



Big Data – Electronic Health Records

Overview, uses & challenges

Ian Douglas – Assoc Prof in Pharmacoepidemiology

DNCDE

Elizabeth Williamson – Assoc Prof in Medical Statistics

Medical Statistics

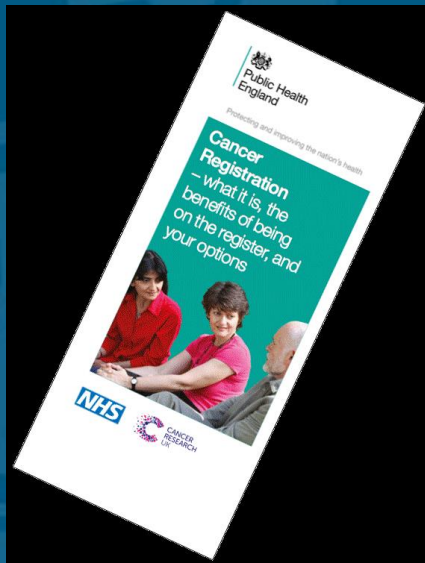
LONDON
SCHOOL of
HYGIENE
& TROPICAL
MEDICINE



What data?



HES: a national treasure



LONDON
SCHOOL of
HYGIENE
& TROPICAL
MEDICINE



BMJ Open Population trends in the 10-year incidence and prevalence of diabetic retinopathy in the UK: a cohort study in the Clinical Practice Research Datalink 2004–2014

Rohini Mathur,¹ Krishnan Bhaskaran,¹ Emma Edwards,² Helen Lee,² Nishi Chaturvedi,³ Liam Smeeth,¹ Ian Douglas¹

To cite: Mathur R, Bhaskaran K, Edwards E, *et al.* Population trends in the 10-year incidence and prevalence of diabetic retinopathy in the UK: a cohort study in the Clinical Practice Research Datalink 2004–2014. *BMJ Open* 2017;7:e014444. doi:10.1136/bmjopen-2016-014444

► Prepublication history and additional material is available. To view please visit

ABSTRACT

Objectives: To describe trends in the incidence and prevalence of diabetic retinopathy (DR) in the UK by diabetes type, age, sex, ethnicity, deprivation, region and calendar year.

Design: Cohort study using the Clinical Practice Research Datalink (CPRD).

Setting: UK primary care.

Participants: 7.7 million patients ≥12 contributing to the CPRD from 2004 to 2014.

Primary and secondary outcome measures:

Age-standardised prevalence and incidence of diabetes, DR and severe DR (requiring photocoagulation) by calendar year and population subgroup. Relative risk of developing DR and severe DR by population subgroup.

Strengths and limitations of this study

- This study constitutes the largest ever sample size to examine trends in the burden of diabetes and diabetic retinopathy (DR) in the UK which allowed for sufficient power to detect relationships between population subgroups, which is often unfeasible in smaller studies where population sizes do not allow for such granular comparisons.
- Since recording of screening of DR was incentivised under Quality and Outcomes Framework (QOF) from 2004 to 2014 and QOF indicators are known to be well recorded by general practitioners and so we anticipate that screening and





Contents lists available at ScienceDirect

Vaccine

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



Impact of the national rotavirus vaccination programme on acute gastroenteritis in England and associated costs averted



Sara L. Thomas^{a,*}, Jemma L. Walker^{a,b}, Justin Fenty^a, Katherine E. Atkins^a, Alex J. Elliot^c, Helen E. Hughes^c, Julia Stowe^b, Shamez Ladhani^b, Nick J. Andrews^{a,b}

^a Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK

^b Immunisation, Hepatitis and Blood Safety Department, Centre for Infectious Disease Surveillance and Control (CIDSC), Public Health England, 61 Colindale Avenue, London NW9 5EQ, UK

^c Real-time Syndromic Surveillance Team, Public Health England, 6th Floor 5 St Philips Place, Birmingham B3 2PW, UK

ARTICLE INFO

Article history:

Received 5 August 2016

Received in revised form 11 November 2016

Accepted 15 November 2016

Available online 20 December 2016

Keywords:

Rotavirus

Gastroenteritis

Vaccination

Electronic health records

Primary health care

ABSTRACT

Background: Introduction of infant oral rotavirus vaccination in the UK in July 2013 has resulted in decreased hospitalisations and Emergency Department (ED) visits for acute gastroenteritis (AGE), for both adults and children. We investigated reductions in AGE incidence seen in primary care in the two years after vaccine introduction, and estimated the healthcare costs averted across healthcare settings in the first year of the vaccination programme.

Methods: We used primary care data from the Clinical Practice Research Datalink and age-stratified time-series analyses to derive adjusted incidence rate ratios (IRR_a) for AGE in the first two years of the post-vaccination era (July 2013–April 2015) compared to the pre-vaccination era (July 2008–June 2013). We estimated cases averted among children aged <5 years in the first year of the vaccination programme by comparing observed numbers of AGE cases in 2013–2014 to numbers predicted from the time series models. We then estimated the healthcare costs averted for general practice consultations



RESEARCH ARTICLE

Phosphodiesterase Type 5 Inhibitors and Risk of Malignant Melanoma: Matched Cohort Study Using Primary Care Data from the UK Clinical Practice Research Datalink

Anthony Matthews, Sinéad M. Langan, Ian J. Douglas, Liam Smeeth, Krishnan Bhaskaran*

Department of Non-Communicable Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, United Kingdom

* krishnan.bhaskaran@lshtm.ac.uk



Abstract

OPEN ACCESS

Citation: Matthews A, Langan SM, Douglas IJ, Smeeth L, Bhaskaran K (2016) Phosphodiesterase Type 5 Inhibitors and Risk of Malignant Melanoma: Matched Cohort Study Using Primary Care Data from the UK Clinical Practice Research Datalink. PLoS

