ELSEVIER

Journal of Clinical Epidemiology ■ (2016) ■

Journal of Clinical Epidemiology

BRIEF REPORT

Improved incidence estimates from linked vs. stand-alone electronic health records

Elizabeth R.C. Millett[a,*], Jennifer K. Quint[b], Bianca L. De Stavola[c], Liam Smeeth[b], Sara L. Thomas[a]

[a]*Department of Infectious Disease Epidemiology, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK*

[b]*Department of Non-communicable Disease Epidemiology, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK*

[c]*Department of Medical Statistics, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, UK*

Accepted 4 January 2016; Published online xxxx

Abstract

Objective: Electronic health records are widely used for public health research, and linked data sources are increasingly available. The added value of using linked records over stand-alone data has not been quantified for common conditions such as community-acquired pneumonia (CAP).

Study Design and Setting: Our cohort comprised English patients aged ≥65 years from the Clinical Practice Research Datalink, eligible for record linkage to Hospital Episode Statistics. Stand-alone general practice (GP) records were used to calculate CAP incidence over time using population-averaged Poisson regression. Incidence was then recalculated for the same patients using their linked GP-hospital admission data. Results of the two analyses were compared.

Results: Over 900,000 patients were included in each analysis. Population-averaged CAP incidence was 39% higher using the linked data than stand-alone data. This difference grew over time from 7% in 1997 to 83% by 2010. An increasingly larger number of pneumonia events were recorded in the hospital admission data compared to the GP data over time.

Conclusion: Use of primary or secondary care data in isolation may not give accurate incidence estimates for important infections in older populations. Further work is needed to establish the extent of this finding in other diseases, age groups, and populations.

Keywords: Pneumonia; Electronic health records; Data linkage; Aged; England/epidemiology; Cohort

1. Introduction

Electronic health records are extensively used in epidemiological research, because of their wide and detailed population coverage. It is increasingly possible to link electronic data sources to enhance available data. For example, linked primary and secondary care data provide more complete information on outcomes, enriched data on covariates such as patients' medical and therapeutic histories, and accurate timing of events such as hospitalizations. The value of linked over stand-alone data has been investigated for conditions such as cardiovascular events, asthma, diabetes, and upper gastrointestinal bleeding [1–4]. However, the potential benefits of linked data for examining the burden of important infectious diseases are unclear.

Community-acquired pneumonia (CAP) causes considerable morbidity among older individuals and can be treated in either primary or secondary care. Large-scale

Funding: This report is independent research arising from a Career Development Fellowship supported by the National Institute for Health Research (http://www.nihr.ac.uk), awarded to Dr Thomas (grant number CDF 2010-03-32). The views expressed in this publication are those of the authors and not necessarily those of the UK National Health Service, the National Institute for Health Research, or the Department of Health. J.K.Q. is funded on a Medical Research Council Population Health Scientist Fellowship (grant number G0902135). L.S. is supported by a Wellcome Trust Senior Research Fellowship in Clinical Science (grant number 098504/Z/12/Z). The funders of the study had no role in study design, data collection and analysis, decision to publish, or preparation of the article.

Conflict of interest: None.

Ethics information: All data were anonymized before receipt by the authors. Ethics approval for the study was given by the Independent Scientific and Advisory Committee (of CPRD), and the London School of Hygiene and Tropical Medicine Ethics Committee.

This work is an updated version of that presented orally at the Scottish Health Improvement Network (SHIP) meeting 2013, and in poster form at London School of Hygiene and Tropical Medicine Open Day 2014.

* Corresponding author. Tel.: +0044 1865 612911.

E-mail address: elizabeth.millett@ndm.ox.ac.uk (E.R.C. Millett).

http://dx.doi.org/10.1016/j.jclinepi.2016.01.005

What is new?

