Using nationwide 'big data' from linked electronic health records to help improve outcomes in cardiovascular diseases: 33 studies using methods from epidemiology, informatics, economics and social science in the ClinicAl disease research using LInked Bespoke studies and Electronic health Records (CALIBER) programme

Harry Hemingway,[1,2]* Gene S Feder,[3]
Natalie K Fitzpatrick,[1,2] Spiros Denaxas,[1,2]
Anoop D Shah[1,2] and Adam D Timmis[2,4,5]

[1]Institute of Health Informatics, University College London, London, UK
[2]Farr Institute of Health Informatics Research, University College London, London, UK
[3]Centre for Academic Primary Care, School of Social and Community Medicine, University of Bristol, Bristol, UK
[4]Barts Health NHS Trust, London, UK
[5]Farr Institute of Health Informatics Research, Queen Mary University of London, London, UK

*Corresponding author

Declared competing interests of authors: none

Published January 2017
DOI: 10.3310/pgfar05040

Scientific summary

Big data to improve outcomes in cardiovascular diseases
Programme Grants for Applied Research 2017; Vol. 5: No. 4
DOI: 10.3310/pgfar05040

NIHR Journals Library www.journalslibrary.nihr.ac.uk

Scientific summary

A digital trace in diverse electronic health records (EHRs) is left by everyone. The UK is arguably the only country in the world with both nationwide EHRs in primary care and ongoing national quality registries in cardiovascular disease (CVD). When linked, such resources have the potential to provide new population-based insights into the quality and outcomes of care.

Coronary heart disease (CHD) is the main cause of preventable, premature mortality in the world; however, there are unacceptable variations in prevention, investigation and treatment within and between health-care systems. However, large-scale outcome studies using samples of nationwide data across the 'patient journey' are lacking. CHD is one disease area in which large-scale, diverse data ('big data') exist and may provide new opportunities for understanding quality of care and outcomes. This programme is among the first to realise the research potential of linking big data across primary and secondary care and disease registries.

Aim

To use nationwide linked EHR approaches to identify opportunities to improve quality of care and outcomes for patients at risk of or with stable coronary artery disease (SCAD) and acute coronary syndrome (ACS) across the patient journey.

Our overall approach was to build an informatics platform and then address key questions at different stages in a person's lifetime to reduce the risk of a major cardiovascular event: from initially healthy populations to initial presentation with chest pain, to the initial acute admission and subsequent phases of management of this long-term condition. We report on 33 inter-related studies (25 of which are published in full elsewhere) which used research designs and methods from epidemiology, health informatics, health economics and ethnographic social science approaches. Our research settings include 230 NHS hospitals and 226 general practices in England and Wales. Our research participants include up to 2 million initially healthy adults, 100,000 people with SCAD and up to 300,000 patients with ACS.

Chapter 2: establishing a framework to study cardiovascular diseases across the patient journey – underlying methodology and approaches

Background

Challenges to harnessing linked EHRs include obtaining data in a research-ready format and the compatibility of information across different sources, and there are methodological weaknesses in prognosis research.

Objectives

To develop a platform for studying CVDs based on linked EHRs (studies 1 and 2). To establish a conceptual framework promoting high-quality prognosis research (study 3).

Methods

We created a novel research platform [ClinicAl disease research using LInked Bespoke studies and Electronic health Records (CALIBER)] based on linkage of four major sources of EHR data, including primary care and national registry data, and curated > 600 EHR disease and risk factor phenotyping algorithms (see www.caliberresearch.org/) (studies 1 and 2). We established the PROGnosis RESearch Strategy (PROGRESS) partnership (study 3).

Results
We created a linked data set encompassing nearly 2 million adults (studies 1 and 2). The PROGRESS partnership highlighted multiple methodological shortfalls in prognosis research and made recommendations to raise standards (study 3).

Conclusions
The CALIBER platform enables the wealth of information in UK EHRs to be exploited for research purposes. This resource, and recommendations for improving the quality of prognosis research, provides an important framework for high-quality research.

Chapter 3: inequalities in the incidence of 12 cardiovascular diseases across the patient journey

Background
Investigations of the demographic determinants of CVD have focused on acute myocardial infarction (AMI), stroke and coronary death. Little is known about their association with other common CVDs.

Objectives
To identify how 12 distinct CVD phenotypes are affected by demographic factors (study 4). To examine the complex relationships between incidence and prognosis of CHD among people of South Asian ethnicity (study 5).

Methods
Analyses of CALIBER linked data and a meta-analysis.

