strict protection, and could therefore mislead people because the information could be used without consent if it were anonymised.

The consultation also revealed that patients and the general public have concerns regarding the risk of ‘data loss’ (i.e. data being inadvertently made accessible to people who should not normally have access), and the ability of the NHS or Government to protect personal data. Consequently, many respondents suggested providing safeguards against the misuse of clinical data. The consultation also covered whether consent should be sought for the additional use of linked anonymised data, whether in a sealed envelope or not. Patients and the general public stated that the use of linked anonymised data increased the risk of identification, and around 30% were sceptical regarding the reliability of the anonymisation.
process, and hence they preferred that consent should be sought before linked anonymised data were to be shared with researchers. In contrast to the other stakeholder groups, most researchers saw the benefits in sharing linked anonymised data, but stressed the need for safeguards, consequently, they were in favour of exploring better ways of pseudonymising data.

Luchenski et al.\(^{54}\) also conducted a study on the perceptions of patients and the public, and their findings suggest that there is general support for the development of a national EPR system for purposes such as personal healthcare, policy and planning, and health research, but some remain sceptical and undecided. The study also found that demographics such as age, ethnicity and education played an important role in the responses provided. The study suggested that knowledge of the content and purpose of EPRs could help to gain the support of those still unsure; patients needed to be guided and safeguards put in place to help them decide.

Generally, patients’ and public attitudes towards data sharing can be said to have changed over time. With advances in technology and the willingness of people to share personal details on social media, patients are probably more open than they have ever previously been to having their data held and accessed electronically. Creating this awareness in patients is primarily the role of GPs, who must act as advocates for their patients, help them understand how their data will be used, and keep them aware of its usage; however, this has often presented a challenge, as many GPs have complained of a lack of guidance regarding what needs to be done.\(^{55}\) Riordan et al.\(^{55}\) looked at the attitudes of patients towards informed consent and the levels of awareness of EPRs. The results revealed that some respondents who reported having heard of EPRs, most frequently cited the media (i.e. newspaper, radio, television) and the NHS as their source of information. It is dangerous to allow only the media to be the source of information on EPR for patients and the general public, as there is the risk that the view presented will be biased and it most definitely will not provide sufficient information to allow patients to make their own decisions. Research has shown that a patient’s decision is based on the perceived sensitivity of the data, the trust they have in the recipient, and the extent to which a patient feels informed about how their data will be used.\(^{55}\)

To date, it seems that the UK Government has faced challenges with gaining the trust of patients and public regarding confidentiality, as there is concern that data will be shared with other companies. Jon Hoeksma, the editor of Insider, reported the results of an online survey
on care.data, which revealed that the public focused their anxiety on drug and insurance companies being able to buy confidential medical data and that such data would be treated as a commercial asset. Responders did not seem to be reassured that users would be legally prohibited from re-identifying individuals. Anonymity and the role that consent has to play remain a significant challenge for the NHS, with the general public still concerned that the pseudonymisation of data can be reversed, thus revealing patient identity. Consent is not required when patient data is anonymised and used for research, but the need for consent arises when identifiable patient-level data is to be used for research, as the Data Protection Act mandates that the use of any identifiable data requires the implicit consent of patients. Price, in an article entitled ‘Respecting patient confidentiality’, described how, the more a care team understands how privileged information is stored, disclosed and used, the less likely it is to abuse patient confidentiality. Better training may be required for healthcare professionals concerning patient confidentiality.

4.3 Privacy Laws and their Impact on Research

Researchers face the challenge of not knowing whether their research complies with all the interwoven clauses and conditions within and between the different legislation. In most cases, research will need to comply with several schedules within a single Act, whilst at the same time meeting the requirements of another Act. Regulators and data controllers have often adopted a conservative approach to legal interpretation, which makes it even more complex and has caused unnecessary restrictions around the use of personal data in research.

The legal framework governing medical research and the use of personal data for research is unclear and is open to different interpretations. The most important UK laws and acts governing research are:

- Data Protection Act 1998
- Human Rights Act
- Common Law of Confidentiality
- Administrative Law
- Section 60 of the Health and Social Care Act 2001
- Health and Social Care Act 2012
4.3.1 Data Protection Act 1998

This Act has a substantial impact on how research is carried out because it safeguards an individual’s rights. The Act covers the use of personal data across a wide range of sectors and will affect how patient data are used in research. In the UK, the use of identifiable data in research is governed by this Act and the common law duty of confidence. The Data Protection Act is based on the European Data Protection Directive (95/46/EC), which is currently being considered and amended by the European Parliament and Council.\textsuperscript{58,59} If the amendment is successful, it will make an already complicated issue for researchers even more complex.

The Act defines ‘personal data’ as that which relates to a living individual who can be identified from that data or any other information which is in the possession of, or likely to come into the possession of the data controller, who determines the purpose for which and the manner in which any personal data are to be processed. This includes any expression of opinion about an individual and any intentions of the data controller or any other person in respect of the individual.\textsuperscript{58,59} The Act defines ‘health records’ as personal records that relate to the physical or mental health of an individual which has been made by or on behalf of a healthcare professional in connection with the care of that individual.\textsuperscript{59} The Act mandates the fair processing of personal data which means individuals should have the opportunity to know what information an organisation holds about them and how that information will be used. Individuals should also be able to check an organisation’s records and make any amendments to the record held on them. The Act recommends that the processing of any personal data must comply with eight principles.

- It must be processed fairly and lawfully, and shall not be processed unless necessary. The processing of data as necessary has to meet one or more of schedules 2 and 3 of the Act.
- Personal data must be obtained only for one or more specified lawful purpose, and cannot be further processed in any manner incompatible with the original purpose.
- Personal data must be adequate, relevant and not excessive in relation to the purpose or purposes for which they are processed.
- Personal data must be accurate and where necessary kept up to date.
- Personal data processed for any purpose shall not be kept for longer than is necessary for that purposes.
Personal data shall be processed in accordance with the rights of the data subject under this Act.

Appropriate technical and organisational measures shall be taken against unauthorised or unlawful processing of personal data and against accidental loss or destruction of, or damage to personal data.

Personal data shall not be transferred to a country or territory outside the European Economic Area unless that country or territory ensures an adequate level of protection for the rights and freedoms of the data subjects in relation to the processing of personal data.

The Act recognises that research and statistical work requires medical information to be processed in ways other than those for which it was originally intended, and therefore gives exemptions from certain provisions of the Act as long as data is not processed in a way that brings substantial damage or distress to the subject and data is not processed in a way that is used to support or decisions in relation to the data subject.

The Act allows such data to be kept indefinitely and it can be exempt from the subject access rights if the results of the work are not available in a form that can reveal the identity of an individual. The Act generally gives data subjects the rights to access personal data about themselves which are held in either computerised or manual form, whenever the record was compiled.

The subject’s right include:

- That they should be informed whether personal data is processed, including that which is held or stored.
- A description of the data held and the purpose for which the data will be processed.
- A copy of the information held.
- Information on who provided the data.

Data subjects have the right to access all records irrespective of when they were created. This was not the case before the existence of the Act; however, requests from data subjects can be refused if a data controller is not satisfied with the identity of the person requesting the information or that it will not be processed properly. Other factors include that the data subject should have consented to the disclosure or that it is reasonable to honour the request without data subject consent.
Using data for research does not breach principle number 2 and is not unlawful on those grounds.\textsuperscript{58,59} Schedules 2 and 3 of the Act recognise consent as the basis of legitimately processing data though schedule 3, which covers sensitive data and specifically mentions that explicit consent must be obtained from a data subject in order to process their data. Explicit consent, as the name implies, involves data subjects clearly and unambiguously expressing their consent, either orally or in writing \textsuperscript{[16]}, whilst schedule 2 simply mentions consent and does not require for it to be explicit, which may be interpreted differently by different researchers. The Act also provides alternatives within schedules 2 and 3 that can qualify personal data for processing. One of the requirements of schedule 3 is that processing may take place if it is necessary for medical purposes and is undertaken by health professionals, including doctors and other health registered professionals, such as nurses, therapists and scientists who are heads of department\textsuperscript{59} or a person who in the circumstances owes a duty of confidentiality which is equivalent to that which would arise if that person were a health professional. The definition of ‘medical purposes’ includes the purposes of preventative medicine, medical diagnosis, medical research, the provision of care and treatment, and the management of healthcare services.\textsuperscript{59}

The Act is therefore open to various interpretations, which have confused researchers but do allow them to process anonymised data lawfully without having to obtain consent. However, this does not exempt researchers from ensuring that the processing of the data is fair. Scientific research generates important benefits by improving our understanding of society, health and disease. The Wellcome Trust is currently working with other stakeholders to ensure that the regulations create a clear legal framework that facilitates research while protecting the interests of data subjects.\textsuperscript{60}

In addition to the Data Protection Act, there are other laws that protect individuals, their data and their right to confidentiality.

\textbf{4.3.2 Human Rights Law 1998}

This guarantees the right to respect for an individual’s private and family life and incorporates the 1950 European Convention on Human Rights into UK law. Medical information on a person is seen an integral part of respect for private and family life. Exemptions can be granted to prevent civil disturbance or to protect public health and possible justifications for public bodies. To ensure that the rights of individuals remain
protected even when it is necessary to share or disclose data, the necessary safeguards need to be put in place to maintain security and must be backed by legally enforceable remedies.

The human rights section on privacy simply states that personal information about an individual (including official records, photographs, letters, diaries and medical records), should be kept securely and not shared without the individual’s permission, except in certain circumstances, such as to protect national security, public safety, the economy, health or morals, and to prevent disorder or crime, whilst protecting the rights and freedoms of other people.61

4.3.3 Common Law Confidentiality

The English common law on confidentiality protects an individual's right to have their personal data kept confidential, especially if it has the necessary qualities of confidence. Those who disclose information in situations where they know or ought to know that the information is confidential are liable under common law, unless disclosure is justified. Information disclosed by a patient to his or her doctor or therapist is regarded as confidential, should be kept that way and is subject to this duty. This includes all information that has the necessary quality of confidence about it, including information that one considers private. There are four circumstances, under which disclosure of confidential information is lawful: 1) where an individual to whom the information relates has consented; 2) where disclosure is in the overriding public interest; 3) where there is a legal duty to do so, for example a court order; and 4) where there is a statutory basis that permits disclosure, such as approval under Section 60 of the Health and Social Care Act 2001.62

Any healthcare provider wishing to disclose a patient's personal information to anyone outside the team providing care will first need to seek that patient’s consent before they are allowed to do so. If a disclosure is to occur, it will have to be based on the four circumstances described above. The NHS allows such information to be disclosed if consent is obtained from a patient to disclose the information or for NHS purposes, for example where the recipient needs the information to improve quality of care, where the information is required by statute or court order, or where passing on the information can be justified for reasons of public interest. Using confidential data for medical research in the interest of the public is therefore not necessarily considered a breach of confidentiality, but boundaries are very unclear and therefore the nature of the law is that the use of confidential data in research should be judged with respect to each individual concerned.62
4.3.4 Administrative Law

Administrative law governs the actions of public authorities, ensuring that the powers they are granted are exercised for the purpose for which they were intended. Researchers are bound to act only within the limits of the powers granted to them.

The approach often adopted by Government to address situations where a disclosure of information is prevented by lack of function, is to create, through legislation, new statutory gateways that provide public sector bodies with the appropriate information disclosure function. However, unless such legislation explicitly requires that confidential patient information be disclosed, or provides for common law confidentiality obligations to be set aside, then these obligations must be satisfied prior to information disclosure and use taking place, for example by obtaining explicit patient consent.63

It is not surprising that researchers find the privacy laws very difficult to interpret. Confidentiality remains an arguable part of the law and even the General Medical Council (GMC) cannot guarantee that researchers or doctors are safe from legal challenges by following the guidelines it provides. The GMC advises researchers to remain aware that they can be held accountable for their decisions regarding the use of confidential information.

4.3.5 Section 60 of the Health and Social Care Act 2001

This Act was enacted in England and Wales to help minimise the confusion surrounding the Data Protection Act and the Human Rights Act, which left research ethics committees interpreting the laws conservatively. It has now been incorporated into Section 251 of the Health and Social Care Act 2006, which was introduced to allow organisations to obtain patient identifiable information for medical purposes, in circumstances where it was impracticable to obtain informed consent from the patients concerned,64 such as when thousands of records are needed or the system is not capable of recording and respecting permission, or where the nature of the research may mean that consent will not be forthcoming. However, researchers have to complete a section 60 application, which was reviewed by the Patient Information Advisory Group (PIAG), and existed up until 31st December 2008. Their responsibility included providing advice to the Government (the Secretary of State for Health) about the use of information concerning patients. The main responsibility of PIAG was to decide whether researchers and NHS professionals should be given permission to see information about identified patients without first obtaining the consent of these patients. PIAG was replaced by the NIGB for Health and Social Care under
Section 158 of the Health and Social Care Act 2008, and responsibility for administering Section 251 powers were transferred to NIGB on 1st January 2009.65

Section 251 provides researchers with the opportunity to certify that research without consent is compatible with common law, but there is confusion about whether it is mandatory for researchers to seek section 60 permission or to rely on the public interest defence, which is permitted by common law, as well as some sections of the Data Protection Act. The responsibility of NIGB is to provide leadership and to promote consistent standards for information governance across health and social care. Section 251 of the Health and Social Care Act 2001 does not cover Scotland, where in 2000 the Confidentiality and Security Advisory Group for Scotland (CSAGS) was set up to provide advice on confidentiality and security of health related information to the Scottish Executive, the public and healthcare professionals. Exemptions allowed the NHS Information Centre to send SUS data to PCTs, which are now being abolished, for various uses, including risk stratification, where analysis of population level data is utilised to identify individuals who may, for example, be at high risk of a hospital admission so they can be dealt with by services such as virtual wards.66

4.3.6 Health and Social Care Act 2012

The NHS lists the reason for this Act as safeguarding its future, as the NHS needs to change to meet the challenges it faces, and only by modernising can the NHS tackle the problems of today and avoid a crisis tomorrow. The Health and Social Care Act 2012 places clinicians at the centre of commissioning, frees up providers to innovate, empowers patients, and provides a new focus on public health.68 The Act itself, unlike the others, gives obvious power to clinicians, which makes several privacy groups nervous, as they believe this could jeopardise confidentiality as many private companies will be commissioned to deliver services and there may be no control over who can view clinical data. The NHS believes that reforms and the Act will give patients a choice over the services they choose, including registering with GPs of their choice, and moving commissioning to GP-led clinical commissioning groups (CCGs) with the aim of ensuring that NHS services are based on local needs and priorities. The NHS has retained the services of the Patient Advice and Liaison Service (PALS) and the Independent Complaints Advocacy Service (ICAS) for any complaints or concerns that patients may have. Although this Act commissions clinicians to be more involved in the running of the NHS and empowers them to make certain decisions that the trusts were previously responsible for, many believe that it will lead to competitive tendering and the
complications of opening contracts to private sectors will lead to fragmenting of the NHS. Similarly, they are not sure that the best care will be delivered to patients if many healthcare providers treat them, and believe that the main priorities of these firms will be to make huge profits from the NHS and which eventually will lead to massive financial implications that will negatively affect the future of the NHS.

