International Journal of Medical Informatics 113 (2018) 17–25

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf

Integrating electronic healthcare records of armed forces personnel: Developing a framework for evaluating health outcomes in England, Scotland and Wales

Daniel Leightley[a,*], Zoe Chui[a], Margaret Jones[a], Sabine Landau[b], Paul McCrone[c], Richard D. Hayes[d], Simon Wessely[a,e], Nicola T. Fear[a,e], Laura Goodwin[f]

[a] King's Centre for Military Health Research, Institute of Psychiatry, Psychology & Neuroscience, King's College London, United Kingdom
[b] Biostatistics & Health Informatics, Institute of Psychiatry, Psychology & Neuroscience, King's College London, United Kingdom
[c] Health Services & Population Research, Institute of Psychiatry, Psychology & Neuroscience, King's College London, United Kingdom
[d] Psychological Medicine, Institute of Psychiatry, Psychology & Neuroscience, King's College London, United Kingdom
[e] Academic Department of Military Mental Health, Institute of Psychiatry, Psychology & Neuroscience, King's College London, United Kingdom
[f] Department of Psychological Sciences, University of Liverpool, United Kingdom

ARTICLE INFO

Keywords:
Hospital episode statistics
Electronic health records
Hospital admission
Secondary care
Big data
Data linkage

ABSTRACT

Background: Electronic Healthcare Records (EHRs) are created to capture summaries of care and contact made to healthcare services. EHRs offer a means to analyse admissions to hospitals for epidemiological research. In the United Kingdom (UK), England, Scotland and Wales maintain separate data stores, which are administered and managed exclusively by devolved Government. This independence results in harmonisation challenges, not least lack of uniformity, making it difficult to evaluate care, diagnoses and treatment across the UK. To overcome this lack of uniformity, it is important to develop methods to integrate EHRs to provide a multi-nation dataset of health.

Objective: To develop and describe a method which integrates the EHRs of Armed Forces personnel in England, Scotland and Wales based on variable commonality to produce a multi-nation dataset of secondary health care.

Methods: An Armed Forces cohort was used to extract and integrate three EHR datasets, using commonality as the linkage point. This was achieved by evaluating and combining variables which shared the same characteristics. EHRs representing Accident and Emergency (A&E), Admitted Patient Care (APC) and Outpatient care were combined to create a patient-level history spanning three nations. Patient-level EHRs were examined to ascertain admission differences, common diagnoses and record completeness.

Results: A total of 6,336 Armed Forces personnel were matched, of which 5,460 personnel had 7,510 A&E visits, 9,316 APC episodes and 45,005 Outpatient appointments. We observed full completeness for diagnoses in APC, whereas Outpatient admissions were sparsely coded; with 88% of diagnoses coded as "*Unknown/unspecified cause of morbidity*". In addition, A&E records were sporadically coded; we found five coding systems for identifying reason for admission.

Conclusion: At present, EHRs are designed to monitor the cost of treatment, enable administrative oversight, and are not *currently* suited to epidemiological research. However, only small changes may be needed to take advantage of what should be a highly cost-effective means of delivering important research for the benefit of the NHS.

Abbreviations: A&E, accident and emergency; APC, admitted patient care; CHI, community health index; DOB, date of birth; EHRs, electronic healthcare records; HES, health episode statistics; ISD, information services division; ICD-10, international classification of diseases, 10th revision; KCMHR, King's College Centre for Military Health Research; NHS, National Health Service; OPCS-4, OPCS classification of interventions and procedures version 4; SAIL, secure anonymised information linkage; UK, United Kingdom

* Corresponding author.
E-mail addresses: daniel.leightley@kcl.ac.uk (D. Leightley), zoe.e.chui@kcl.ac.uk (Z. Chui), margaret.jones@kcl.ac.uk (M. Jones), sabine.landau@kcl.ac.uk (S. Landau), paul.mccrone@kcl.ac.uk (P. McCrone), richard.hayes@kcl.ac.uk (R.D. Hayes), simon.wessely@kcl.ac.uk (S. Wessely), nicola.t.fear@kcl.ac.uk (N.T. Fear), laura.goodwin@liverpool.ac.uk (L. Goodwin).

https://doi.org/10.1016/j.ijmedinf.2018.02.012
Received 7 June 2017; Received in revised form 13 February 2018; Accepted 17 February 2018