Key findings

- Use of linked primary-secondary care health data provided markedly higher incidence estimates of community-acquired pneumonia compared to stand-alone general practice (GP) records for the same group of English older adults.
- Comparison of the data sources revealed diverging incidence estimates over time, rising from 7% higher in 1997/98 to 83% higher in 2010/11 when using the linked data compared to the stand-alone GP data.

What this adds to what was known?

- The benefits of the use of linked electronic health records (compared to single data sources) have been demonstrated for conditions such as cardiovascular diseases; this is the first article to demonstrate the benefits for an important, common infection.

What is the implication and what should change now?

- Use of primary or secondary care data in isolation may not give accurate estimates of burden of disease for important infections in older populations.
- Further work is needed to establish if this trend is seen in other infections and diseases.

studies of CAP incidence trends have commonly used either stand-alone general practice (GP) records, potentially excluding patients who present to hospital if practices record hospitalized events suboptimally, or stand-alone hospital records which exclude cases treated in the community. Two recent studies used large linked GP and hospital data sets to assess disease burden of CAP but did not assess the added value of using the linked data [5,6].

We thus investigated the utility of linked primary/secondary care data in better determining trends in CAP disease burden in England among those aged ≥65 years by comparing incidence of CAP derived from stand-alone primary care data with that from linked primary-secondary care data. Each analysis used essentially the same cohort of patients over the same time period, using the same analytical approach.

2. Methods

The Clinical Practice Research Datalink (CPRD) is a nationally representative UK primary care dataset, containing a range of information including Read-coded diagnoses [1]. Hospital Episode Statistics (HES) contain inpatient records with ICD10-coded diagnoses, including admission and discharge dates. CPRD and HES records are linked at a patient-level for consenting English practices. By March 2011, CPRD contained >12 million patient records, with HES-linkage available for 65% of English CPRD practices (around 5% of the English population) [7].

Practices and patients joined CPRD throughout the study period, providing dynamic cohorts of patients. To ensure comparability of the two data sources, a near-identical group of patients were used in both analyses. Patients included in the study were eligible for record linkage, were aged ≥65 years, and contributed ≥1 day of follow-up. Follow-up started at the latest of the study start date (April 1, 1997), the patient's 65th birthday, the date the practice met CPRD quality standards or 28 weeks after patient registration (to exclude historical illnesses retrospectively reported) [6]. Follow-up ended at the earliest of the study end date (March 31, 2011), death, the practice's last data collection date, or the date the patient left the practice.

We have previously described in detail definitions for pneumonia illness episodes in CPRD and HES, using pneumonia and other lower respiratory tract infection records [6]. In brief, records for which pneumonia was recorded in CPRD (stand-alone and linked data) or as the admitting diagnosis (primary code of the first episode) in HES (linked data only) within 28 days of each other or of a record for lower respiratory tract infection were considered to be part of the same episode. The incident date of the episode was the date of the first of these pneumonia codes.

In both analyses, pneumonia illness episodes which started ≤14 days after a hospitalization were assumed to be hospital-acquired (HAP) and were excluded; episodes with no such hospitalization record were classed as community acquired. The method for defining hospitalizations, and thus distinguishing between CAP and HAP, differed between the two analyses. In the stand-alone CPRD data, hospitalization records were identified using Read codes and other relevant fields in the GP files. In the linked cohort, the 14-day period started at the discharge date of any hospital admission.

Patients were not considered ''at-risk'' of pneumonia during any pneumonia episode (CAP or HAP) or for 28 days after the last record in the episode, and this time was excluded from the denominator in both cohorts. A key difference in the linked data analysis was the capacity to also exclude the duration of any hospital admission and the subsequent 14 days from person-time at risk of a community-acquired infection and thus obtain more accurate denominator data. This was not possible in the stand-alone data as hospital admission, and discharge dates were not available.

Population-averaged Poisson models were used to calculate the incidence of CAP across clusters of CAP episodes per patient. Rates were calculated stratified by year, age group, and sex.