Results
We identified inequalities in the rates of initial presentation and lifetime risks of different CVDs by age, sex, deprivation and ethnicity (study 4). Incidence of stable CHD among South Asian patients was markedly higher than among white patients, although their prognosis was significantly better (study 5).

Conclusions
Our findings are crucial for understanding the global burden of CVDs and argue against the use of composite measures of cardiovascular health.

Chapter 4: risk factor modification among people without established cardiovascular disease

Background
Reported failures to meet national guideline targets and exploit treatment opportunities to optimise risk reduction among people without established disease may represent important missed opportunities to protect people against CVD.

Objectives
To investigate the association between type 2 diabetes mellitus and blood pressure (BP) and 12 CVDs (studies 6 and 7). To identify whether or not targets for risk factor control are being met among people at risk of CVD (study 8).

Methods
Analysis of CALIBER linked data.

Results

Associations between type 2 diabetes and 12 CVD outcomes differed in terms of direction and magnitude of effects, and there was substantial heterogeneity in associations with BP across CVDs and ages (studies 6 and 7). We identified a continuing failure to meet guideline targets for BP control (study 8).

Conclusions

Important opportunities to reduce cardiovascular risk are being missed, including more vigorous clinical management to meet BP targets.

Initial presentation with stable chest pain

Chapter 5*: initial presentation with stable chest pain: opportunities for earlier diagnosis*

Background

Delayed diagnosis represents a missed opportunity for treatment and is common in patients presenting with chest pain.

Objectives

To determine the extent and impact of missed opportunities in diagnosis (studies 9 and 10). To develop automated methods for extracting diagnostic information from EHR free text (studies 11 and 12).

Methods

Analyses of CALIBER data (studies 9 and 10). Development of two free-text data-mining algorithms for extracting clinically meaningful entities from clinical narrative (studies 11 and 12).

Results

More than 70% of people presenting with chest pain in primary care received a diagnosis of unattributed chest pain (studies 9 and 10). More CVD events occurred in this group than in the group of patients diagnosed with angina. Unheralded AMI was uncommon. Algorithms to extract information from free text showed promising performance (studies 11 and 12).

Conclusions

Opportunities to identify patients with angina, and to initiate the prevention of disease, are being missed. Free-text data-mining algorithms may be able to identify early diagnostic indicators in EHRs before a formal diagnosis is recorded.

Chapter 6*: initial presentation with stable chest pain – opportunities using a clinical decision support system – the Optimising the Management of Angina programme*

Background

The diagnosis of patients with stable chest pain is challenging. Rapid access chest pain clinics (RACPCs) may be missing significant numbers of cases.

Objectives

To pilot a multifaceted intervention [the Optimising the Management of Angina (OMA) programme] including a computerised clinical decision support system (CDSS) targeting initial specialist management of patients (study 13).

Methods

Mixed quantitative (before and after) and qualitative study. The OMA CDSS was based on National Institute for Health and Care Excellence guidance and expert panel recommendations.

Results
The OMA programme had a negligible effect on the decision-making of clinicians in RACPCs, indicating that a larger evaluation is not justified. Barriers to implementation included the ambivalence of clinicians towards the guidance, insufficient integration into hospital EHRs and the inability to extract clinical information at scale from hospital information systems.

Conclusions
Decision-making in the chest pain clinic is complex. Our approach provided an in-depth understanding of the ways in which clinicians interacted with the CDSS and could support the development of tools more likely to be adopted in practice.

Stable chest pain and stable coronary artery disease

Chapter 7: *stable chest pain and stable coronary artery disease – a novel cohort (the Clinical Cohort in Coronary disease Collaboration)*

Background
Consented cohorts recruited in clinical settings linked to *-omic* and biomarker data can add value to EHR data in the study of diseases.

Objectives
To establish a novel prognostic cohort of patients with suspected SCAD with linked genetic, biomarker and health data (study 14).

Methods
Consecutive patients undergoing investigation for chest pain at four UK NHS hospitals provided blood samples and completed a baseline health questionnaire. Clinical information was manually extracted from hospital electronic databases. Participants were followed up for cardiovascular events using national registry databases.

Results
A total of 3345 patients were recruited.

Conclusions
The wealth of data in the Clinical Cohorts in Coronary disease Collaboration is a resource that will facilitate research into the causes and consequences of SCAD.

Chapter 8: *stable coronary artery disease – use of single prognostic factors*

Background
Questions remain about the quality of evidence being generated on biomarkers and their clinical usefulness.