4.3.7 Caldicott Principles 1997

The *Review of the Uses of Patient-Identifiable Information* in 1997, chaired by Dame Fiona Caldicott, devised six general principles of information governance that could be used by all NHS organisations in England regarding access to patient information. A key recommendation of the Caldicott Report was the establishment of Caldicott ‘guardians’ of patient information throughout the NHS. A Health Service Circular (HSC1990/012) advised on the appointment of guardians to ensure that information governance is effective. Organisations that access patient data, such as acute trusts, ambulance trusts, mental health trusts, CCGs, special health authorities, commissioning support units and area teams, were all mandated to appoint Caldicott guardians. A Caldicott guardian ‘is a senior person responsible for protecting the confidentiality of a patient and service-user information and enabling appropriate information sharing.’69 A Caldicott guardian also has a strategic role, which involves representing and championing information governance requirements and issues at board or management team level, and where appropriate, at a range of levels within an organisation's overall governance framework.70 A guardian is usually a board level health professional or their deputy, and they are expected to develop local protocols for information disclosure, restrict access to patient information by enforcing strict need-to-know principles, and to regularly review and justify the uses of patient information.

4.4 Data Security

In addition to trying to understand the requirements of the data protection laws, the NHS and researchers face the challenge of trying to ensure that these laws are not broken. Patient data is extremely sensitive and if it is to be kept electronically, the public will need to be reassured that it is secure and in the hands of those who really need to have access to it. EPRs require strictly controlled access and guidelines are needed about who can access the records and, if their role requires them to have access to the health records, who can update records. For the successful implementation of EPRs the NHS will need well-thought-out standards to determine by whom and how records are accessed. HSCIC has listed some principles of
information security, which require all reasonable care to be taken to prevent inappropriate access, modification or manipulation of data from taking place. The NHS defines three cornerstones, confidentiality, integrity and availability, as the way to achieve good security and to prevent the inappropriate use of data: information must be secured against unauthorised access, confidentiality; information must be safeguarded against unauthorised modification, integrity; and information must be accessible to authorised users at times when they require it, availability.\textsuperscript{71}

The NHS is also obliged to do all it can through its security to meet the data protection laws, and has a number of bodies in place to check that these requirements are met. These bodies are detailed in the NHS information security code of practice for monitoring of NHS performance and through existing arrangements which have an interest in NHS information security management. The Audit Commission regularly conducts studies into information security management and related information governance issues. The DoH collects performance details as part of the annual information governance assessment, and these inform the work of both the Healthcare Commission and the Audit Commission, thereby ensuring that the NHS is closely monitored and meets the requirements of confidentiality and security where patient data is concerned. However, many breaches still occur,\textsuperscript{72} and the Big Brother Watch Report on NHS breaches, published in November 2014,\textsuperscript{73} stated that there had been at least 7,255 breaches in data security within the NHS. This is equivalent to 2,418 breaches per year, 201 breaches per month, 46 breaches per week and 6 breaches per day. The details of these breaches include:

- At least 50 instances of data being posted on social media
- At least 143 instances of data being accessed for ‘personal reasons’
- At least 124 instances of cases relating to IT systems
- At least 103 instances of data loss or theft
- At least 236 instances of data being shared inappropriately via email, letter or fax
- At least 251 instances of data being inappropriately shared with a third party
- 115 cases of staff accessing their own records

The public are therefore very nervous about such NHS breaches and do not trust that their data is secure within the NHS, which poses an obstacle to the successful implementation of EPRs.

Neame\textsuperscript{74} lists five sources from which privacy breaches principally arise:
Inadequate identification and authentication of individuals, allowing significant numbers of authorised system users to pose as someone else, using privileges belonging to another person.

Ready accessibility of electronically stored information, where restricted record access and read/write privileges are available to vastly more users than those with a need to know and with the authorisation of the patient concerned; inadequate logging of user activity and monitoring of logs for abuses.

Inappropriate disclosure, where data have been exported from an institution, for example on paper, on memory media, in mobile devices/laptops, in communications without authorisation and whilst inadequately secured, and/or passed to a recipient where privacy measures are inadequate.

Reporting requirements, some statutory and deriving from primary legislation, others arising out of departmental directives, and all requiring the disclosure of personal information provided in confidence.

Poor security, failing to protect a system against external hackers and malware, permitting users to access instant messaging services and system back doors (e.g. implemented for remote management which permits external access to a system often by-passing security controls).

The data breaches committed by the NHS and the subsequent huge fines have been a source of concern, leading the Information Commissioner to welcome the change in the law that will give his office the right to force NHS authorities to be audited for compliance with the Data Protection Act. From 1st February 2015, the Information Commissioners Office (ICO) became able to subject public healthcare organisations to a compulsory audit, which previously only central government offices have been subjected to. These audits will look at how the NHS handles patients' personal information, and can review areas including security of data, records management, staff training, and data sharing.

4.5 Chapter Summary

The concerns around confidentiality, the perception of patients and the general public, coupled with the complications associated with interpreting the privacy laws currently in place, can be seen as stumbling blocks by researchers when it comes to the use of patient data for research purposes. Researchers would benefit from more clarity on the privacy laws and some guidance on how to enter data and ensure consistency across the board. Patients are
willing to have their data used for research, but they need to be assured that the data remains
safe and will not be shared without their consent. The NHS has faced huge fines for data
security breaches in the past, which has not helped with public confidence.
CHAPTER 5 - OVERVIEW OF EPRS IN EUROPE, COMPARING WITH THE UK

5.1 Introduction

The previous chapter described the challenges associated with the use of EPRs in the UK. Issues around privacy and confidentiality, coupled with challenges resulting from the privacy laws which are open to interpretation by researchers, seem to be influencing the implementation of the full potential of EPRs.

This chapter summarises the efforts being made in other EU countries regarding the use of EPRs. This chapter does not provide a full comprehensive view of the work that is being undertaken in other EU member states, but instead presents an overview of the progress being made. The EU countries that are reviewed are Denmark, Estonia, France, Germany, the Netherlands and Sweden. The EU Commission is looking for member states to not simply implement EPRs into their healthcare systems, but to ensure that data is transferrable across borders. Countries are therefore making an effort to achieve this goal and the challenges being faced in the UK are common to most of the EU countries discussed.

In mid-2008 the EU Commission set 2015 as the target year for EPR interoperability, to ensure that key EPR datasets could cross European borders, and do so in conformity with medical rules and other relevant legal frameworks. The EU Commission encouraged member states to invest in beneficial e-health systems and services, as well as offering necessary support to help member states and to encourage the sharing of good practice. Interoperable electronic health records (EPRs) and ePrescribing systems are high on the political agenda in Europe and beyond. The EPR IMPACT study was commissioned by DG Connect, with the goal of supporting ongoing initiatives and the implementation work by the European Commission, member state governments, private investors, and other actors. The study aimed to improve awareness of the benefits and to provide new empirical evidence of the socio-economic impact and lessons learnt from successfully implemented systems, through the following objectives:

- Identify and apply state-of-the-art methods for detailed qualitative and quantitative socio-economic impact analyses for implemented interoperable EPR and ePrescribing systems across Europe and beyond.
- Disseminate the gained evidence and lessons learnt to the wider public.
- Develop policy recommendations to foster the faster diffusion of such eHealth applications.

The results of the EPR IMPACT study reported give grounds for optimism concerning the success, value and deployment of interoperable EPR and ePrescribing systems across Europe. The socio-economic gains to society eventually exceed the costs, although they take between six and eleven years to do so, and a typical development can reach annual socio-economic returns of up to 400%.

5.2 Estonia

Estonia has implemented a nationwide EPR system that integrates data from Estonia’s different healthcare providers to create a common record for each patient. It was built on Estonia’s X-Road network, a secure gateway service architecture that hosts 3000 eservices available to Estonian citizens. The system actually retrieves data as needed from the various providers, who may be using different systems, and presents it in a standard format.

The system is a powerful tool that allows doctors to access a patient’s records easily from a single electronic file. A doctor can read test results as soon as they are entered, including image files, such as X-rays. All end users of Estonia’s EPR system can access their personal records and those of their children by logging into the Patient Portal via an electronic ID card. Patients can review their past visits to doctors and current prescriptions, control which doctors have access to their files, and even receive general health advice. Doctors and patients have equal viewing access, and in an emergency situation, a doctor can use a patient’s ID card to read time-critical information, such as blood group, allergies, recent treatments, ongoing medication or pregnancy status.

5.3 The Netherlands

Several million Euros were invested by the Dutch government and about 10 years was spent trying to develop a federated EPR, but this failed in 2011 due to privacy concerns. Instead of abandoning the project completely, ownership and operation of the project was transferred to a non-profit organisation, which is solely controlled by healthcare providers, patients and health insurance companies. This organisation tackled the privacy concerns and changed the opt-out to opt-in for patient participation in the national information exchange.
There is a planned development that will run from 2013-2016 whereby patient access to both health record access logs and the health information made available will be part of the agreed development plan. Healthcare providers must make available ‘a client service desk’ for health record consultations. Patients can then contact the service desk to select which healthcare organisations can be provided with information about them or the ones they wish to withdraw access from, and they will also get the opportunity to access their records at these service desks.

5.4 Sweden

Sweden rolled out the first stage of the Swedish National Patient Summary initiative in 2009. The aim of the project, known as the National Patient Summary (NPÖ) is to improve patient security and quality of care by rolling out the national electronic health record system in stages. Just 1 year after commencing the implementation, the EPR was deployed in one county, and this was the first of the country’s 21 healthcare regions. Health Share, the system chosen, allows the creation of summary views, showing a patient’s medical record on a national basis for the NPO. Patients and clinicians will be able to share information on the web-based system from wherever they are, as long as they have the authority to do so.77,78

The national patient overview provides a modern tool to enable efficient co-operation between county councils, local authorities, and other healthcare providers. One particular point of interest is that individual citizens will eventually have access to their own medical records via the internet.

5.5 Germany

It has been the German Ministry of Health’s intention to introduce an EPR system for over 10 years but this has not yet been implemented.77,79 There are still ongoing discussions about data security, and politicians, health professionals, and health insurers have raised concerns. In 2011 a compromise solution was introduced via a national electronic card known as the Elektronische Gesundheitskarte, or eGK, which allows patients to access both administrative data and a number of e-health services on offer.79 The aim was for the eGK to stand at the heart of a network of e-health services, delivered locally. The project began in 2004 and suffered a number of setbacks before being halted in 2010 due to the concerns of many doctors and IT security specialists regarding access to patients’ health data. The first version of the eGK, which launched in 2006, stored medical details directly onto the chip of each
individual’s card, but this was deemed too risky with the potential to impact on privacy. The card was re-launched in a new form in 2010 allowing the linking of various healthcare providers (doctors, hospitals, pharmacies, health insurance companies) by means of telematics. By default, the card now contains only social security and insurance administrative data, enabling medical costs to be reimbursed. Tested in six German states since October 2011, the card is now available to all citizens.

5.6 France

The Dossier Medical Personnel (DMP) has been implemented in France based on a mandate instigated by French national legislation and is available to all patients through web services. DMP was launched in January 2011 and was gradually rolled out across all French territories through voluntary adoption by patients and healthcare professionals.

DMP is a free service that aims to improve the coordination of healthcare in France and is supported by healthcare professionals. DMP has now become part of patient expectations for care and makes patient information readily available, facilitating communication between healthcare professionals and patients. It will form the infrastructural and technical base for numerous e-health services, including those proposed by public authorities and the private sector. Patients will be able to have direct access to their records through an internet DMP portal called ‘patient web access’ which was created by ASIP Santé, the shared healthcare information systems agency. Patients can add files or documents to their records in the same way as healthcare professionals can add files or documents. Patients can also verify that those accessing their records have the right to do so, and DMP has helped patients in France to feel in control of their medical records. Pharmacists are also required to feed into the French pharmaceutical record scheme.

5.7 Denmark

Denmark has implemented a national e-health portal which provides a shared infrastructure that allows all parties within the health sector to collaborate across professional and IT-related boundaries with the individual patient at its centre. The system is a public, internet based portal that collects and distributes healthcare information between the general public and healthcare professionals. The system is reported to have transformed the Danish healthcare system from a silo-orientated structure to a structure that is more patient-centred. All Danish people have had access to this system since 2003, and healthcare professionals are
able to communicate freely with their patients and patients’ families. Every Danish citizen has their own personal page; however they need to identify themselves before they can gain access to it. They can obtain up-to-date access to their personal records, can send secure emails, book appointments, renew medicine prescriptions, and look at waiting lists for operations or to see a specialist. They can also view the quality ratings of hospitals, register as an organ donor, and access local diseases management for in-out patients.

The portal provides 150,000 Danish health professionals with the information that they require to make appropriate decisions for their patients. They are able to obtain up-to-date information about their patients from this portal because most hospitals and laboratories also access the portal. The healthcare system is based on two main principles: 1) free and equal access to public healthcare, which includes general and specialised practitioner services and all public hospital services, as well as private co-payment, which includes dentists and out-of-hospital medicines; and 2) universal coverage, meaning that residents in Denmark are entitled to public healthcare benefits in kind.

Individual hospitals have their own local EPRs, which contain their own patient information only. Information from other hospitals or laboratories is not included in these local EPRs but can be found via Sundhed.dk, which is an important supplement to local EPRs. A study conducted by Kierkegaard showed that Denmark’s healthcare system and use of e-health may be regarded as excellent in terms of world standards. The case study performed concluded that the country still has room for improvement, as Denmark suffers from electronic medical record (EMR) fragmentation as a result of the state’s failure to issue any common standards for interoperability and a technical platform to ensure health information exchange.

