Primary Care data signposting

CPRD, THIN and other databases

Evangelos (Evan) Kontopantelis[1]

[1]Institute of Population Health, University of Manchester

Oxford, 21 Sep 2015

Outline

The Clinical Practice Research Datalink

CPRD

- Established in 1987, with only a handful of practices
- Since 1994 owned by the Secretary of State for Health
- In July 2012:
  - 644 practices (Vision system only: in Eng mainly London, SE, SC, NW, WM; see /pubmed/23913774)
  - 13,772,992 patients (≈5m active)
  - covering ≈7.1% of the UK population
- Access to the whole database is offered and costs ≈£130,000 pa
- Offers the ability to extract anything adequately recorded in primary care and construct a usable dataset

The Health Improvement Network database

THIN

- Established in 2003 as a collaboration between In Practice Systems Ltd and CSD Medical Research UK (EPIC)
- Now part and parcel of UCL
- In May 2014:
  - 562 practices (Vision system only, 50-60% overlap with GPRD)
  - 11.1m patients (3.7m active)
  - covering ≈6.2% of the UK population
- Usually offered under a 4-year license which costs £119,000
- Similar structure to CPRD and possibly more efficient patient matching for socio-demographic characteristics

QResearch

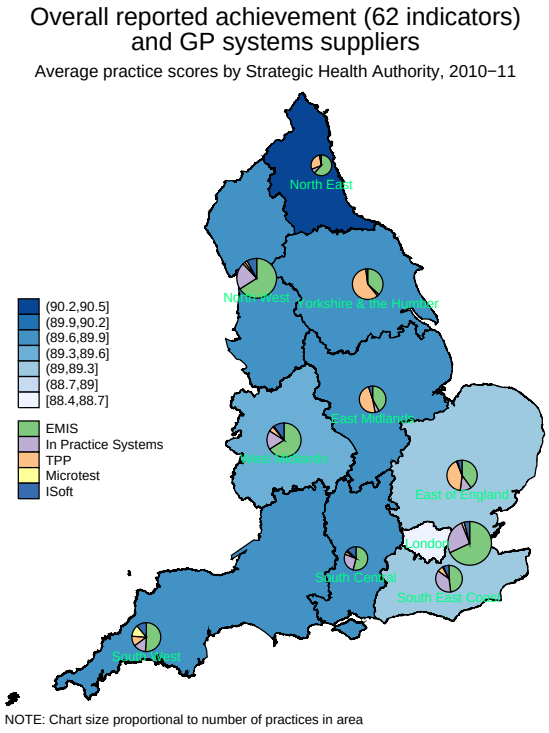
- Collaboration with the University of Nottingham
- In May 2014 reports:
  - 754 practices (EMIS systems: biggest UK provider)
  - over 13m patients (??m active)
  - covering ≈7% of the UK population?
- Datasets limited to 100k patients for externals
- Publication list, 90-95%: Vinogradova, Coupland and/or Hippisley-Cox