Objectives
To evaluate currently used and potentially informative biomarkers as prognostic indicators in SCAD (studies 15–19).

Methods
Analyses of CALIBER data and literature-based evidence syntheses.

Results
Simple blood (haemoglobin) (study 15) and physiological [heart rate (study 16), systolic blood pressure (study 19)] markers were associated with a range of prognostic outcomes. Published evidence of associations

between C-reactive protein levels and cardiovascular events was of variable quality, and the degree of association may have been overstated (study 17). Chromosomal marker Ch9p21 was convincingly associated with an increased risk of a first but not subsequent cardiovascular events (study 18).

Conclusions
Biomarkers may be useful prognostic factors in SCAD, although poor methodology impairs the interpretation of studies.

Chapter 9*: stable coronary artery disease – risk prediction and cost-effective targeting of interventions*

Background
Validated prognostic models and information on resource use drawn from clinical data may assist the management of SCAD patients and policy-making.

Objectives
To develop prognostic models for SCAD patients (study 20). To model health-care resource utilisation and costs in SCAD patients (study 21).

Methods
Prognostic models were developed from CALIBER data and validated with independent data (study 20). Lifetime health-care costs were calculated by applying the risk prediction model to resource use estimates derived from CALIBER (study 21).

Results
Prognostic models based on widely available clinical data could identify patients at risk of poor outcome with a high degree of accuracy (study 20). Long-term costs and resource use associated with SCAD were high (study 21).

Conclusions
Prognostic models have the potential in early stages of risk assessment to identify patients for intensive management or further investigation. Cost-modelling can inform resource allocation decisions and the cost-effectiveness of new treatments.

Acute myocardial infarction

Chapter 10*: acute myocardial infarction – missed opportunities for treatment during and after hospitalisation*

Background
Registry-based research offers a way to evaluate care once disease has manifested.

Objectives
To evaluate missed opportunities for care in hospital and after discharge (studies 22–25). To compare outcomes in post-AMI survivors across four countries (studies 22 and 24).

Methods
Analyses of national registry data from the UK, Sweden, France and the USA.

Results
Rates of prescribing and the use of invasive procedures were lower in the UK than in Sweden or the USA (studies 22 and 24). Short-term mortality was greater in the UK than in Sweden. More than 50% of

eligible ST-segment elevation myocardial infarction patients did not receive nine evidence-based care components. Greater numbers of missed opportunities for care were associated with higher mortality (study 23). Opportunities to encourage smoking cessation were commonly missed after discharge from hospital (study 25).

Conclusions
Despite guideline recommendations, opportunities to improve outcomes for ACS patients are being missed. International comparisons may identify key areas that could be targeted to improve outcomes.

Chapter 11: *acute myocardial infarction – evidence of treatment benefits beyond those provided by clinical trials (further opportunities to improve outcomes)*

Background
Clinical trials cannot provide evidence of long-term effectiveness or the impact of interactions between drugs; these are issues that can be addressed using EHR data.

Objectives
To determine the use of secondary prevention medications after AMI (studies 26–29).

Methods
Analyses of CALIBER data.

Results
Use of fewer than five classes of secondary prevention drugs was associated with increased mortality, although effects varied according to the class of drug omitted (study 26). Beta-blocker use in patients with chronic obstructive pulmonary disease was associated with reduced mortality (study 27). Proton pump inhibitors are unlikely to interfere with the clinical action of clopidogrel (e.g. Plavix®, Sanofi-Aventis, Gentilly, France) (study 28). Discontinuation of clopidogrel after AMI is common and is associated with adverse outcomes (study 29).

Conclusions
'Real-world' data can provide important information beyond that generated by clinical trials to guide prescribing decisions.

Chapter 12: *acute myocardial infarction – variation between hospitals in short-term mortality outcomes*

Background
Although care and outcomes vary widely between hospitals, previous studies have not accounted for case mix, compared national health-care systems or evaluated processes of care in depth in the UK using qualitative methods.

Objectives
To investigate the extent of, and reasons for, interhospital variation in mortality after admission for ACS (studies 30–33).

Methods
An analysis of UK and Swedish national registry data (studies 30–32). An ethnographic study in 10 UK hospitals (study 33).

Results
Thirty-day mortality varied substantially between UK hospitals (study 30). About 1 in 20 hospitals had an excess mortality of $\geq 40\%$, and up to four seriously underperforming hospitals could be identified with

confidence. We found no clear relationship between the number of patients admitted to individual hospitals or the quality of BP management in primary care and 30-day mortality rates for AMI patients (study 31). Interhospital variation in 30-day mortality was greater in the UK than in Sweden, with the lowest mortality in hospitals adhering to guideline-recommended procedures and treatments (study 32). Five processes were associated with high-quality care (study 33).