5.8 Comparison between the UK and Europe

The challenges faced by the NHS are not unique to the UK but can be seen across the six countries that were reviewed. Some countries, like Denmark, are slightly ahead with their implementation of some form of central system that is accessible to patients and their families. However, the concerns regarding data security and privacy experienced in the UK seem to be a global problem. Pathfinder sites, which are pilots for the care.data plans, are certainly a step in the right direction. NHS Scotland has introduced a key information summary (KIS) and continues to develop it for the provision of a summary of demographic, allergy and medication information for millions of patients in Scotland. Craig et al. report
that patients are willing to consent to share data with unscheduled care clinicians, and patients benefitting from this include those with complex or multiple conditions, who are at high risk of using emergency services. Similarly, Wales has also developed a system of individual health records (IHR) which summarise key information from a patient’s medical record.86 Swansea acute GP services have reported that the system has benefitted their work by helping them to diagnose in urgent situations and deliver treatment faster.87

5.9 Chapter Summary

EU member states are also working towards the implementation of EPRs. From the brief review undertaken, it is clear that many countries begin by wanting to implement a single national system but face numerous challenges and eventually have to compromise by implementing something that is not quite what was originally intended; the main cause for concern is privacy. Denmark seems to have made more progress compared to other countries but improvements are still needed. This review is not a comprehensive description of the work that is currently being carried out in the EU, but is intended to show that the challenges being faced in the UK regarding EPRs are common across EU countries. It can be concluded that implementing an EPR system is a continuous process and one which will continue to evolve over time.
CHAPTER 6 - METHODS

6.1 Introduction

The previous chapter summarised the work being undertaken in other EU countries and looked at how work being undertaken in the UK compares with that in other EU countries. The challenges experienced in the UK concerning privacy are common to most of EU countries trying to implement EPRs.

From the literature review it can be concluded that EPRs currently exist and are actively being used for research purposes. The fact that they are not interoperable and have several limitations has propelled the NHS to seek a central national system that is believed to have the potential to take healthcare and research to the next level. Although researchers face other challenges, including a lack of set standards for data entry, interpretation of privacy laws, and how the general public views their data being stored electronically or shared, this has not stopped them from using existing data available in primary and secondary databases to conduct research. A lot of work has been carried out on single systems, such as CRPD and HES, but there have been no comparisons of the existing large systems to look at the collective impact that they are having on research and how this has grown over time.

This chapter will summarise the methods used to carry out the bibliometric analysis of three primary care databases, CRPD, THIN, QResearch, and the secondary care database, HES, to look at the combined impact of these databases.

6.2 Hypothesis and Research Question

EPRs exist in one form or other and are being used by researchers. Studies have shown over the time that data contained in databases, like CPRD and HES, is being more actively used for research, but many limitations still exist. This thesis focuses on the combined impact that the EPRs contained in the chosen primary care databases (CPRD, THIN and QResearch) and the secondary care database (HES) are having on research. The study examines whether there has been an increase in the use of data in the chosen databases for research purposes, and whether the impact they are making on their own or jointly is tangible.

6.2.1 Hypothesis

Hypothesis 1: The use of EPRs is advancing research in many medical specialty areas and is
continuing to grow over time.

Hypothesis 2: Advancement in research is seen not only via large databases, such as HES and CPRD, but also through smaller databases, like THIN and QResearch.

6.2.2 Research Questions

- In spite of the lack of a central national system, are the current healthcare databases being used for research?
- Are the advancements seen through the use of these databases (increased use by researchers) unique to just one type of database, e.g. CPRD and HES (which have existed for longer) or can interest be seen for the other selected databases (THIN and QResearch)?
- What has been their combined impact over time?

6.3 Aim and Objectives

The primary aim of the research is to show that EPRs can be used to advance clinical, epidemiological and health services research. The specific objective to achieve this is:

- To conduct a bibliometric review to analyse the combined research outputs and the longitudinal growth in the number of publications using the three primary care databases CPRD, QResearch and THIN, and the secondary care database HES, to show the benefits of EPRs in research and the progress that they are making over time, even though the UK is yet to implement one central national system.

6.4 Overview of Method

A bibliometric analysis was carried out on the primary care databases (CPRD, QResearch and THIN). Publications using data from each of the databases were extracted and analysed and an overview of the method used is described below and illustrated in Figure 6.1.

For CPRD, publications were extracted which had been published between 1993 and 2013 (a 20-year period), using the Science Citation Index (SCI) of the Thomson Scientific Institute for Scientific Information (Web of Science). Conference abstracts and posters were not included in the data. The same method was used to extract QResearch publications from 2004 to 2013 (a 9-year period). Data for THIN was obtained from the Department of Primary Care
and Population Health, University College London (UCL) and verified using the Web of Science. The data was provided in a Microsoft Excel format and contained details of the author, title of the article, the journal in which the article was published, article reference, and the year of publication. Publications provided by UCL dated from 2004 to 2013. The number of times publications were cited, their speciality area, and the names of the journals they were published in were then extracted using the Web of Science for all three databases.

To review the impact of HES data on research, publications generated from the use of HES data from 1996-2014 (a 19 year period) were extracted from PubMed and analysed. Extracted publications were examined using the Web of Science for details of the publication’s speciality area.

### 6.5 Search Criteria

Specific criteria were used to search for publications on the Web of Science and are listed in Table 6.1. Although data was collected for earlier years for CPRD and HES (1993-2013 and 1996-2013, respectively), the comparison and combined impact was assessed only from 2004 -2013 because data for THIN and QResearch in the earlier years does not exist.

#### Table 6.1: Search Criteria used in Web of Science

<table>
<thead>
<tr>
<th>Database</th>
<th>Source of original data</th>
<th>Search terms in Web of Science</th>
<th>Years searched</th>
<th>Filtering method</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPRD</td>
<td>Web of Science</td>
<td>Clinical Practice Research Database, CPRD, &amp; GPRD</td>
<td>1993-2013</td>
<td>Posters and reports were excluded</td>
</tr>
<tr>
<td>QResearch</td>
<td>Web of Science</td>
<td>QResearch</td>
<td>2004-2013</td>
<td>Posters and reports were excluded</td>
</tr>
<tr>
<td>THIN</td>
<td>UCL</td>
<td>Publication authors and year used as search criteria to verify</td>
<td>2004-2013</td>
<td>Posters and reports were excluded</td>
</tr>
<tr>
<td>HES</td>
<td>Web of Science</td>
<td>HES Secondary Care Database</td>
<td>1996-2014</td>
<td>Posters and reports were excluded</td>
</tr>
</tbody>
</table>
6.6 Specialty Area Grouping for Primary Care Databases

The speciality areas of publications for all three databases were categorised into four groups and speciality areas were classed as follows:

- Speciality areas with ≥ 50 publications = Group 1
- Speciality areas with < 50 but ≥ 20 publications = Group 2.
- Speciality areas with publications < 20 but > 10 publications = Group 3
- Speciality areas with ≤ than 10 publications = Group 4

All three databases were assessed for overall growth, and their speciality areas were compared for similarities. To fairly compare the three databases, publication analysis had to focus on the overlapping period of data availability, namely 2004 to 2013. The publications in the three databases were compared, looking at the average number of publications per year, individual growth, and cumulative growth over the 10-year period. The three ‘most published in’ journals were also recorded across the databases.
6.7 Specialty Area Grouping for the Secondary Care Database

Speciality areas were categorised into four groups.

- Speciality areas with \( \geq 50 \) publications = Group 1
- Speciality areas with \(< 50 \) but \( \geq 20 \) publications = Group 2
- Speciality areas with \(< 20 \) articles but \( > 10 \) articles = Group 3
- Speciality areas with \( \leq 10 \) articles = Group 4

The average number of publications per year was calculated and cumulative growth over the 19 year period was also looked at, as well as the journals that were most published in.

6.7.1 Rationale for Speciality Areas

Publications were grouped as explained above, looking at ‘greater than or equal to 50 articles and less than 10 articles’, so that a good overview could be obtained. If the group \( > 100 \) has been used then this would have limited what could be reported, and some speciality areas with smaller publications would have been overlooked.

6.8 Chapter Summary

The goal of the research was to evaluate the research output from three large primary care databases and a secondary database. The aim was to show that the impact on research of EPRs currently present in the primary and secondary care databases is positive, and their combined impact is even more so. The use of the Web of Science presented some limitations, but ensured that the publications contained in all the databases were evaluated in the same manner in order to reduce bias. Findings from the bibliometric analysis are presented in Chapter 7 and limitations of the research are further discussed in Chapter 8.
CHAPTER 7 - RESULTS

7.1 Introduction

The previous chapter looked at the methods used to carry out the bibliometric analysis which assessed the research output from three primary care databases, CPRD, THIN and QResearch, and the secondary database, HES. This chapter summarises the findings of the bibliometric analysis performed and lists the number of publications for each of the databases per year, the speciality areas the publications were assigned, and their combined impact. The three primary care databases were evaluated and their individual results are presented, together with an evaluation of their combined impact. The journals that were ‘most published in’ for all three primary care databases are also presented. The HES database was assessed and analysed separately, and the number of publications per year, speciality areas and the journals most published are presented.

7.2 Primary Care Databases Results

7.2.1 CPRD Database

A total of 1,140 publications categorised into 28 speciality areas were extracted for CPRD. The results shown in Table 7.1 present the number of publications from 1993 to 2013.

Table 7.1: Publications utilising CPRD data between 1993 and 2013

<table>
<thead>
<tr>
<th>CPRD Speciality Area Grouping</th>
<th>Number of Publications</th>
<th>Total No. of Publications</th>
<th>% Publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPRD Group 1</td>
<td>≥50 publications</td>
<td>412</td>
<td>36.1</td>
</tr>
<tr>
<td>CPRD Group 2</td>
<td>&lt; 50 publications, ≥ 20</td>
<td>229</td>
<td>20.1</td>
</tr>
<tr>
<td>CPRD Group 3</td>
<td>&lt; 20 publications, ≥ 10</td>
<td>117</td>
<td>10.3</td>
</tr>
<tr>
<td>CPRD Group 4</td>
<td>≤10 publications</td>
<td>382</td>
<td>33.5</td>
</tr>
</tbody>
</table>

The largest speciality areas with ≥ 50 publications (Group 1), were pharmacology (284 publications), health services research (71 publications), and public health (57 publications). The average number of articles published per year over the 20-year period from 1993–2013 was 54, with the most publications in 2013 (144 publications). CPRD has shown consistent growth in publications output since its data first started being used for research. As expected, there were relatively few publications in the early years; publications from 1993 to 1997 were
10, 9, 8, 7 and 17, respectively. The publication rate increased in 1998 and 1999, with 30 articles in each year, and has continued to grow ever since.

Between 2004 and 2013 (10 year period), the total number of publications listed in the CPRD database was 825, which shows significant growth within this period, with 72.3% of CPRD publications having been published in the last 10 years. The highest number of publications was published in the journal *Pharmacoepidemiology and Drug Safety*, representing 4.5% (52) of all CPRD publications.

### 7.2.2 QResearch

In total, 77 articles categorised into 13 speciality areas were published from studies conducted using QResearch data between 2004 and 2013 and the results are presented in Table 7.2.

Table 7.2: Publications utilising QResearch data between 2004-2013

<table>
<thead>
<tr>
<th>QResearch Speciality Area Grouping</th>
<th>Number of Publications</th>
<th>Total No. of Publications</th>
<th>% Publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>General &amp; Internal Group 1</td>
<td>≥50 publications</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>QResearch Group 2</td>
<td>&lt; 50 publications, ≥ 20</td>
<td>35</td>
<td>45.5</td>
</tr>
<tr>
<td>QResearch Group 3</td>
<td>&lt; 20 publications, ≥ 10</td>
<td>26</td>
<td>33.8</td>
</tr>
<tr>
<td>QResearch Group 4</td>
<td>&lt;10 publications</td>
<td>16</td>
<td>20.8</td>
</tr>
</tbody>
</table>

The average number of publications per year over the 9-year period is around 8, with most articles being published in 2011 (17 publications) and 2012 (20 publications). The highest number of articles was published in the *British Journal of General Practice*, which represented 24.6% (19) of the total publications for data from QResearch.

### 7.2.3 THIN Database

In total, 394 articles categorised into 32 speciality areas were published for studies conducted using THIN data between 2004 and 2013 and the results are presented in Table 7.3.

Table 7.3: Publications utilising THIN data between 2004-2013

<table>
<thead>
<tr>
<th>THIN Speciality Area Grouping</th>
<th>Number of Publications</th>
<th>Total No. of Publications</th>
<th>% Publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>THIN Group 1</td>
<td>≥50 publications</td>
<td>167</td>
<td>42.4</td>
</tr>
<tr>
<td>THIN Group 2</td>
<td>&lt; 50 publications but ≥ 20</td>
<td>75</td>
<td>19.0</td>
</tr>
<tr>
<td>THIN Group 3</td>
<td>&lt; 20 publications but ≥ 10</td>
<td>105</td>
<td>28.6</td>
</tr>
<tr>
<td>THIN Group 4</td>
<td>&lt;10 publications</td>
<td>47</td>
<td>11.9</td>
</tr>
</tbody>
</table>
The largest speciality areas for publications from THIN with > 50 publications (Group 1) were pharmacology (116) and general medicine (51). The average number of articles published per year over the 9-year period was around 39, with the most articles published in 2010 (87 publications). The journal *Pharmacoepidemiology and Drug Safety* accounted for 21.3% (95) of all publications, and the remaining 78.7% of articles were published in journals which contained between 1 and 14 other articles utilising THIN data. The majority of journals that published articles using THIN data only published one article.

### 7.2.4 Combined Impact

The results presented in Table 7.4 and Figure 7.1 show an increase in the number of publications using data from all three databases. Over the 9-year period, publications using THIN and QResearch data have slowly increased over time, while CPRD publications have increased substantially in the last 4 years. All three databases produced a combined total of 1,296 publications over the 9-year period (2004-2013), with CPRD representing 63.6% (825), THIN 30.4% (394) and QResearch 5.9% (77) of the total collective number of publications across these electronic health databases. The growth across all three databases was consistent over the 9-year period, with the highest number of publications listed in 2013.

**Table 7.4: Annual number of publications published per database between 2004 and 2013**

<table>
<thead>
<tr>
<th>Year</th>
<th>CPRD</th>
<th>THIN</th>
<th>QResearch</th>
<th>Total No. of Publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004</td>
<td>70</td>
<td>2</td>
<td>2</td>
<td>74</td>
</tr>
<tr>
<td>2005</td>
<td>68</td>
<td>9</td>
<td>2</td>
<td>79</td>
</tr>
<tr>
<td>2006</td>
<td>61</td>
<td>14</td>
<td>2</td>
<td>77</td>
</tr>
<tr>
<td>2007</td>
<td>61</td>
<td>30</td>
<td>4</td>
<td>95</td>
</tr>
<tr>
<td>2008</td>
<td>72</td>
<td>45</td>
<td>7</td>
<td>124</td>
</tr>
<tr>
<td>2009</td>
<td>85</td>
<td>43</td>
<td>4</td>
<td>132</td>
</tr>
<tr>
<td>2010</td>
<td>69</td>
<td>87</td>
<td>9</td>
<td>165</td>
</tr>
<tr>
<td>2011</td>
<td>94</td>
<td>43</td>
<td>17</td>
<td>154</td>
</tr>
<tr>
<td>2012</td>
<td>101</td>
<td>72</td>
<td>20</td>
<td>193</td>
</tr>
<tr>
<td>2013</td>
<td>144</td>
<td>49</td>
<td>10</td>
<td>203</td>
</tr>
<tr>
<td>TOTAL</td>
<td>825</td>
<td>394</td>
<td>77</td>
<td>1296</td>
</tr>
</tbody>
</table>
Figure 7.1: Cumulative growth of primary care database publications from 2004-2013

7.2.5 Publication Journals

The journal most published in across the three databases was *Pharmacoepidemiology and Drug Safety*. As shown in Table 7.5, most publications from the CPRD and THIN database occurred in this journal.