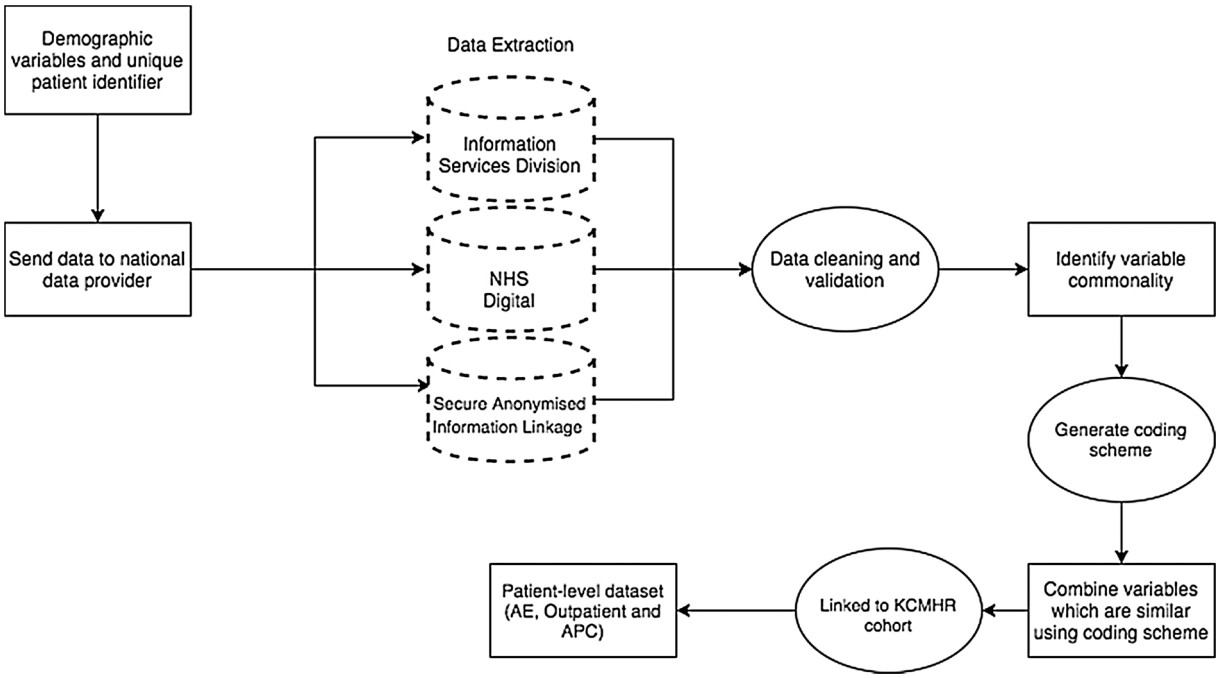

Fig. 1. Overview of the data linkage process and formation of patient-level dataset.

1. Background

Routinely collected Elxectronic Healthcare Records (EHRs) can be used to evaluate disease prevalence, tendencies, and to perform epidemiological analyses [1]; investigate quality of care and to improve clinical decision-making, which can influence patient outcome and care [2,3]. In recent years, there has been a growth in the use of EHRs in the field of Big Data analytics, which is the analysis and evaluation of large datasets, to answer specific research questions [4,5]. The term "Big Data" [6] has been used in a number of health settings, it is synonymous with the meaningful analyses of EHRs to identify health movements and associations [7]. Further, Schneeweiss et al. [8] summarises the potential applications into two key areas: generation of knowledge to improve the effectiveness of treatment; and to predict the outcome of treatment and diagnoses.

Globally, the development and use of EHRs is increasing, alongside Big Data innovation in a number of fields [8–10]. Standardised EHR systems are difficult to implement in larger countries, those which have complex political structures or multiple private entities [11], such as the United States. In the United Kingdom (UK), health and social care is devolved to national Government and local agencies using propriety systems which are not interconnected or able to identify patients who relocate from a nations [11–13]. In the UK, it is estimated that 53,000 individuals migrate from England and Wales to Northern Ireland and Scotland; with 46,800 individuals migrating from Northern Ireland and Scotland to England and Wales each year [14]. At present, migration statistics are not reported separately, making it difficult to determine cross-border migration. EHRs which represent the same patient across the three nations show great promise and could be used to identify risk, inform healthcare policy, service provision and improve health and social outcomes [13,15,16]. Using a cross-border cohort might provide the foundation of creating such a system. While several studies have sought to create national datasets of health and social care [4,17], to our knowledge, no studies exist which integrate EHRs from multiple nations into a single repository for research.

There are multiple challenges in integrating EHRs in the UK. Firstly, there is no system or framework which uniquely identifies a person (*e.g.* unique ID number for each citizen) across public services such as welfare, housing, education and health [18]. Those registered with the National Health Service (NHS) of England and Wales are assigned a unique 10-digit number, which is used as the sole identifier in healthcare [19]. Similarly, in Scotland a person is assigned a Community Health Index (CHI) number which is used in the same manner. Second, EHRs provided by England [20], Wales [4,17] and Scotland [21] are distinctive in structure, collection and management. Third, EHR data from each of these data sources include various data types, from structured information such as drug prescriptions, diagnoses and treatment, to unstructured data such as clinical notes and patient self-reported illness [11]. Finally, there is a dearth of literature which describes how to undertake EHR data linkage from multiple data sources [22].

Current research efforts have been directed towards modelling and predicting single conditions (*e.g.* depression, diabetes and epilepsy) and outcomes [3,16,23,24]. In this work, we study a UK Armed Forces cohort [25,26] (which includes individuals from England, Scotland and Wales) that has been linked separately to three sources of national secondary healthcare data. These data have then been integrated to form a patient-level dataset across 3 nations which include A&E, Outpatient Care and Admitted Patient Care (APC). The subsequent dataset is unique, population-based longitudinal dataset that contains a wide range of physical and mental health indicators of serving and ex-serving UK Armed Forces personnel. The objectives of this work are; 1) to develop and describe a methodology which integrates EHRs of Armed Forces personnel in England, Scotland and Wales 2) examine healthcare utilisation within these data and 3) evaluate characteristics of admission, record completion and diagnoses.

2. Methods

The proposed framework is illustrated in Fig. 1, and is represented by the following stages:

1. Demographic identifiers of Armed Forces personnel are sent to each data provider;
2. EHRs are extracted and sent to the research team by data providers;
3. EHRs are cleaned and validated;
4. Variables are linked using commonality;
5. EHRs for each patient are integrated to generate a patient-level