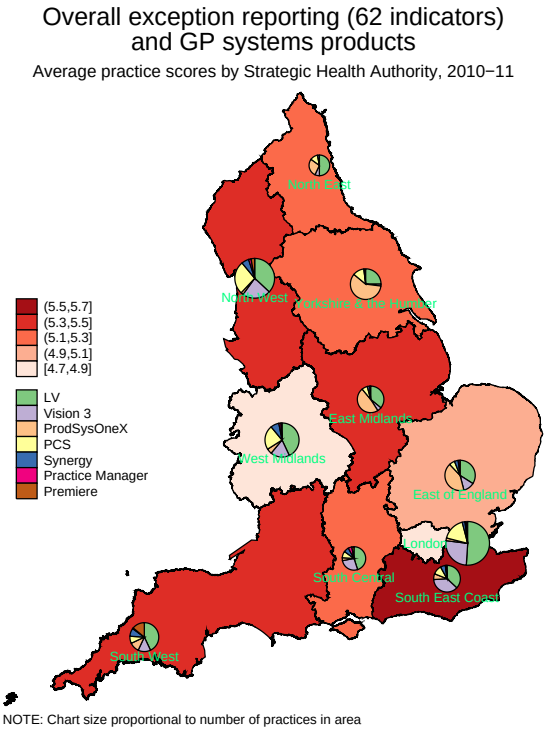
ResearchOne

- Collaboration between TPP and the University of Leeds
- In May 2014 reports:
  - ??? practices (SystmOne: Yorkshire&H, East Mid, East Eng, NE)
  - GP, Community Care, Hospital Care.
  - 30m research records
  - covering ≈?% of the UK population
  - costs?
- New potentially important player
- Uniformity of SystmOne and central databases for TPP systems likely to provide better quality data at lower cost

GP clinical systems

Quality of care and choice of clinical computing system, BMJ Open 2013





Export format

from SQL

- Broken down to numerous tables, due to volume of the data
- Text files need to be imported into powerful analysis/database management software
- Some of the information available:
  - Patient birthyear, sex, marital status, smoking/drinking status, height, weight and BMI
  - Clinical, referral, therapy, test and immunisation events
- All events are entered in codes (lookup tables available)
- Everything (likely to be recorded by a GP) can be identified, provided one knows which codes to look for and in which tables
- BUT a manual search on all the codes is not possible and automated processes are required

Primary Care Databases structure

based on CPRD

- Event files
  - Clinical: all medical history data (symptoms, signs and diagnoses)
  - Referral: information on patient referrals to external care centres
  - Immunisation: data on immunisation records
  - Therapy: data relating to all prescriptions issued by a GP
  - Test: data on test records
- Look-up files
  - Medical codes: READ codes, ≈100k available
  - Product codes: ≈80k available
  - Test codes: ≈300 available

Diabetes example

- Size of the tables prohibits looking at codes one by one
- Instead we use search terms to identify potentially relevant codes in the look-up tables and create draft lists

Example (Search terms for diabetes)

- String search in Medical codes: 'diab' 'mell' 'iddm' 'niddm'
- READ code search in Medical codes file: 'C10' 'XaFsp'
- String search in Product codes file: 'insulin' 'sulphonylurea' 'chlorpropamide' 'glibenclamide'

Diabetes example

...continued

- Clinicians go through the draft lists and select the relevant codes
- Three sets of codes are created, corresponding to:
  - QOF criteria
  - Conservative criteria
  - Speculative criteria
- Using the finalised code lists we search for events in the Clinical, Referral, Immunisation, Therapy and Test files
- Process involves much work in code writing, hence use of an appropriate statistical package like Stata or R is essential

Primary Care Databases tools

CPRD/THIN based but applicable to all

- Search commands
  - pcdsearch in Stata and Rpcdsearch in R
  - code list extraction algorithm
  - *Modelling conditions and health care processes in Electronic Health Records: an application to Severe Mental Illness with the Clinical Practice Research Datalink, under review*
- Code lists
  - clinicalcodes.org
  - Website with freely available developed code lists
  - *ClinicalCodes: An Online Clinical Codes Repository to Improve the Validity and Reproducibility of Research Using Electronic Medical Records, PLOS ONE 2014*
- Data extraction
  - rEHR (github.com)
  - R package for manipulating and analysing EHR data
  - *rEHR: An R package for manipulating and analysing Electronic Health Record data, under review*