Table 7.5: Top three journals most often published in across all three databases from 2004-2013

<table>
<thead>
<tr>
<th>Journal</th>
<th>CPRD Publications</th>
<th>QResearch Publications</th>
<th>THIN Publications</th>
<th>Total No. of Publications</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Pharmacoepidemiology and Drug Safety</em></td>
<td>52</td>
<td>4</td>
<td>95</td>
<td>151</td>
</tr>
<tr>
<td><em>British Medical Journal</em></td>
<td>19</td>
<td>18</td>
<td>14</td>
<td>51</td>
</tr>
<tr>
<td><em>British Journal of General Practice</em></td>
<td>23</td>
<td>19</td>
<td>1</td>
<td>43</td>
</tr>
</tbody>
</table>

7.3 Secondary Care Database Results

7.3.1 Publication Growth

In total, 534 studies that used HES data published between 1996 and 2014 were extracted from PubMed. As listed in Table 7.4, the results show an increase in the number of publications over the 19-year period, with the most growth in the last 5 years (Figure 7.2). The average number of publications over the 19-year period was 28 per year, and the greatest increase during this period was between 2013 and 2014, with 73 and 113 publications respectively. All 534 publications were categorised into a total of 44 publication groups, and
as listed in Table 7.7, speciality areas were assigned to four groups depending on the number of publications.

**Table 7.6: Number of publications using HES data between 1996 and 2014**

<table>
<thead>
<tr>
<th>Year</th>
<th>No. of Publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996</td>
<td>1</td>
</tr>
<tr>
<td>1997</td>
<td>3</td>
</tr>
<tr>
<td>1998</td>
<td>8</td>
</tr>
<tr>
<td>1999</td>
<td>2</td>
</tr>
<tr>
<td>2000</td>
<td>7</td>
</tr>
<tr>
<td>2001</td>
<td>5</td>
</tr>
<tr>
<td>2002</td>
<td>10</td>
</tr>
<tr>
<td>2003</td>
<td>12</td>
</tr>
<tr>
<td>2004</td>
<td>19</td>
</tr>
<tr>
<td>2005</td>
<td>12</td>
</tr>
<tr>
<td>2006</td>
<td>19</td>
</tr>
<tr>
<td>2007</td>
<td>40</td>
</tr>
<tr>
<td>2008</td>
<td>27</td>
</tr>
<tr>
<td>2009</td>
<td>30</td>
</tr>
<tr>
<td>2010</td>
<td>39</td>
</tr>
<tr>
<td>2011</td>
<td>45</td>
</tr>
<tr>
<td>2012</td>
<td>69</td>
</tr>
<tr>
<td>2013</td>
<td>73</td>
</tr>
<tr>
<td>2014</td>
<td>113</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>534</strong></td>
</tr>
</tbody>
</table>

**Figure 7.2: Cumulative growth of publications using HES data from 1996-2014**
7.3.2 Speciality Areas

*Group 1 (≥ 50 publications)*

Group 1 consisted of 102 publications, with 52 publications assigned to the general surgery speciality area and 50 to the health statistics speciality area.

*Group 2 (<50 but >20 publications)*

Group 2 consisted of 205 publications, and the majority of the publications fell into this category. Speciality area examples included oncology (39), infectious diseases (38), orthopaedic surgery (31), vascular surgery (29), obstetrics and gynaecology (28) and paediatrics (22).

*Group 3 (≤20 but ≥10 publications)*

Group 3 accounted for 143 publications, and consisted of speciality areas including gastroenterology (20), ophthalmology (17), cardiology (16), urology (15), neurology (15) and public health (14).

*Group 4 (< 10 publications)*

Group 4 accounted for 84 publications and speciality areas included respiratory (8), pharmacology (6), nephrology (6), dentistry (5), haematology (4) and hepatology (3).

There was one non-medical speciality area, management, which was assigned to an article looking at the management of large-scale change within the NHS.

<table>
<thead>
<tr>
<th>Articles Per Group</th>
<th>Total Articles in Speciality Area</th>
<th>% Publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgery &amp; Health Statistics - Group 1 (&gt;50 articles)</td>
<td>102</td>
<td>19.1</td>
</tr>
<tr>
<td>Group 2 (&lt; 50 articles but &gt; 20)</td>
<td>205</td>
<td>38.3</td>
</tr>
<tr>
<td>Group 3 (&lt;20 articles but ≥ 10)</td>
<td>143</td>
<td>26.7</td>
</tr>
<tr>
<td>Group 4 (&lt; 10 articles)</td>
<td>84</td>
<td>15.7</td>
</tr>
</tbody>
</table>

7.3.3 Publication Journals

The journals that published articles utilising HES data were some of the most reputable journals. Table 7.8 details the three ‘most published in journals’. The *British Medical Journal* published the most articles compared to the other journals for HES publications. However, most journals published just one article, accounting for 93 of the publications reviewed.
Table 7.8: Journals with highest number of HES data publication

<table>
<thead>
<tr>
<th>Journals</th>
<th>No. of Publications</th>
<th>% Publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>British Medical Journal</td>
<td>48</td>
<td>9</td>
</tr>
<tr>
<td>British Journal of Surgery</td>
<td>29</td>
<td>5.4</td>
</tr>
<tr>
<td>Journal of Public Health</td>
<td>20</td>
<td>3.7</td>
</tr>
</tbody>
</table>

7.4 Chapter Summary

An increase in publication over the past years was observed in the findings of this study, showing that researchers are increasingly using data available in these databases to advance research. This use of the database was not simply of the larger databases, like HES and CPRD, as an increase in publications could also be seen for even the smallest of the four databases analysed, QResearch.

The results show that the three primary care databases produced a combined total of 1,296 publications over the 9-year period (2004-2013), with CPRD representing 63.6% (825), THIN 30.4% (394) and QResearch 5.9% (77) of the total collective number of publications across these electronic health databases. A further 534 studies used HES data for research purposes between 1996 and 2014 over a 19-year period. The publications were categorised into speciality areas to allow for easy analysis of the information. It was also found that researchers in some speciality areas were more inclined to use the data from particular databases; for example, more pharmacology studies were conducted using CPRD data compared to data from other databases, while surgery and health statistics studies tended to use data from the HES database.
8.1 Summary of Key Findings

Storing patient records electronically is the way forward and there are enormous benefits associated with the storage of records in this format. Paper records have many limitations that will deter the advancement of research and the improvement of healthcare as a whole. The evidence suggests that stored health records are already advancing research and have contributed highly to many speciality areas. For EPRs to be further advanced, there needs to be complete transparency of what data collected will be used for and strict control of who accesses the data. Patients will be willing to have their data used for research if they feel part of the whole process. Healthcare providers need to be encouraged to enter data accurately and standards need to be instigated to guide them on what is required. Confusion around the privacy legal framework needs to be clarified so that clinicians can understand what is required. The government also needs to ensure that there is adequate education for the general public on the use of personal data stored electronically, so that the benefits of EPRs can be harnessed.

8.1.1 Literature Search Findings

Information from electronic health records for research purposes can indeed advance healthcare services, and evidence of this can be seen in the many initiatives that are already in place. The DoH recognises the potential of EPRs and has therefore encouraged the advancement of EPRs for research purposes, but there are many hurdles that need to be overcome. Reviewing the literature from various sources highlighted that cohesive guidelines, set processes, and an understanding of the legal framework within the healthcare system is missing or insufficient. This has led to a lack of harmony within the systems or processes employed by researchers and general practices. Researchers are often unclear on what their limitations are, as they face various obstacles in trying to carry out their research, and GPs are sometimes unaware of the processes they need to follow when information from patient records are to be made available to research.

The lack of standards and processes creates confusion around how patient data should be handled and processed. Many patients have therefore lost confidence in the healthcare system as far as privacy is concerned, and do not trust that their privacy and data are well protected. Various safeguards, such as the anonymisation of data, use of sealed envelopes, etc., are in
place but are generally not understood by patients. Many patients fear that data held on them are also housed by systems that could easily be hacked or accessed by NHS staff not directly involved in their care. The NHS has proposed many initiatives to advance storing health records electronically but it seems that until the root cause of this problem is tackled, then the NHS may remain in a vicious cycle of starting initiatives to progress EPRs, which would allow patients to have compatible digital records so that their health information can follow them around the health and social care system, but then having to compromise this aim due to a lack of patient trust in data security. If compatible digital records were to be achieved, then in the vast majority of cases, whether a patient needed a GP, a hospital or a care home, the professionals involved in their care could see their history at the touch of a button and share crucial information.

The data held within general practices has great potential for advancing research but also significant limitations if processes cannot be made robust and concise and rolled out nationally, and there will continue to be hindrances in advancing an EPR system. Before processes are defined, there is a need for both patients and physicians to be completely committed to the concept of EPRs and to want and appreciate the benefits that they offer. Available research suggests that, apart from the complexities concerning a lack of guidance, physicians have some reservations about EPRs in general, such as the fact that a patient’s computer records should promote the seamless transfer of care from one clinical team to another, which would best be accomplished by using a conversational or narrative format rather than chunks of information scattered across divergent screen pages. Physicians also do not want to be burdened with the data entry of information and believe that this will be time-consuming; they also believe that trying to enter structured data can subtly change the meaning of the item coded. In addition, physicians are concerned about the costs associated with implementing electronic databases in their practices and are concerned about keeping the required codes updated and standardised.

The concerns of patients are centred primarily on trust, although research exists to show that many are open to the idea of EPRs and accept that we now live in a technology age; however, patients will feel more comfortable if there are tight processes and checks in place to ensure that their privacy is protected. Patients want information on what their data will be used for, and more specifically, who will access it. A study carried out by Pyper et al., on the views of patients regarding access to electronic health records in primary care, showed that patients were concerned about confidentiality and security, and especially the linking of their
electronically stored data with other organisations. Some were concerned about accuracy and completeness, and whether the information in records could be inaccurate, misleading or judgmental, with some who had seen their records reporting that they contained omissions or errors. Patients were also keen to see their records and wanted to be involved in their development; this desire to access records is not unrealistic and can be accomplished through careful consideration of what patients want.

EU member states, like Denmark, that have successfully implemented EPRs have all given access to patients. Other research has shown that patients are receptive to the establishment of EPRs, with most respondents preferring that partial or anonymous data should be shared rather than their whole complete health records. If patients and the public are open to EPRs, then the issue that needs to be addressed is trust. The government must work hard, not only to show the benefits of EPRs, but also to convince patients and the general public beyond reasonable doubt that their data is secure and will not be transferred to anyone without their consent.

Patients’ involvement in shared decision-making is desirable and transactional services may be a better way to achieve this than investment in records access. The NHS is working towards this by using the lessons learnt from past failed EPR initiatives in the proposed care.data system. The care.data pathfinder sites should hopefully provide or reveal areas of improvement before care.data is finally rolled out. If care.data is successful, then it will collate GP records and link to hospital records on a national scale, significantly increasing the volume and depth of data available for research and other uses. In time, wider linkage to other information, such as social care, dental records, and biobanks, will progress. There is also a need for one set of guidelines that will promote consistency across the NHS. There is no doubt that this will be challenging, but these issues must be overcome in order for the benefits of EPRs in research to be seen. A review is needed of what is done at the general practice level in order to get to the root cause of these challenges so that workable processes and guidelines can be put in place. Patients and the general public need to be educated regarding the benefits of research and to understand the safeguards that are in place. There also needs to be transparency about what data may be used without consent for research, and patients will need to be assured that data used will not be identifiable.

Williams et al. suggest that dynamic consent could be the way forward, as this provides a participant-centred approach to consent, enabling patients to readily provide or withdraw
their consent over time. The procedure for dynamic consent also allows patients to be provided with details of what their data is to be used for. Kaye et al.\textsuperscript{91} define dynamic consent as both a specific project and a wider concept that offers a new approach to consent, and one that is designed to meet the needs of the twenty-first century research landscape. Kaye et al. further explain that at the heart of dynamic consent is a personalised, digital communication interface that connects researchers and participants, placing participants at the heart of decision-making. The dynamic consent approach facilitates two-way communication, which should in turn stimulate a more engaged, informed and scientifically literate participant population, where individuals can tailor and manage their own consent preferences.

The practicalities of this include the ability for patients to consent to new projects with ease or to alter their consent preferences in real time as their circumstances change. Access for changes to contact information and personal preferences, and to find out how their samples have been used, would be via a secure, password-protected interface. Patients can set preferences about the kind of information they receive, how often they receive it, and in what format (e.g. via text messages, emails and letters). In turn, researchers can customise the interface to meet the needs, resources and capabilities of the research enterprise. The stand-alone online communication interface could also be integrated with other information systems. While dynamic consent appears to be a good solution, it too comes with its own challenges and will require a shift in the current mind-set of both the public and clinicians. It would also require everyone using this method to be ‘technology savvy’, which may exclude the older population. Both Williams et al.\textsuperscript{90} and Kaye et al.\textsuperscript{91} agree that calls for such a cultural change will pose ethical questions concerning social exclusion. A prototype has actually been developed by the UK Ensuring Consent and Revocation Project, known as the EnCoRe project, which was developed in the context of three biobanks, the Oxford Radcliffe Biobank, the Oxford Musculoskeletal Biobank, and the Oxford Biobank.\textsuperscript{92}

There will not be one straightforward solution to resolving the issue of privacy as far as patients are concerned, and any procedure will never be a ‘one hat fits all’ solution. The processes that will enable patients to feel more confident that their data is safe and will help them embrace EPRs, will have to be tailored to suit the needs of individuals. Many clinical researchers have faced consent challenges because they use a standard informed consent template and expect it to work globally. Established NHS standards are needed, but there must be room for flexibility to allow for cultural adaptation. If individuals feel listened to and feel that a well-thought-out process is in place to obtain their consent, then they are much
more likely to be open to giving their consent. The issue of consent must thus be solved by looking at a more tailored approach that suits the research, and most especially the patient.