Conclusions

Underperformance was not wholly explained by differences in case-mix or treatment factors. Qualitative work may be able to identify aspects of care giving, contributing to variation in outcomes.

Overall conclusions and recommendations for research

Linking national EHRs at different stages across the patient journey offers increasing opportunities to understand and improve diagnosis, risk stratification and cost-effective treatment of people at risk of or with coronary disease.

Using CALIBER, we addressed questions about missed opportunities across the patient journey, identifying areas where, despite national guidance, suboptimal clinical management results in a worse prognosis for patients. Through international comparisons, we show that these missed opportunities contribute at least in part to differences in mortality between countries.

Our findings suggest that a systematic drive to improve quality of care may have more impact than measures targeting the poorest-performing institutions.

We illustrated how linked EHR data can provide real-world evidence supplementing data obtained in clinical trials on combinations of secondary prevention medications.

Our novel consented cohort study highlighted the challenges of extracting information from hospital informatics systems. Our biomarker studies suggest that some prognostic factors, including simple clinical measures, have the potential to improve targeting of clinical interventions in SCAD.

Our economic modelling highlighted the financial implications of growing numbers of SCAD patients. As well as cost projections, our models provide a way to assess the probable cost-effectiveness of new interventions.

Although these findings are in CVD, many might easily be extended in approach to any disease, both acute and chronic, and both common and rare.

Nonetheless, the majority of NHS data remain inaccessible to research and this hampers efforts to improve efficiency and the quality of care, and to drive innovation.

We propose three priority directions for further research. First, there is an urgent need to ‘unlock’ more detailed data within hospitals, to expand the coverage of the UK’s 65 million people with primary care data available for research, and to expand and make sustainable the range of record linkages (e.g. to a wider range of disease registries). Second, there should be major expansion in the underlying methods and applications of scaled approaches to using EHRs to design and carry out trials, and interpret the implementation of trial results. This is important for delivering learning health systems and patient benefit. Third, large-scale, disease agnostic genetic and biological collections linked to such EHRs are required in order to deliver precision medicine and to innovate the discovery of new drug targets or to repurpose existing drugs.

Study registration

CALIBER studies are registered as follows: study 2 – NCT01569139, study 4 – NCT02176174 and NCT01164371, study 5 – NCT01163513, studies 6 and 7 – NCT01804439, study 8 – NCT02285322, and studies 26–29 – NCT01162187. OMA is registered as Current Controlled Trials ISRCTN54381840.

Funding

The National Institute for Health Research (NIHR) Programme Grants for Applied Research programme (RP-PG-0407-10314) (all 33 studies) and additional funding from the Wellcome Trust (study 1), Medical Research Council Partnership grant (study 3), Servier (study 16), NIHR Research Methods Fellowship funding (study 19) and NIHR Research for Patient Benefit (study 33).

Programme Grants for Applied Research

ISSN 2050-4322 (Print)

ISSN 2050-4330 (Online)

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: nihredit@southampton.ac.uk

The full PGfAR archive is freely available to view online at www.journalslibrary.nihr.ac.uk/pgfar. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Programme Grants for Applied Research* journal
Reports are published in *Programme Grants for Applied Research* (PGfAR) if (1) they have resulted from work for the PGfAR programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Programme Grants for Applied Research programme

The Programme Grants for Applied Research (PGfAR) programme, part of the National Institute for Health Research (NIHR), was set up in 2006 to produce independent research findings that will have practical application for the benefit of patients and the NHS in the relatively near future. The Programme is managed by the NIHR Central Commissioning Facility (CCF) with strategic input from the Programme Director.

The programme is a national response mode funding scheme that aims to provide evidence to improve health outcomes in England through promotion of health, prevention of ill health, and optimal disease management (including safety and quality), with particular emphasis on conditions causing significant disease burden.

For more information about the PGfAR programme please visit the website: http://www.nihr.ac.uk/funding/programme-grants-for-applied-research.htm

This report

The research reported in this issue of the journal was funded by PGfAR as project number RP-PG-0407-10314. The contractual start date was in January 2009. The final report began editorial review in March 2015 and was accepted for publication in April 2016. As the funder, the PGfAR programme agreed the research questions and study designs in advance with the investigators. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The PGfAR editors and production house have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the final report document. However, they do not accept liability for damages or losses arising from material published in this report.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).