Primary Care Databases tools

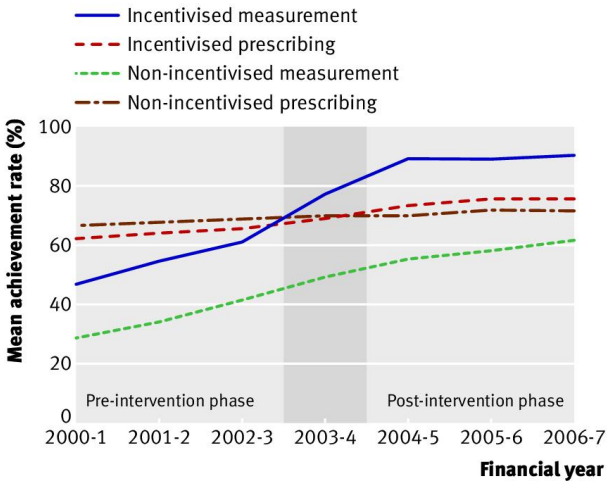
CPRD/THIN based but applicable to all

- Power calculations
  - ipdpower in Stata
  - mixed-effects power calculation through simulations
  - *Simulation-Based Power Calculations for Mixed Effects Modelling: ipdpower in Stata, JSS in print*
- Cleaning BMI
  - mibmi in Stata
  - Cleaning and multiple imputation for missing BMI data
  - *Longitudinal multiple imputation approaches for Body Mass Index: the mibmi command, Stata Journal under review*
- General Multiple imputation
  - twofold in Stata
  - Multiple imputation for longitudinal datasets
  - *Application of multiple imputation using the two-fold fully conditional specification algorithm in longitudinal clinical data, Stata Journal 2014*

Non-incentivised aspects of care

Sample of 148 representative practices from the CPRD

- Achievement rates improved for most indicators in the pre-incentive period
- Significant initial gains in incentivised indicators but no gains in later years
- No overall effect on improvement rate for non incentivised aspects in 2004-5
- But by 2006-7 achievement rates significantly below those predicted by pre- trends




BMJ

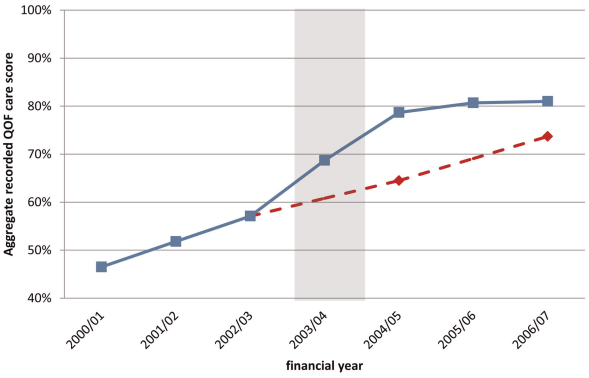
RESEARCH

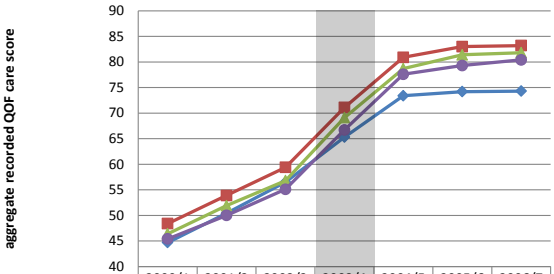
Effect of financial incentives on incentivised and non-incentivised clinical activities: longitudinal analysis of data from the UK Quality and Outcomes Framework

Patient level diabetes care

Sample of 148 representative practices from the CPRD

- In 2004-5 quality improved over-and-above this pre-incentive trend by 14.2%
- By 2006-7 improvement above trend smaller at 7.3%
- Levels of care varied significantly for sex, age, years of previous care, number of co-morbid conditions





	2000/1	2001/2	2002/3	2003/4	2004/5	2005/6	2006/7
new diagnoses	44.7	50.4	56.5	65.3	73.4	74.2	74.3
1-4 years	48.4	53.9	59.4	71.1	80.9	83	83.2
5-9 years	46.4	51.9	56.8	69.1	78.7	81.4	81.8
10+ years	45.4	50	55.1	66.7	77.6	79.3	80.4