There is also sufficient evidence to suggest that healthcare practitioners themselves could benefit from better guidance, standards, and some help regarding interpretation of the privacy laws. Currently, healthcare practitioners remain confused about when consent is considered to be implied, or when it can be waived on the grounds of public interest.\textsuperscript{95} It is recommended that informed consent should be sought for any medical research, whether it involves direct contact with participants or simply the ability to access their records. The law, however, is not clear. Both the law of confidentiality and the Data Protection Act allow circumstances in which personal health information may be accessed and used for medical research without explicit consent or full anonymisation. Since the Human Rights Act 1998 came into force, privacy interferences of this kind are allowed if the research investigates an important question, i.e. the research is in the public interest and is of a public nature. However, the Data Protection Act does not clearly define what this could possibly mean and how clinicians are to measure the value of ‘public interest’ or even the meaning of ‘public nature’.\textsuperscript{93,94} Consequently, clinicians are seeking guidance which the DoH and bodies like the Wellcome Trust are constantly providing. Anonymisation and pseudonymisation will continue to be the way forward to protect confidentiality, but this does not detract from the fact that study/research designs need to be carefully considered and require early discussions with data protection bodies. Physicians may also need to be part of working groups or committees looking at the legal framework concerning medical research.

The major issue that patients and the general public have is a lack of understanding of how their data is used, the need to decide whether or not it can be used, and who exactly will see the data. This appears to be the major problem associated with EPRs and is why they have not been fully implemented nationally; it must be accepted that securing the data, even after consent has been obtained, is very important to patients. Standard security processes will have to be improved upon to avoid security breaches that lead to large fines for the NHS. Clear instructions are needed on who can read or amend data, and there needs to be tight management to ensure that if a person no longer has reason to access patient information then they no longer can. There must be a responsible person who decides who should or should not have access to the list of people who have been given access to the data and a clinician on the access control list must be marked as being responsible. Patients would be more confident if they knew who can access their records, so the responsible clinician must share with the
patients the names on the access control list and update patients when there are new additions to the list or when someone else takes over their care.

The ability to delete information from a patient’s record must also be strictly controlled. No one should be able to have the ability to delete clinical information until an appropriate time period has expired. An audit trail must also be kept of all deletions and all accesses must be carefully recorded. As with many things controlled by human beings, there will be the potential for errors. It has been known for patients to have incorrect information entered into their records, and patients must have the ability to check their record from time to time to ensure that the information contained is accurate. Systems that hold EPRs must allow for all the functionality that keeps patient confidentiality well protected, and must be frequently audited to ensure that it meets the required standards.

There is still a lot of work to be done regarding the implementation of EPRs nationally and it will take years to get to the point where a system is fully optimised; however, there will still be the need to modify processes to ensure that they meet the needs of the evolving population and advances in technology. The work on care.data is being carefully thought through following the historical mistakes in previous projects. The four pathfinder areas (Blackburn with Darwen, Somerset, Hampshire, and Leeds) are still under review, with patient materials, such as the practice tool kit, being the focus. The amount of work that has been put into this project is clearly showing that the initial plan to roll it out was far from realistic. According to information provided by the NHS in 2015, Blackburn with Darwen was ready to start fair processing (the time when patients have to make a decision as to whether to opt out) at the end of June 2015, while Somerset and West Hampshire was aiming to start at the beginning of September 2015, and Leeds had not confirmed when it would commence testing communications but was also working towards the beginning of September 2015. A total of 104 practices across Blackburn with Darwen, Somerset and West Hampshire have signed up to care.data. At the time of writing, formal accountability for proceeding with the programme still sits with the Senior Responsible Officer, Tim Kelsey, and until Dame Fiona Caldicott advises she is satisfied with the programme’s proposals and safeguards, no data will be shared. There were plans to undertake a data extraction between September and November 2015, but this would depend on how far processing testing communications had progressed. GP systems were undertaking the work required for extraction to begin to take place.
The complexities concerning how the NHS functions may also need to be looked at to ensure the smooth working of a system like the planned care.data. The NHS as an organisation is becoming very complicated in terms of its institutional fragmentation, and it is now far from a single public section entity. Transferring data between the sections of the NHS may mean trying to work with distinct legal entities, processes, and standard operating procedures. Trust preferences may further complicate matters and trying to harmonise processes between different divisions of a hospital is very complex, let alone a giant organisation such as the NHS.

8.1.2 Bibliometric Analysis of Primary Care Databases

A bibliometric analysis was conducted to demonstrate that electronic health records are being used in research and the longitudinal growth in the number of publications that harness three primary care databases was examined. The review looked at publications using data from the three main UK primary care databases, CPRD, QResearch and THIN. Other databases derived from primary care data do exist, for example the PACT database and QMAS, but were not included in this review because they do not contain data based on individual patient records. A more recently developed database, ResearchOne, was also excluded because it is only now being used for research.

The review clearly showed an increase in the number of studies conducted using data from all three databases over time. The combined number of publications based on data from these three databases from 2004 to 2013 was 1,296 publications. The publications covered a large number of specialty areas, providing evidence of the widespread usage of this data. CPRD data was found to be most commonly used by researchers conducting pharmacology studies, implying its importance for pharmaco-epidemiological research and its inclusion in the UK medicines regulator. These studies accounted for 27% of the publications over the 20-year period reviewed for CPRD. In contrast, QResearch attracted researchers who conducted more studies in the general and internal medicine specialty area, which accounted for 45% of the publications for QResearch data. Finally, the THIN database was used for pharmaco-epidemiology and drug safety studies, which accounted for 22% of the publications generated from THIN data. Overall, publications from all three databases demonstrated that researchers from varying specialty areas are showing a keen interest in the use of electronic health databases for research, and over time this is likely to increase. Studies conducted in the past have also highlighted that electronic health databases, such as CPRD, promote scientific
productions in many ways, for example studies conducted by Chen et al. and William et al.96,97

The clinical impact of these databases can be significant, as studies based largely on CPRD data and entirely from database research have contributed to the evidence for management, investigation, and referral draft consultation documents published by NICE, as described by Kousoulis et al.98 The wealth of information available in primary care electronic health databases is invaluable, and plays a key role in healthcare improvement in the UK. Studies that have provided useful information for many disease areas have used data available from both primary and secondary databases. The use of the data available in these databases has increased significantly over time and more recently, as seen in this review, has attracted further international interest. This is evident in the authorship of the publications available from such studies. These studies have provided clinicians with useful tools to help make the best decisions for the care of their patients.

Monitoring bodies have also been beneficiaries of the data contained in these databases because of the surveillance information that they provide. With the ability of CPRD to now link to other databases, such as HES - outpatient and admitted, mortality data from the Office of National Statistics (ONS), cancer registry data from the National Cancer Intelligence Network (NCIN), cardiovascular disease registry data from the Myocardial Ischemia National Audit Project (MINAP), and socioeconomic data from the Lower Layer Super Output Area (LSOA) level, the benefits are expanding.99,100,101

The complete success of electronic health databases and their impact on research is entirely dependent upon the quality of the data that has been entered into them.102 Maintaining the accuracy, consistency and completeness of the data they contain has always been a challenge, and one that is still seeking a solution. Evidence of the accuracy and validity of the data from clinical databases is mixed, and varies between clinical areas, individual databases and the use of data.103,104,105 GPs are the gatekeepers of this data and will need to be fully committed to reporting data, increasing the quality of data reported, and ensuring that patients’ privacy is well protected.106

8.1.3 Bibliometric Analysis of a Secondary Care Database

HES is primarily an administrative database that collects detailed records of each episode of hospital care received by a patient in NHS England. Its uses include supporting payment to
hospitals, benchmarking performances against other trusts, and as a source of data for researchers. This review looked at the research output of HES over the past 19 years and the analysis shows that there has been slow but consistent growth in the use of HES data for research, with an output of just one publication in 1996 increasing significantly to 534 publications by 2014.

The review showed that HES data is widely used by researchers in different speciality areas, with in total 44 speciality areas being represented. Researchers from different backgrounds deem data captured in HES to be valuable and sufficiently informative to contribute to their field. A large proportion of HES data publications were identified in journals such as the British Medical Journal, the British Journal of Surgery and the Journal of Public Health; however, the majority of journals appeared to have published between one and two articles, and 93 journals published an article using HES data just once.

There are some known limitations to this study such as exclusion of some HES publications. As only studies extracted from PubMed were used in this study, there could be other HES publications listed in non-PubMed indexed journals that may have been overlooked. Some of the articles extracted could also have been published in non-peer-reviewed journals. Using publications extracted from PubMed alone, however, was still sufficient to indicate the steady growth in the use of HES data for research purposes. Significant growth over the last five years is also indicative of the fact that as technology advances and the ability to share, store and transfer data becomes easier, this will lead to a rise in popularity of this method of research. It seems that the use of large clinical databases for research is becoming increasingly common worldwide.107,108,109,110

With HES now able to link to other data, such as ONS mortality data,111 then this presents further opportunities for researchers because this linkage captures deaths outside hospital. HES is now also linked to patient reported outcome measures (PROMS), which contain data from questionnaires completed by patients before and after hip replacement, knee replacement, groin hernia and varicose vein surgery.112 HES data are now also linked to patient records in the CPRD, which pulls in data from primary care, thereby extending the scope of the data available to researchers. The comprehensive nature of HES, with the large amount of data it holds, makes it a strong research tool, but its potential can only be harnessed if the quality of the data it holds is complete and accurate, and if it is linked to other systems that can make the data it contains even more robust.
Concerns have been raised about the lack of involvement or engagement of clinicians in the process of data collection. Clinicians who enter data require a lot of encouragement to do so and Williams et al. \(^{113}\) questioned the validity of HES because they believed that there was no uniformity in the quality of data it provided. The authors concluded that this lack of uniformity was because physicians were not sufficiently engaged in the process of data collection for HES. It does seem, though, that there have been improvements in HES data quality in recent years,\(^{114}\) and bodies such as the Royal College of Physicians (RCP) are committing to training physicians in order for them to be more engaged with HES and other clinical and administrative databases.\(^{115}\) Although coding staff in hospitals are very effective at accurately coding and entering information, the information that clinicians provide in patient notes and discharge summaries can often be incomplete or unclear for the purposes of coding. This has been cited as a possible ‘weak link’ in the data quality chain.\(^{115}\) The RCP is committed to training physicians in this discipline and recommends local sharing of clinically useful analyses with doctors to help deliver benefits and improve data quality. Hospital trusts also have a part to play by collaborating with clinicians when there is the submission of data to SUS for HES, and Spencer\(^{116}\) highlighted that clinicians need regular access to data with an interface that is easy to use. The HSCIC and the Academy of Medical Royal Colleges (AOMRC) have established a joint initiative to help clinicians improve the quality of data entered in HES and in 2011 it set three major goals towards attaining good quality data in HES,\(^{117}\) which was certainly a step in the right direction.

### 8.2 Research Limitations

Certain limitations of this study need to be acknowledged.

- The publications obtained in each database varied in amount. The CPRD, being the largest of the three, contained more publications; over 90% more when compared to QResearch and 46.1% more than THIN. The THIN database when compared to QResearch contained 82% more publications.
- It was also possible that some publications could have been missed during extraction from the Web of Science. THIN publications were primarily obtained from the Department of Primary Care & Population Health, University College London (UCL) and then checked against the Web of Science. Publications that could not be verified via the Web of Science had to be dismissed. These publications perhaps could have
increased numbers of certain speciality areas or at least given a higher number of research outputs.

- Blanket categorisations (speciality areas) of the Web of Science were used, which may not always be reflective of a publication’s true audience. Furthermore, publications published in non-peer-reviewed journals are not accounted for in this work.
- The research could have been more robust if a survey had been conducted to look at the perception of patients concerning EPRs. This was decided against because there was thought to be enough research to show what the general concerns are regarding EPRs.
- Finally, another large primary care database, ResearchOne, was not included in this study, as it has only recently been established.

In spite of these limitations, the growth in publications derived from these databases is clear. The use of electronic health databases in this manner remains invaluable and researchers are beginning to realise their benefits. The increase in publication for the combined databases, from 74 in 2004 to 1296 in 2013, is clearly indicative that this method of research is becoming widely popular. It also highlights the importance of perfecting any limitations that this research method may present. Opportunities to use patient records for secondary uses are also increasing with advances in technology allowing routinely collected data to be easily stored and shared.

8.3 Contribution to Science

I have been able to show from this research project that the combined impact of the four databases reviewed is quite high, and research using data from these databases has increased over the years. There have been numerous reviews and research studies on databases like CPRD and HES, but not on the THIN database or QResearch, looking at their impact on research but not on the progressive impact of these databases, neither has there been any work to compare the research outputs of these databases with each other. This work thus contributes to research in this field and provides a good overview of the current systems in place within the NHS, as well as the attempts that have been made towards achieving a single central national system in the UK.
8.4 Recommendations

The review of the progress of how electronic health records advance research remains an ongoing process. It would be beneficial to be able to assess the ongoing growth of the healthcare databases CPRD, THIN, QResearch with the addition of ResearchOne. Bibliometric analysis of the HES database was conducted from 1996 to 2014, and rapid growth was shown in the last 3 years from 2012 to 2015, from 69 to 113 publications. It would be beneficial to see if there has been significant growth in the research output beyond 2015. An increase in studies using databases such as those looked at in this study is bound to be seen in future years, as the drive for big data continues and if the plan to move the UK towards a single system, such as care.data, is successful and data begins to be uploaded, then it will be beneficial to conduct studies to see if the objectives are being met.

The NHS lists the benefits of care.data as follows\textsuperscript{119} and these provide a useful range of benefits to measure against.

Associated benefits of care.data:

- Improved monitoring of outcomes when primary data and secondary data are linked. This would help healthcare professionals on the route to and from diagnosis and help with the better care of patients.
- Improved monitoring of performance. This would help the NHS to monitor the types of patients that are using A&E frequently and determine if they could have been seen in general practices.
- Early diagnosis of illness. Extracting data of this nature would help identify the number of patients recorded on QOF diseases registers compared with the expected prevalence.
- Improving the contribution of primary care to CCG outcomes. care.data would give the opportunity for correlations between the outcomes at individual practices and the CCG.
- Improved data quality. care.data has the ability to understand more about ethnic categories and referrals within GP records.
- Monitoring and understanding of trends. Trends in A&E attendance, unplanned admissions and readmissions can be thoroughly monitored and understood.
- Help ensure that NHS organisations receive the correct payments for the services they provide.
Once the functions of care.data have been established and there is understanding of its benefits and risks, then areas of improvement can be identified. As with every system which exists today, improvements and enhancements can be made. No perfect system can be created, but one that meets the current needs of the patients, public and healthcare professionals will be a good place to start.

8.5 Conclusions

The review of publications that harnessed data from CPRD, QResearch and THIN showed strong evidence that these electronic healthcare databases are promoting scientific research in many ways, and the growth in publications shows that researchers are now conducting more studies using these databases and are beginning to realise their full potential. To continue to promote academic research, GPs need to continue to provide complete and accurate data, and set standards are needed to encourage GPs’ enthusiasm and willingness to enter the required data.