Background

Laboratory evidence suggests that reduced phosphodiesterase type 5 (PDE5) expression increases the invasiveness of melanoma cells; hence, pharmacological inhibition of PDE5 could affect melanoma risk. Two major epidemiological studies have investigated this and



REGULATION (EU) No 1235/2010 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

LONDON
SCHOOL *of*
HYGIENE
& TROPICAL
MEDICINE



REVIEW

Near real-time vaccine safety surveillance using electronic health records—a systematic review of the application of statistical methods†

Andreia Leite^{1*}, Nick J. Andrews² and Sara L. Thomas¹

¹*Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London, UK*

²*Statistics, Modelling and Economics Department, Public Health England, London, UK*

ABSTRACT

Purpose Pre-licensure studies have limited ability to detect rare adverse events (AEs) to vaccines, requiring timely post-licensure studies. With the increasing availability of electronic health records (EHR) near real-time vaccine safety surveillance using these data has emerged as an option. We reviewed methods currently used to inform development of similar systems for countries considering their introduction.

Methods Medline, EMBASE and Web of Science were searched, with additional searches of conference abstract books. Questionnaires were sent to organizations worldwide to ascertain unpublished studies. Eligible studies used EHR and regularly assessed pre-specified AE to vaccine(s). Key features of studies were compared descriptively.

Results From 2779 studies, 31 were included from the USA (23), UK (6), and Taiwan and New Zealand (1 each). These were published/conducted between May 2005 and April 2015. Thirty-eight different vaccines were studied, focusing mainly on influenza (47.4%), especially 2009 H1N1 vaccines. Forty-six analytic approaches were used, reflecting frequency of EHR updates and the AE studied. Poisson-based maximized sequential probability ratio test was the most common (43.5%), followed by its binomial (23.9%) and conditional versions (10.9%). Thirty-seven of 49 analyses (75.5%) mentioned control for confounding, using an adjusted expected rate (51.4% of those adjusting), stratification (16.2%) or a combination of a self-controlled design and stratification (13.5%). Guillain-Barré syndrome (11.9%), meningitis/encephalitis/myelitis (11.9%) and seizures (10.8%) were studied most often.

Conclusions Near real-time vaccine safety surveillance using EHR has developed over the past decade but is not yet widely used. As more countries have access to EHR, it will be important that appropriate methods are selected, considering the data available and AE of interest.



ANALYSIS

Pragmatic randomised trials using routine electronic health records: putting them to the test



OPEN ACCESS

What to prescribe for a patient in general practice when the choice of treatments has a limited evidence base? **Tjeerd-Pieter van Staa and colleagues** argue that using electronic health records to enter patients into randomised trials of treatments in real time could provide the answer

Tjeerd-Pieter van Staa *head of research and honorary professor of epidemiology*^{1,2,3}, Ben Goldacre *research fellow*³, Martin Gulliford *professor of public health*⁴, Jackie Cassell *professor of primary care epidemiology*⁵, Munir Pirmohamed *NHS chair of pharmacogenetics*⁶, Adel Taweel *senior lecturer in software engineering*⁴, Brendan Delaney *Guy's and St Thomas' charity chair in primary care research*⁴, Liam Smeeth *professor of clinical epidemiology*³

¹General Practice Research Database (GPRD), Medicines and Healthcare products Regulatory Agency, 151 Buckingham Palace Road, London SW1W 9SZ, UK; ²Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, the Netherlands; ³London School of Hygiene and Tropical Medicine, London, UK; ⁴Department of Primary Care and Public Health Science, King's College, London, UK; ⁵Division of Primary Care and Public Health, Brighton and Sussex Medical School, University of Brighton, Brighton, UK; ⁶Wolfson Centre for Personalised Medicine, Institute of Translational Medicine, University of Liverpool, Liverpool, UK



Missing/mis-measured data

Messy data!

- Uncaptured exposures
 - e.g. over-the-counter medications, hospital in-patient and out-patient medications
 - Misclassification
- Missing data
 - Empirical research comparing EHR with other data (Marston et al, 2010, PDS)
 - Smoking, alcohol MNAR, Blood pressure MAR
 - Measurement process correlated with health status
- Multiple Imputation
 - Flexible, powerful
 - Combinations of MI and other approaches?
 - Complete-case, missing indicator
 - Repeated irregular measurements - added complexity



Unmeasured confounding

- Lots of uncaptured variables
 - E.g. Diet, lifestyle
- Instrumental Variables
 - Physician preference, distance to facility
- Regression discontinuity
 - Exploit thresholds
 - E.g. treat with statin if 10-year CVD risk $>10\%$.
 - E.g. policy to have a named GP for people > 75 .
- (PS) calibration
 - Link subsample to richer data
- Sensitivity analysis
- Better use of captured data
 - High-dimensional PS, machine learning



Time-varying data

Single timepoint exposure

- Defining “start of follow-up” for control subjects
- What to do when patients change exposure status

Exposures varying over time

- Time-varying confounding
- MSM, G-computation, G-estimation, etc.
- Keyser et al., 2014, Eur J Epidemiol
 - Found Cox models gave similar answers to MSMs (much more complex)
 - Lack of data on the precise time-dependent confounders
 - EHR data unsuitable?

Other issues

Data linkage issues

- Linkage errors can introduce bias into the analysis
- Important for analyst to have information on quality of the linkage (e.g. match rank (deterministic) or match weight (probabilistic))
- Interesting theoretical work
- But... we rarely do the linkage

Determining drug exposure periods

- Very complex
- Different (deterministic) algorithms can lead to quite different exposure periods
- Information on prescriptions only