ORIGINAL RESEARCH

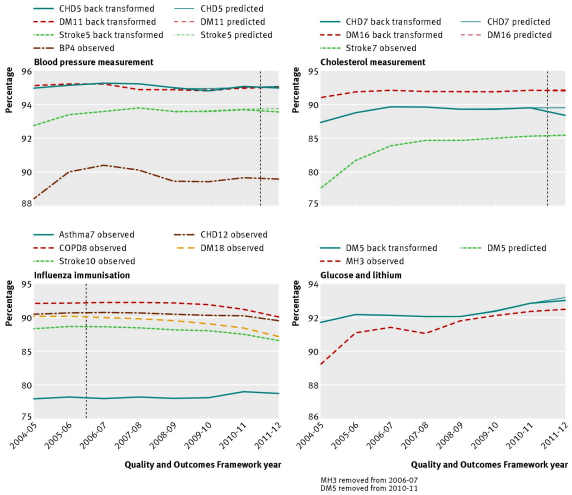
Recorded quality of primary care for patients with diabetes in England before and after the introduction of a financial incentive scheme: a longitudinal observational study

Evangelos Kontopantelis,[1] David Reeves,[1] Jose M Valderas,[2,3] Stephen Campbell,[1] Tim Doran[1]

Withdrawing incentives

644 CPRD practices, 2004-5 to 2011-12

- Financial incentives partially removed for aspects of care for patients with asthma, CHD, diabetes, stroke and psychosis
- Mean levels of performance generally stable after the removal of incentives, mainly in the short term
- Health benefits from incentive schemes may be increased by periodically replacing existing indicators with new ones



BMJ

BMJ 2014;348:g330 doi: 10.1136/bmj.g330 (Published 27 January 2014) Page 1 of 17

RESEARCH

Withdrawing performance indicators: retrospective analysis of general practice performance under UK Quality and Outcomes Framework

OPEN ACCESS

Evangelos Kontopantelis *senior research fellow*[1,2], David Springate *research associate*[1,3], David Reeves *reader*[1,4], Darren M Ashcroft *professor*[5], Jose M Valderas *professor*[6], Tim Doran *professor*[6]

Quality and Outcomes Framework

QOF datasets

- Pay for performance scheme that started in 1/4/2004
- Costs over £1bn pa
- Voluntary scheme but participation over 99.9%
- Freely available on Health & Social Care Information Centre (HSCIC), by financial year:
  - NHS practice code and list size
  - Prevalence on 15 key chronic conditions (e.g. diabetes, asthma, CHD, COPD etc)
  - Practice level performance on various clinical indicators for these conditions
  - Practice level exception rates for each indicator

General Medical Services

GMS datasets

- Data from around 2000
- Information on general practices
- Available on request (not free but cheap) from the HSCIC, by calendar year:
  - NHS practice code, list size, contract type, full address (including postcode, sha, pct, lsoa)
  - Number of GPs, FTE, names, country/area qualified, sex, age
  - Patient counts by age group and sex
- Part of the Workforce theme: more info for other health professions

Patient Satisfaction

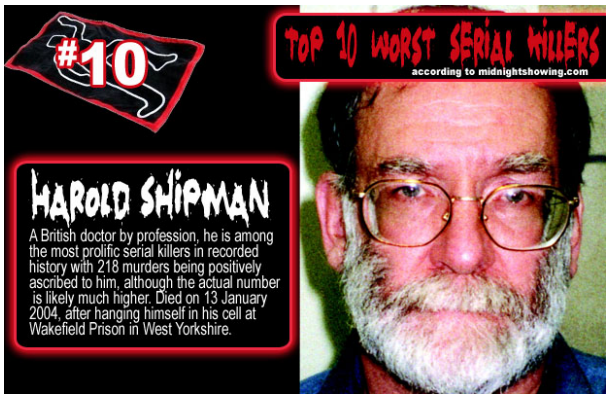
GP Patient Survey

- Data from 2007
- Run by Ipsos MORI, data collected twice a year
- Stratified random sampling of patients to collect data on satisfaction with GP services
- Data freely at the practice and higher levels, weighted (to match patient population) and unweighted satisfaction scores on:
  - access, making an appointment, waiting times speaking to GP or nurse, ease of access
  - last GP and last nurse appointment, opening hours, overall experience
  - and many more domains

Primary Care Mortality

PCM database

- Data from 2006
- Managed by the HSCIC and accessible remotely
- Monthly and annual extracts of individual record level data on deaths supplied by ONS:
  - registered GP/practice, patient details e.g. age, causes of death, NHS no
- Data for use by Local Authorities and NHS organisations only