The use of HES data for research is increasing, and while the 534 publications produced within the 19-year period does suggest steady and consistent growth, concerns exist regarding the accuracy and completeness of the data entered. Clinicians are continuously being urged to become engaged and realise the clinical benefits of HES. If clinicians do become more engaged with HES data and there is more trust in the accuracy of the data, then its use for research is likely to increase further, particularly as HES becomes linked to additional data sources, such as PROMS, CPRD and mortality data, together with technological advancements which streamline the process of extracting data for prospective researchers. If HES-based analyses are accurate and incorporated into an existing quality improvement framework, alongside the clinical databases themselves, then clinicians would be able to assess and see for themselves how useful these datasets are in the delivery of improved quality data where HES is concerned.

Electronic health records for research purposes can indeed advance healthcare services. Evidence can be seen in the many initiatives that are already in place to further promote EPRs and the research output from databases such as HES and CPRD, but there are many hurdles that need to be overcome. Reviewing work from various sources has highlighted that cohesive guidelines, set processes, and an understanding of the legal framework within the healthcare system remain missing or are insufficient. Researchers are often unclear regarding what their limitations are, as they face various obstacles in trying to carry out their research,
while GPs are sometimes unaware of the processes they need to follow when information from a patient record is to be made available for research. The lack of standards and processes creates confusion around how patient data should be handled and processed.

All these issues will need to be addressed but there is no doubt that advancing EPRs is the way forward, otherwise the UK will be left behind as the rest of Europe works towards the EU Commission mandate. Public confidence will have to be strengthened by good and credible technical developments, thereby enhancing the possibilities for operating high quality health information systems that can also be used for research, whilst at the same time providing a high standard of data protection and security.
REFERENCES

1) Robert Hoyt, Benefits of switching to an electronic health record (EPR) – Available From URL http://www.practicefusion.com/health-informatics-practical-guide/


19) Care.data guide for GP practices. Available from URL:

20) The Royal College of Practitioners (RCGP), the British Medical Association (BMA) & HSCIC – Joint Statement. Available from URL:
https://www.google.co.uk/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwjTkei2drNAhXD8RQKHT0FCfgQFggcMAA&url=https%3A%2F%2Fwww.bma.org.uk%2F-%2Fmedia%2FFFiles%2FPDFs%2FPPractical%2520advice%2520at%2520work%2FEt hics%2FCare%2520data%2Fcaredatointstatement.pdf&usg=AFQjCNHuVuUMlItipnjkwBPYPyDMS-UFjRFA&bvm=bv.126130881,d.d24

21) Care.data How did it go so wrong? Available from URL:
http://www.bbc.co.uk/news/health-26259101


23) Care.data NHS Available from URL:
http://www.nhs.uk/NHSEngland/thenhs/records/healthrecords/Pages/care-data.aspx


28) Care.data will close the gap between outsiders and those in the loop; Available from URL- http://www.hsj.co.uk/comment/darzi-caredata-will-close-the-gap-between-outsiders-and-those-in-the-loop/5067686.article


33) Hospital Episode Statistics (HES): Available from URL http://www.hscic.gov.uk/hes


39) Research One, FAQ. Available from URL- http://www.researchone.org/about/


43) HES for Physicians 2007 Available from URL http://www.uhce.ox.ac.uk/hessepho/reports/HESForPhysicians.pdf


on Electronic Health Records and Their Uses in the United Kingdom: Cross-Sectional Survey.


Section 251 and the Confidentiality Advisory Centre (CAG). Available from URL- http://www.hra.nhs.uk/about-the-hra/our-committees/section-251/


78) Sweden chooses InterSystems for NPO. Available from URL-
   http://www.ehi.co.uk/news/ehi/3582/sweden-chooses-intersystems-for-npo

79) Germany and the challenges of rolling out e-health on a large scale Available from URL 
   out-e-health-a-large-scale.

80) IHE Implementation Case Study: French Electronic Health Record Program. 
   Available From URL 
   http://www.ihe.net/uploadedFiles/Content/case_study_france_EPR.pdf

81) KantTa National Archive Health of Health Information. Available from 

82) eHealth in Denmark eHealth as a part of a coherent Danish healthcare system. 
   Available from URL- http://www.sum.dk/~media/Filer%20-
   %20Publikationer_i_pdf/2012/Sundheds-IT/Sundheds_IT_juni_web.ashx


84) New electronic health record rolls out across Scotland. Available From URL 
   record-rolls-out-across-scotland/#.VlNR0HbhAdU.

85) Craig, J. An evaluation of the impact of the key information summary on GPs and 
   out-of-hours clinicians in NHS Scotland. Scott Med J. August 2015 vol. 60 no. 3126-
   131.

86) Individual Health Records- Available from URL 
   http://www.wales.nhs.uk/nwis/page/52553

87) Individual Health Record continues to benefit Swansea Acute GP service. Available 

88) Walsh, S.H. The clinician’s perspective on electronic health records and how they can 
References


References

111) Hospital Episode Statistics Linkage to the Mental Health and Learning Disabilities Data Set. Available from URL- http://www.hscic.gov.uk/hesmhldds


116) Spencer, A. Clinicians and trusts have a role in data quality. BMJ (Clinical research ed). 2012; 344:e2926.


The UK could significantly enhance its health research capability by making effective use of data from electronic patient records for secondary research. This would have major benefits for many types of research; for example through enhancing the understanding of the natural history and epidemiology of diseases, and optimising healthcare delivery. In recent years, we have already seen many examples of research from electronic patient records that could have had a major impact on healthcare delivery and health policy. These studies include the development of algorithms for measuring cardiovascular risk studies evaluating the impact of pay for performance schemes on health inequalities and the development of case-mix measurement tools for examining medical practice variations.

However, significant concerns remain around the ethical and legal issues of using information from patients’ medical records without their consent. There is also uncertainty amongst general practitioners (GPs) and the other custodians of medical records around what processes should be used when information from patient records is made available for research. This lack of consistency and confidence around how these records should be accessed led to the Wellcome Trust hosting a meeting in May 2008 to address these issues. The report that was published after the meeting was based on three overarching principles: safeguarding patient confidentiality and privacy, the role of the GP and healthcare professional as patient advocate, and the need to improve public awareness and understanding. The report provides advice on best practice for ways in which patient records may be used in research, targeted at GPs and other primary care professionals.

The Wellcome Trust report highlights the importance of protecting patient confidentiality and safeguarding privacy, which requires clearly defined processes and controls on the use of patient data. In the past, it has often been common practice for members of a research team to run searches on GP electronic patient record systems to identify patients who may be eligible for a study. However, this practice would appear to be a breach of the guidance stating that people not directly involved in a patient’s clinical care should not have access to identifiable data about them. Although the idea of an ‘approved researcher’ who is also bound by the same duty of confidentiality as the clinical team and with the same penalties has been suggested, this would also not seem compatible with best practice on confidentiality. Hence, systems that use members of the clinical team or technology-based solutions may be the best way to take these recommendations forwards.

In some parts of England, Comprehensive Local Research Networks, working in collaboration with the Primary Care Research Network, have established schemes whereby practices can be reimbursed for carrying out such searches; and for then either giving research staff anonymised data or writing to patients on their behalf to seek consent to be approached about a study. Thanks to the financial
support from the National Institute for Health Research (NIHR), where these schemes are in place, practices no longer need to bear the financial burden of helping research teams to carry out searches on their electronic patient record system. When combined with training on best practice for accessing records for secondary uses and the importance of accurate clinical coding, these schemes could have benefits for the primary healthcare team as well as the research team.

A second method of giving researchers access to anonymised data for research has been through the use of ‘data warehouses’ or similar systems for holding data in a central repository. The best known examples of this approach are the large primary care databases, such as the General Practice Research Database and QRESEARCH, and the Weekly Returns Service. These databases are extensively used for research and have made major contributions to research in such areas as drug safety and disease epidemiology. We are now seeing locally developed systems for data warehousing being implemented in many primary care trusts. Although initially developed for health service delivery and monitoring, these systems also have great potential for research use, particularly where anonymised data is needed; for example, for assessing the feasibility of a study. Some primary care trusts are also using such systems to develop integrated records to improve service delivery; for example, by linking data from primary care and secondary care for people with diabetes. These linked data sets can bring additional benefits to researchers by adding information from external patient record systems, such as from diagnostic laboratories. Because data extraction is generally automated, clearly defined data sets that do not breach any ethical guidelines can be extracted, with minimal workload or disruption for general practices. In the longer term, we await the rollout of the Research Capability Programme in England and similar schemes in the devolved nations. These programmes will be responsible for data linkage of different NHS patient record systems and for ensuring that only anonymised data is released to researchers, unless consent is obtained to use identifiable patient data.

In the process of using data for secondary purposes, GPs and healthcare professionals must act as an advocate for the patient and retain ultimate responsibility for ensuring confidentiality and appropriate access to data. GPs need to be aware that any research must comply with research governance standards; that is, it should have obtained ethical approval and have been approved by the Research Governance Lead in the local primary care trust, and have an approved sponsor (generally the academic institution which is hosting the study). A key aim for primary healthcare teams is to improve public awareness and understanding of the use of their records in research.8 The development of an NIHR Research Practice Incentive scheme in which general practices will receive financial support for hosting studies on the NIHR portfolio will help facilitate this process. If successful, the new scheme will see up to 30% of practices in England taking part in research studies. As part of implementation, practices need to make clear to patients that they are part of such a scheme, through publicity material on practice leaflets and websites, as well as within the practice itself.

The Wellcome Trust report provides useful guidance to researchers and clinicians on why secondary research using data from electronic records in primary care is important. Combined with advances in NHS Information Technology systems, particularly the Research Capability Programme, and financial support from the NIHR, we need to make this potential a reality and ensure that the UK remains a world leader in primary care informatics.

REFERENCES


6 Mathers N, Perrin N and Watt G. Using patient records from general practice for research. Informatics in Primary Care 2009;17:


CONFLICTS OF INTEREST

Angela Gibson-White is an employee of GSK and is undertaking a PhD on the use of data from electronic patient records for research. Azeem Majeed has received funding for research on electronic patient records from the MRC, EPSRC, NIHR, Wellcome Trust, Commonwealth Fund, and Department of Health.

ADDRESS FOR CORRESPONDENCE

Ms Angela Gibson-White Tel: +44 (0)20 7594 3368 Email: angela.gibson-white08@imperial.ac.uk
APPENDIX 2 – Submitted for Publication: Research Outputs of Primary Care Databases in the United Kingdom: Bibliometric Analysis

Submitted to the Journal of Innovation in Health Informatics, April 2015 - awaiting response

Department of Primary Care and Public Health, Imperial College London W6 8RP, UK
Angela Gibson-White, Doctoral Student
Usama Syed, Medical Student
Fahmida Mannan, Medical Student
Zain Chaudhry, Medical Student
Azeem Majeed, Professor of Primary Care

Kings College London, London WC2R 2LS
Shirin Ahmed, Medical Student

Medicines & Healthcare products Regulatory Agency, Clinical Practice Research Datalink (CPRD), London SW1W 9SZ
Antonis Kousoulis, Head of Business Development

CONFLICTS OF INTEREST
Azeem Majeed has received funding for research on electronic patient records from the MRC, EPSRC, NIHR, Wellcome Trust, Commonwealth Fund, and Department of Health. Antonis Kousoulis is an employee of the Clinical Practice Research Datalink.

ADDRESS FOR CORRESPONDENCE
Ms Angela Gibson-White
Department of Primary Care and Public Health
Reynolds Building
Imperial College London
London W6 8RP
Tel: +44 (0)20 7594 2771

Email: angela.gibson-white08@imperial.ac.uk

Abstract

Background: The data held in electronic health databases is now commonly used by researchers. The United Kingdom has several such electronic health databases derived from primary care records. The three major ones are the ‘Clinical Practice Research Datalink’ (CPRD), ‘The Health Improvement Network’ (THIN) and ‘QResearch’. Over time, research outputs generated from data contained in these databases have increased substantially, but as yet these have not been collectively reviewed.
**Objective:** This study compares research outputs from CPRD, THIN and QResearch to assess growth in publications over a 10-year period (2004-2013). CPRD was also reviewed separately over 20 years as a case study.

**Methods:** Publications that used data from CPRD and QResearch were extracted using the Science Citation Index (SCI) of the Thomson Scientific Institute for Scientific Information (Web of Science). Data for THIN was obtained from University College London and validated in Web of Science. All three databases were analysed for their growth in publications, the speciality areas and the journals in which their data have been published.

**Results:** The three databases collectively produced 1,296 publications over a nine year period, with CPRD representing 63.6% (n=825 papers), THIN 30.4% (N=394) and QResearch 5.9% (n=77). Pharmacoepidemiology and General Medicine were the specialities more often featured. Almost three-quarters of CPRD publications have been published in the past 9 years.

**Conclusion:** There is strong evidence that these databases are facilitating and enhancing scientific research and are growing from year to year. The three databases can be even more powerful research tools if the National Health Service and general practitioners can provide accurate and comprehensive data for inclusion in these databases.

**Introduction:**

Data collected in electronic medical records for a patient in primary care can span from birth to death and can have enormous benefits in improving healthcare and public health, and for research. Several systems exist in the United Kingdom (UK) to facilitate the use of research data generated from consultations between primary care professionals and their patients. General Practitioners play a gatekeeper role in the United Kingdom’s National Health Service (NHS) because they are responsible for providing primary care services and for referring patients to see specialists.

In more recent years these databases have been supplemented (through data linkage) with additional data from areas such as laboratory investigations, hospital admissions and mortality statistics. Data collected in primary care research databases are now increasingly used for research in many areas, and for providing information on patterns of disease [1]. These databases have clinical and prescription data and can provide information to support pharmacovigilance, including information on demographics, medical symptoms, therapy (medicines, vaccines, devices), and treatment outcomes [1]. The major primary care research databases in the UK include the ‘Clinical Practice Research Datalink’ (CPRD), ‘QResearch’ and ‘The Health Improvement Network’ (THIN). For all three systems, the information relating to symptoms, diseases, consultations and other clinical events are recorded using the Read code system. The data made available to researchers are anonymised and strong patient identifiers such as name, address and postcode, date of birth, and NHS number are removed.

The Clinical Practice Research Datalink (CPRD) is jointly funded by the NHS National Institute for Health Research (NIHR) and the Medicine and Healthcare products Regulatory Agency (MHRA) [2]. It is one of the largest databases of longitudinal medical records derived from primary care in the world [3]. The collection of information began in 1987 under the previous name General Practice Research Database (GPRD). GPRD was initially part of the Value Added Medical Products (VAMP), a company that pioneered the design and marketing of a general practice office computer system,
allowing the recording of individual patient medical recording. The database was later transferred to government control. CPRD has been providing nearly 30 years of longitudinal data. As of December 2014, the database contained data for over 13.5 million patients, of which approximately 5.7 million are currently active. As well as primary care data, CPRD now links to a number of other data sets such as ‘Hospital Episode Statistics’ (HES) and mortality data from the Office for National Statistics (ONS). It is increasingly being used to enhance clinical trial efficiency (protocol optimization, feasibility and recruitment), through working with the general practitioners, and can provide data for both industry and academic researchers. Access to the data is subject to protocol approval by the MHRA Independent Scientific Advisory Committee (ISAC). Over 1,500 research reports published in peer-reviewed journals have used data from the CPRD and have had direct impacts on public health and disease speciality areas.

QResearch is a large primary care database derived from the anonymised health records of over 12 million patients. The data currently comes from over 950 general practices using the EMIS clinical computer system that is used throughout the UK. Although the data contains socio-economic details of patients based on their postcode, it does not hold any identifiable data and access to it is only opened to academic researchers who have ethical approval to receive datasets. QResearch has led many projects such as QFlu which was used for monitoring and tracking the prevalence of the swine flu outbreak in 2009, reporting to the Health Protection Agency. One of the limitations of QResearch is that although it has links to external databases such as Hospital Episode Statistics, the anonymisation process in compiling the database means that there is no way to identify patients.

The Health Improvement Network (THIN) is collaboration between two companies; In Practice Systems Ltd (INPS), who developed Vision software that is used by GPs in the UK to manage patient data, and CSD Medical Research UK who then provide access to the data for use in medical research. THIN data collection started in 2003 and over 500 Vision practices have so far joined the scheme. THIN data currently contains the electronic medical records of 11.1 million patients (3.7 million active patients). This covers 6.2% of the UK population. In addition to the main consultations being recorded, most patient data in THIN are linked to postcode-level area-based socioeconomic, ethnicity and environmental indices. The data are based on the patients’ postcodes so that variables at ward level are available. The patient is identified only by a code allocated by the GP system and cannot be identified outside the practice.

Aim:

The aim of this study was to conduct a bibliometric review to analyse the research outputs and the longitudinal growth in the number of publications that harness these three primary care databases; CPRD, QResearch and THIN from 2004 until 2013, and also to look at the growth of CPRD on its own from 1993 to 2013 as a case study.

Methods:

To analyse the impact of the three primary care databases (CPRD, QResearch and THIN), publications using data from each of the databases were extracted and analysed. For CPRD, publications were extracted from 1993 to 2013 (20 year period) using the Science Citation Index (SCI) of the Thomson Scientific Institute for Scientific Information (Web of Science). Conference abstracts and posters were not included in the data. The same method was used to extract QResearch publications from 2004 to 2013 (9 year period). Data for THIN was obtained from the Department of
Primary Care & Population Health, University College London (UCL) and verified using the Web of Science. The data was provided in an excel format which contained details of the author, title of article, the journal in which the article was published, article reference and the year of publication. Publications provided by UCL dated from 2004 to 2013.

The number of times publications were cited, their speciality area and the names of the journals they were published in were then extracted using the Web of Science for all three databases. The speciality areas of publications in all three databases were then categorised into four groups.

Speciality areas were classed as follows for all three databases

- Speciality areas with ≥ 50 publications = Group 1
- Speciality areas with < 50 but ≥ 20 publications = Group 2.
- Speciality areas with publications < 20 but ≥ 10 publications = Group 3
- Speciality areas with < than 10 publications = Group 4

All three databases were assessed for overall growth and their speciality areas were compared for similarities. To fairly compare the three databases, publication analysis had to focus on the overlapping period of data availability, which is 2004 to 2013. The publications in the three databases were compared, looking at the average number of publications per year, individual growth and cumulative growth over the 9 year period. The three “most published in” journals were also recorded across the databases.

Results

The CPRD Database

1,140 publications categorised into 28 speciality areas were extracted for CPRD. Results represented in Table 1 show number of CPRD publications from 1993 to 2013.

Table 1 - Publications published in CPRD between 1993 and 2013

<table>
<thead>
<tr>
<th>CPRD Area Grouping</th>
<th>Speciality Area</th>
<th>Number of Publications</th>
<th>Total Publications</th>
<th>% Publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPRD Group 1</td>
<td>≥50 publications</td>
<td>412</td>
<td>36.1</td>
<td></td>
</tr>
<tr>
<td>CPRD Group 2</td>
<td>&lt; 50 publications, ≥ 20</td>
<td>229</td>
<td>20.1</td>
<td></td>
</tr>
<tr>
<td>CPRD Group 3</td>
<td>&lt; 20 publications, ≥ 10</td>
<td>117</td>
<td>10.3</td>
<td></td>
</tr>
<tr>
<td>CPRD Group 4</td>
<td>&lt;10 publications</td>
<td>382</td>
<td>33.5</td>
<td></td>
</tr>
</tbody>
</table>

The largest speciality areas with > 50 publications, (Group 1), were made up from the following speciality areas: Pharmacology (284 publications), Health Services Research (71 publications), and Public Health (57 publications).

The average number of articles published per year over a 20-year period from 1993 – 2013 was 54, with the most listed in 2013 (144 publications). CPRD has shown consistent growth in publications outputs since its data started being used for research. As expected, there were relatively few publications in the early years. Publications from 1993 to 1997 were 10, 9, 8, 7 and 17 publications.
respectively. This rate picked up in 1998 and 1999, with 30 articles in each year, and it has continued to grow since this point.

Between 2004 and 2013 (9-year period), the total number of publications listed in the CPRD database was 825, which shows significant growth in this period. 72.3% of the CPRD publications were published in the last 9 years. The highest number of publications was published in the Pharmacoepidemiology and Drug Safety journal, which represents 4.5% (52) of the publications in CPRD. Table 5 shows the journals that CPRD papers were most frequently published in.

QResearch

Seventy-seven articles categorised into 13 speciality areas were published from studies conducted with QResearch data between 2004 and 2013 that were extracted from Web of Science. Results are listed in table 2.0 below.

Table 2. Publications published from QResearch between 2004-2013

<table>
<thead>
<tr>
<th>THIN Area Grouping</th>
<th>Number of Publications</th>
<th>Total Publications</th>
<th>% Publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Internal Group 1</td>
<td>≥50 publications</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>QResearch Group 2</td>
<td>&lt; 50 publications, ≥ 20</td>
<td>35</td>
<td>45.5</td>
</tr>
<tr>
<td>QResearch Group 3</td>
<td>&lt; 20 publications, &gt; 10</td>
<td>26</td>
<td>33.8</td>
</tr>
<tr>
<td>QResearch Group 4</td>
<td>&lt; 10 publications</td>
<td>16</td>
<td>20.8</td>
</tr>
</tbody>
</table>

The average publications published per year over the 9-year period is around 8, with most articles published between 2011 (17 publications) and 2012 (20 publications).

The highest number of articles was published in the British Journal of General Practice, which represented 24.6% (19) of the publications in QResearch.

THIN Database

Three hundred-ninety four (394) articles categorised into 32 speciality areas from studies conducted with THIN data between 2004 and 2013 were extracted. Results are listed in table 3.

Table 3. Percentage of publications published in THIN between 2004-2013

<table>
<thead>
<tr>
<th>THIN Area Grouping</th>
<th>Number of Publications</th>
<th>Total Publications</th>
<th>% Publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>THIN Group 1</td>
<td>≥50 publications</td>
<td>167</td>
<td>42.4</td>
</tr>
<tr>
<td>THIN Group 2</td>
<td>&lt; 50 publications but ≥ 20</td>
<td>75</td>
<td>19.0</td>
</tr>
<tr>
<td>THIN Group 3</td>
<td>&lt; 20 publications but ≥ 10</td>
<td>105</td>
<td>28.6</td>
</tr>
<tr>
<td>THIN Group 4</td>
<td>&lt; 10 publications</td>
<td>47</td>
<td>11.9</td>
</tr>
</tbody>
</table>

The largest speciality areas with publications from THIN, with > 50 publications (Group 1) were the following: pharmacology (116) and General Medicine (51). The average number of articles published per year over the 9-year period was around 39, with the most publications listed in 2010 (87
publications). The Pharmacoepidemiology and Drug Safety journal published 21.3% (95) of publications. The remaining 78.7% articles were published in journals that published between 1 and 14 articles. The majority of journals that published articles on THIN data only published once.

**Combined Growth**

Results represented in table 4 and figure 1 showed an increase in publications using data from all three databases. Over the 9 year period, Publications for THIN and QResearch have slowly increased over time. CPRD publications have increased substantially in last 4 years. All 3 databases produced a combined total of 1,296 publications over the 9-year period (2004-2013) with CPRD representing 63.6% (825), THIN 30.4% (394) and QResearch 5.9% (77) of the total collective number of publications across these electronic health databases. The growth across all three databases has been consistent over the 9-year period, with the highest number of publications listed in 2013.

**Table 4. Annual number of publications published per database between 2004 and 2013**

<table>
<thead>
<tr>
<th>Year</th>
<th>CPRD</th>
<th>THIN</th>
<th>QResearch</th>
<th>Total Publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004</td>
<td>70</td>
<td>2</td>
<td>2</td>
<td>74</td>
</tr>
<tr>
<td>2005</td>
<td>68</td>
<td>9</td>
<td>2</td>
<td>79</td>
</tr>
<tr>
<td>2006</td>
<td>61</td>
<td>14</td>
<td>2</td>
<td>77</td>
</tr>
<tr>
<td>2007</td>
<td>61</td>
<td>30</td>
<td>4</td>
<td>95</td>
</tr>
<tr>
<td>2008</td>
<td>72</td>
<td>45</td>
<td>7</td>
<td>124</td>
</tr>
<tr>
<td>2009</td>
<td>85</td>
<td>43</td>
<td>4</td>
<td>132</td>
</tr>
<tr>
<td>2010</td>
<td>69</td>
<td>87</td>
<td>9</td>
<td>165</td>
</tr>
<tr>
<td>2011</td>
<td>94</td>
<td>43</td>
<td>17</td>
<td>154</td>
</tr>
<tr>
<td>2012</td>
<td>101</td>
<td>72</td>
<td>20</td>
<td>193</td>
</tr>
<tr>
<td>2013</td>
<td>144</td>
<td>49</td>
<td>10</td>
<td>203</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>825</strong></td>
<td><strong>394</strong></td>
<td><strong>77</strong></td>
<td><strong>1296</strong></td>
</tr>
</tbody>
</table>

**Figure 1: Cumulative growth of database publications from 2004-2013**
The journal most published in, across the three databases was the Pharmacoepidemiology & Drug Safety journal. As represented in table 5, most publications from CPRD and the THIN database occurred in this journal.

Table 5: Top 3 journals most published in across all 3 databases from 2004-2013

<table>
<thead>
<tr>
<th>Journal</th>
<th>CPRD Publications</th>
<th>QResearch Publications</th>
<th>THIN Publications</th>
<th>Total Publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pharmacoepidemiology &amp; Drug Safety journal</td>
<td>52</td>
<td>4</td>
<td>95</td>
<td>151</td>
</tr>
<tr>
<td>British Medical Journal</td>
<td>19</td>
<td>18</td>
<td>1</td>
<td>51</td>
</tr>
<tr>
<td>British Journal of General Practice</td>
<td>23</td>
<td>19</td>
<td>1</td>
<td>43</td>
</tr>
</tbody>
</table>

Discussion:

This review looked at the publications using data from the three main UK primary care databases: CPRD, QResearch and THIN. Other databases derived from primary care data do exist; such as the Prescribing Analysis and Cost (PACT database) and the Quality Management and Analysis System (QMAS), but they have not been included in this review because they do not contain data based on individual patient records. A more recently developed database, SystmOne, was also excluded because it is only now being started to be used for research.

This review clearly showed an increase in studies conducted with data from all three databases over time. The combined number of publications based on data from these three databases from 2004 to 2013 was 1,296 publications. The publications covered a large number of specialty areas, providing evidence of the widespread usage of this data. CPRD data was found to be used most commonly by researchers conducting studies in the Pharmacology specialty area, implying its importance for pharmacoepidemiological research and its inclusion in the UK medicines regulator. This accounted for 27% of the publications over the 20 year period reviewed for CPRD. QResearch attracted researchers who conducted more studies in the General and Internal Medicine specialty area. This accounted for 45 % of the publications for QResearch. Finally, THIN database showed that researchers in the Pharmacoepidemiology and drug safety specialty area conducted more studies using its data, accounting for 22% of the publications generated from THIN data. Overall, publications from all three databases showed that researchers from varying specialty areas showed a keen interest in the use of electronic health databases for research and, over time, this is bound to increase.

However, some limitations do exist with this study because the publications obtained in each database varied in amount. The CPRD, being the largest of the three, contained more publications; over 90% more when compared to QResearch and 46.1% more than THIN. The THIN database when compared to QResearch contained 82% more publications. It was also possible that some publications could have been missed during extraction from Web of Science. The THIN publications were primarily obtained from the Department of Primary Care & Population Health, University College London (UCL) and checked against Web of Science. Also, the blanket categorisations (specialty areas) of the Web of Science were used, which may not always be reflective of a publication’s true audience. Further, publications published in non-peer reviewed journals would not be accounted for in this
work. Finally, another primary care database, ResearchOne, was not included in this study, as it has only recently been established.

In spite of these limitations, the growth in publications derived from these databases is clear. The use of electronic health databases in this manner remains invaluable and researchers are beginning to realise their benefits. The increase in publication for the combined databases from 74 in 2004 to 1296 in 2013 is clearly indicative that this method of research is becoming widely popular. It also highlights the importance of perfecting any limitations that this research method may present. Opportunities to use patient records for secondary uses are also on the increase with advances in technology allowing routinely collected data to be easily stored and shared.

Studies conducted in the past have also highlighted that electronic health databases such as CPRD will promote scientific productions in many ways (11). For example the clinical impact of these databases can be significant; namely studies based largely on CPRD data and entirely from database research, have contributed to the evidence for management, investigation, and referral draft consultation document by the National Institute for Health and Care Excellence (NICE) (3, 12). The wealth of information available in electronic health databases from primary care and their uses is invaluable, and plays a key role in healthcare improvement in the UK. Studies that have provided useful information for many disease areas have used data available from both primary and secondary databases. The use of the data available in these databases has increased significantly over time and more recently, as seen in this review, has attracted more international interest. This is evident in the authorship of publications available from such studies. These studies have provided researchers with great tools to help make the best decisions for the care of their patients.

Monitoring bodies have also been beneficiaries of the data contained in these databases because of the surveillance information that they provide. With the ability of CPRD to now link to other databases such as the Hospital Episode Statistics (HES) - Outpatient and Admitted, Death data from Office of National Statistics (ONS) Mortality data, Cancer registry data from the National Cancer Intelligence Network (NCIN), Cardiovascular disease registry data from the Myocardial Ischemia National Audit Project (MINAP) and Socioeconomic data at the Lower Layer Super Output Area (LSOA) level, the benefits are expanding (14, 15, 16).

The complete success of electronic health databases and the impact on research is completely dependent on the quality of the data that has been entered into them (17). Accuracy, consistency and completeness of the data they contained have always been a challenge, one that is still seeking a solution. Evidence for the accuracy and validity of the data from clinical databases is mixed and varies between clinical areas, the individual databases and the use of data (6, 18, 19 20). General Practitioners will need to be fully committed to reporting data, increasing the quality of data reported and ensuring patient’s privacy are well protected by being gatekeepers of the data (21).

Conclusion

Based on the review of publications that harness data in CPRD, QResearch and THIN, there is strong evidence that these electronic healthcare database are promoting scientific research in many ways, their growth in publications have shown that researchers are now conducting more studies using these databases and are beginning to realise their full potential. To continue to promote academic research, general practitioners will need to continue to provide complete and accurate data; set standards will
also need to be provided to General Practitioners to encourage enthusiasm and willingness to enter the required data.

References:


9) Julie Glanville, Tony Kendrick, Rosalind McNally, John Campbell, Research output on primary care in Australia, Canada, Germany, the Netherlands, the United Kingdom and the United States: Bibliometric analysis. BMJ 2011; 342 10.1136


APPENDIX 3 – Submitted for Publication: Research Outputs of England’s Hospital Episode Statistics (HES) Database: Bibliometric Analysis

Submitted to the Journal of Innovation in Health Informatics, April 2015 - awaiting response

Department of Primary Care and Public Health, Imperial College London, London W6 8RP, UK
Angela Gibson-White, Doctoral Student
Usama Syed, Medical Student
Fahmida Mannan, Medical Student
Zain Chaudhry, Medical Student
Azeem Majeed, Professor of Primary Care

Kings College London, London WC2R 2LS
Shirin Ahmed, Medical Student

ADDRESS FOR CORRESPONDENCE
Ms Angela Gibson-White
Department of Primary Care and Public Health
Imperial College London. Faculty of Medicine
W6 8RP.
Tel: +44 (0)20 7594 2771

Email: angela.gibson-white08@imperial.ac.uk

ABSTRACT

Background: Hospital administrative data, such as that provided by the Hospital Episode Statistics (HES) database in England, are increasingly being used for research and quality improvement. To date, no study has yet tried to quantify and examine trends in the use of HES for research purposes.

Objective: To examine trends in the use of HES data for research.

Methods: Publications generated from the use of HES data were extracted from PubMed and analysed. Publications from 1996 to 2014 were then examined further in the Science Citation Index (SCI) of the Thompson Scientific Institute for Science Information (Web of Science) for details of research speciality area.

Results: 534 studies, categorised into 44 speciality areas, were extracted from PubMed. The review showed an increase in publications over the 18 year period with an average of 28 publications per year. The highest number of publications was in the General Surgery specialty area.

Conclusion: The use of HES data for research is becoming more common. Increase in publications over time shows that researchers are beginning to take advantage of the potential of HES data. Although HES is a valuable database, concerns exist over the accuracy and completeness of the data.
entered. Clinicians need to be more engaged with HES for the full potential of this database to be
harnessed.

INTRODUCTION

Hospital administrative data, such as that provided by the Hospital Episode Statistics (HES) database in
England, are increasingly being used for research and quality improvement. The Hospital Episode
Statistics database was developed in 1987 and was designed to collect a detailed record of each
episode of care received by a patient, either from NHS inpatient providers or independent sector
providers commissioned by the NHS to deliver care.

HES data are compiled from data sent by over 400 NHS trusts and foundation trusts, including acute
hospital trusts, primary care trusts (PCTs) and mental health trusts in England. Independent sector
organisations also send data to HES for activity commissioned by the English NHS. The data is stored
as a large collection of separate records, one for each episode of care, in a secure data warehouse. The
Health and Social Care Information centre (HSCIC) liaises closely with NHS organisations to
ensure that data submitted is complete and accurate.

The HES database houses data containing patient details such as: age, sex, NHS number and location
of residence; episode of care (Hospital name, GP referral, emergency/elective admission details;
clinical information (Diagnosis, operative procedures, consultants). HES data are largely designed for
non-clinical purposes, including hospital remuneration following delivery of care to patients. Some
other secondary uses of the data available in HES include: monitoring trends and patterns in hospital
activity, general medical research and statistical functions, developing and monitoring government
policies as well as providing the basis for national health indicators of clinical quality such as hospital
readmission rates to examine health outcomes and to improve patient experience.

HES is not a live system and it usually takes between 9 and 12 months for a full financial year to be
available through on-line HES or the safe haven extract. So, for example, complete 2014/15 data
would not be available until around December 2015. Similar databases to HES exist in Wales,
Northern Ireland and in Scotland. HES is increasingly being used by researchers due its
comprehensive coverage of NHS-funded hospital activity in England. To date, no study has yet tried
to quantify or examine trends in the use of HES for research.

The aim of this study was to analyse the research outputs and the growth in the number of
publications using the Hospital Episode Statistics (HES) database in England.

METHODS

To review the impact of HES data on research, publications generated from the use of HES data from
1996-2014 (19 year period), were extracted from PubMed and analysed. Extracted publications were
examined in the Science Citation Index (SCI) of the Thomson Scientific Institute for Scientific
Information (also known as Web of Science) for details of the publication’s speciality area as assigned
by the Web of Science.

Speciality areas were categorised into 4 groups.

- Speciality areas with \( \geq 50 \) publications = Group 1
Speciality areas with < 50 but > 20 publications = Group 2
Speciality areas with ≤ 20 articles but ≥ 10 articles = Group 3
Speciality areas with < 10 articles = Group 4

The average number of publications per year was calculated and cumulative growth over the 19 year period was also looked at, as well as the journals that were most published in.

RESULTS

Publication Growth

534 studies that used HES data between 1996 and 2014 were extracted from PubMed. As listed in Table 1, there was an increase in the number of publications over the 19 year period, with the largest growth in the last 5 years. The average number of publications over the 19-year period was 28 publications per year. The greatest increase in this period was between 2013 and 2014; 73 and 113 publications respectively. All 534 publications were categorised into a total of 44 publication groups. As listed in Table 2, speciality areas were categorised into 4 groups, depending on the number of publications listed under them.

Speciality Areas

Group 1- (≥ 50 publications)

Group 1 was made of 102 publications, with the General Surgery speciality area accounting for 52 publications and the Heath Statistics speciality area having 50 publications assigned to it.

Group 2- (<50 but >20 publications)

Group 2 accounted for 205 publications, with the majority of publications falling into this category. Speciality area examples included Oncology (39), Infectious Diseases (38), Orthopaedic Surgery (31), Vascular Surgery (29), Obstetrics & Gynaecology (28), Paediatrics (22) etc.

Group 3- (<20 but ≥10 publications)

Group 3 accounted for 143 publications. Speciality area examples included Gastroenterology (20), Ophthalmology (17), Cardiology (16), Urology (15), Neurology (15), Public Health (14) etc.

Group 4-(< 10 publications)

Group 4 accounted for 84 publications. Speciality area examples included Respiratory (8), Pharmacology (6), Nephrology (6), Dentistry (5), Haematology (4) Hepatology (3) etc.

There was 1 non-medical speciality area represented under Management; this was from a publication looking at the management of large-scale change in the NHS.
Table 3 details three of the “most published-in journals”. The British Medical Journal published more articles than all other journals of HES publications. Most journals, however, published just one article, with 93 examples of this in our review.

**DISCUSSION**

HES is primarily an administrative database that collects detailed records of each episode of hospital care received by a patient in England’s NHS. Its uses include supporting payment to hospitals; benchmarking performances against other trusts and as a source of data for researchers. This review looked at the research output of HES over the past 19 years. Our analysis shows that in the last 19 years there has been a slow but consistent growth in the use of HES data for research, with an output of just one publication in 1996 increasing significantly to 534 publications by 2014.

The review showed that HES data is widely used by researchers in different speciality areas, with 44 speciality areas being represented. Researchers of different backgrounds deem data captured in HES to be valuable and educative enough to contribute to their field. Journals such as the British Medical Journal, British Journal of Surgery and the Journal of Public Health published a large proportion of the HES data publications, but the majority of journals appeared to have published between one and two articles. 93 journals published an article just once.

There are some known limitations with this study such as exclusion of some HES publications. As only studies extracted from PubMed were used in this study, there could be other HES publications listed in non-PubMed indexed journals that may have been overlooked. Some of the articles extracted could also have been published in non-peer reviewed journals. Using publications extracted from PubMed alone, however, is still sufficient to indicate the steady growth in the use of HES data for research. Significant growth over the last 5 years is also indicative of the fact that, as technology advances and the ability to share, store and transfer data becomes easier, this will lead to a rise in popularity of this method of research. It seems that the use of large clinical databases for research is becoming increasingly common worldwide.

With HES now able to link to other data such as the Office for National Statistics (ONS) mortality data this gives even more of an opportunity to researchers because the linkage captures deaths of people in the HES database who died outside of hospital. HES is now also linked to Patient Reported Outcome Measures (PROMS), which contains data from questionnaires completed by patients before and after hip replacement, knee replacement, groin hernia and varicose vein surgery.

HES data are now also inked to the patient records in the Clinical Practice Research Database (CPRD) which pulls in data from primary care, extending the scope of available data to researchers. The comprehensive nature of HES, bestowed by the large amounts of data that it holds, makes it a strong research tool; but its potential can only be harnessed if the quality of the data it holds is complete and accurate. Concerns have been raised about the lack of involvement or engagement of clinicians in the process of data collection. Clinicians who enter the data still need a lot of encouragement to do so. J.G Williams et al in their article “Hospital Episode Statistics: time for clinicians to get involved” questioned the validity of HES because they believed that there was no uniformity in the quality of data it provided. They concluded that this lack of uniformity was because physicians are not sufficiently engaged in the process of data collection for HES. It does seem, though, that there have
been improvements in HES data quality in recent years\textsuperscript{12}. These improvements have continued over the years and bodies such as the Royal College of Physicians are committing to training physicians in order for them to be more engaged with HES and other clinical and administrative databases \textsuperscript{13}. Although coding staff in hospitals are very effective at accurately coding and entering information, the information clinicians provide in patient notes and discharge summaries can often be incomplete or unclear for the purposes of coding. This has been cited as a possible ‘weak link’ in the data quality chain\textsuperscript{13}. The Royal College of Physicians are committed to training physicians in this discipline and recommend that there be local sharing of clinically useful analyses with doctors to help deliver benefits and improve data quality. Hospital trusts also have a part to play by collaborating with clinicians when there is submission of data to secondary user services for HES. Clinicians still need regular access to data with an interface that is easy to use \textsuperscript{20}. The Health and Social Care Information Centre (HSCIC) and Academy of Medical Royal Colleges (AOMRC) have a joint initiative to help clinicians improve the quality of data entered in HES \textsuperscript{22}.

**CONCLUSION**

The use of HES data for research is increasing. The 534 publications produced within the 19-year period does suggest steady and consistent growth, but concerns still exist about the accuracy and completeness of the data entered. Clinicians are continuously being urged to get engaged and realise the clinical benefits of HES. The Academy of Medical Royal Colleges (AOMRC) is working towards improving the quality of data in HES and in 2011 set out three major goals towards improving the quality of HES data\textsuperscript{6}. If clinicians do get more engaged with HES data and there is more trust in the accuracy of the data, then its use for research is likely to increase further, particularly as HES becomes linked to additional data sources such as PROMS, CPRD and mortality data and technological advancements streamline the process of extracting data for prospective researchers.

**TABLES & FIGURES**

**Table 1 Number of publications using HES data between 1996 and 2014**

<table>
<thead>
<tr>
<th>Year</th>
<th>Publication</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996</td>
<td>1</td>
</tr>
<tr>
<td>1997</td>
<td>3</td>
</tr>
<tr>
<td>1998</td>
<td>8</td>
</tr>
<tr>
<td>1999</td>
<td>2</td>
</tr>
<tr>
<td>2000</td>
<td>7</td>
</tr>
<tr>
<td>2001</td>
<td>5</td>
</tr>
<tr>
<td>2002</td>
<td>10</td>
</tr>
<tr>
<td>2003</td>
<td>12</td>
</tr>
<tr>
<td>2004</td>
<td>19</td>
</tr>
<tr>
<td>2005</td>
<td>12</td>
</tr>
<tr>
<td>2006</td>
<td>19</td>
</tr>
<tr>
<td>2007</td>
<td>40</td>
</tr>
<tr>
<td>2008</td>
<td>27</td>
</tr>
<tr>
<td>2009</td>
<td>30</td>
</tr>
<tr>
<td>2010</td>
<td>39</td>
</tr>
<tr>
<td>2011</td>
<td>45</td>
</tr>
</tbody>
</table>
Table 2 Speciality Areas of Publications Using HES data between 1996 and 2014

<table>
<thead>
<tr>
<th>Articles Per Group</th>
<th>Total Articles in Speciality Area</th>
<th>% Publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgery &amp; Health Statistics - Group 1 (≥50 articles)</td>
<td>102</td>
<td>19.1</td>
</tr>
<tr>
<td>Group 2 (&lt; 50 articles but ≥ 20)</td>
<td>205</td>
<td>38.3</td>
</tr>
<tr>
<td>Group 3 (≤ 20 articles but ≥ 10)</td>
<td>143</td>
<td>26.7</td>
</tr>
<tr>
<td>Group 4 (&lt; 10 articles)</td>
<td>84</td>
<td>15.7</td>
</tr>
</tbody>
</table>

Table 3 Journals with highest number of HES data Publications.

<table>
<thead>
<tr>
<th>Journals</th>
<th>Publications</th>
<th>% Publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>British Medical Journal</td>
<td>48</td>
<td>9.0</td>
</tr>
<tr>
<td>British Journal of Surgery</td>
<td>29</td>
<td>5.4</td>
</tr>
<tr>
<td>Journal of Public Health</td>
<td>20</td>
<td>3.7</td>
</tr>
</tbody>
</table>

REFERENCES


13. Royal College of Physicians, September 2006. Engaging clinicians in improving data quality in the NHS; Key findings and recommendations from research conducted by the Royal College of Physicians’ iLab . Available from https://www.rcplondon.ac.uk/sites/default/files/ilab-summary-report_1.pdf