Census 2011 datasets

but also 2001, 1991 etc

- Information aggregated at various levels, as low as lower super output area (LSOA) level
- Freely available from the ONS websites, including:
  - Counts by age groups and sex
  - Health
  - Ethnicity
  - Religion
  - Occupation
  - Qualifications
  - Household-accommodation

Deprivation datasets

Index of Multiple Deprivation (IMD): 2004, 2007, 2010, 2014

- Important covariate, available at the 2001 LSOA level
- England only (although there is a Welsh IMD as well)
- Free at the Neighbourhood Statistics ONS website
- Aggregate of 7 domains:
  - Income
  - Employment
  - Health deprivation
  - Education and skills
  - Housing
  - Crime
  - Environment
- 2010 range was 0.5-87.8 (9.8 and 30.2 for 25th and 75th centiles)

Mortality datasets

From 1998

- As counts available at the LSOA level (2001 or 2011) but special request to the ONS mortality team
- As standardised mortality rates freely available but at electoral ward level or higher from the main ONS website
- Specific mortality causes available:
  - using ICD-10 codes from 2001, ICD-9 before
  - counts at the LSOA level can be broken down by sex and age-group

Admitted patient care dataset

and outpatient

- Data more or less available from 1989
- Patient-level data, with various organisational markers:
  - GP, SHA, PCT, site of treatment
- Available upon request from the HSCIC, including:
  - patient characteristics (incl IMD), admissions, discharges, episodes, clinical, maternity, psychiatric
- Additional sensitive info: dob, NHS number, patient residence postcode, LSOA etc
- Data for outpatient care available from 2003: similar but less detailed

Critical care data

- Data available from 2008
- Add-on dataset which should be matched with inpatient dataset, on request from the HSCIC
- Includes:
  - critical care dates
  - admission type
  - support info
  - critical care levels
  - discharge info

Accident and Emergency data

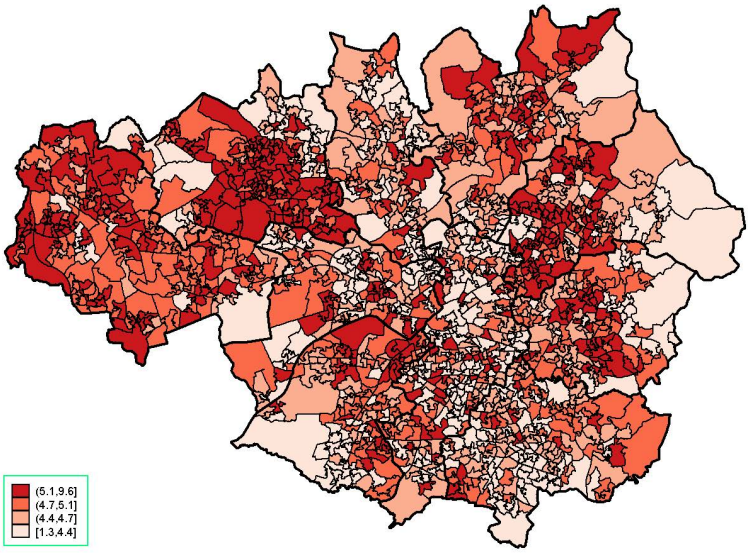
- Data available from 2007
- Similar covariate and organisation info as inpatient-outpatient datasets Available upon request from the HSCIC, with info on:
  - attendances
  - clinical diagnosis
  - clinical investigation
  - clinical treatment
- Additional sensitive info: dob, NHS number, patient residence postcode, LSOA etc

Lookup tables

- To combine datasets reported at different levels
- Usually the postcode is the best start, if known
- The UK Data Service (previously UK Borders) contains tables to help merge data at various levels, at 1991, 2001, 2011 or 2013 boundaries:
  - PCTs
  - Wards
  - LSOAs
  - SHAs
  - Clinical Commissioning Groups (CCGs formerly PCTs)
  - NHS Area Teams

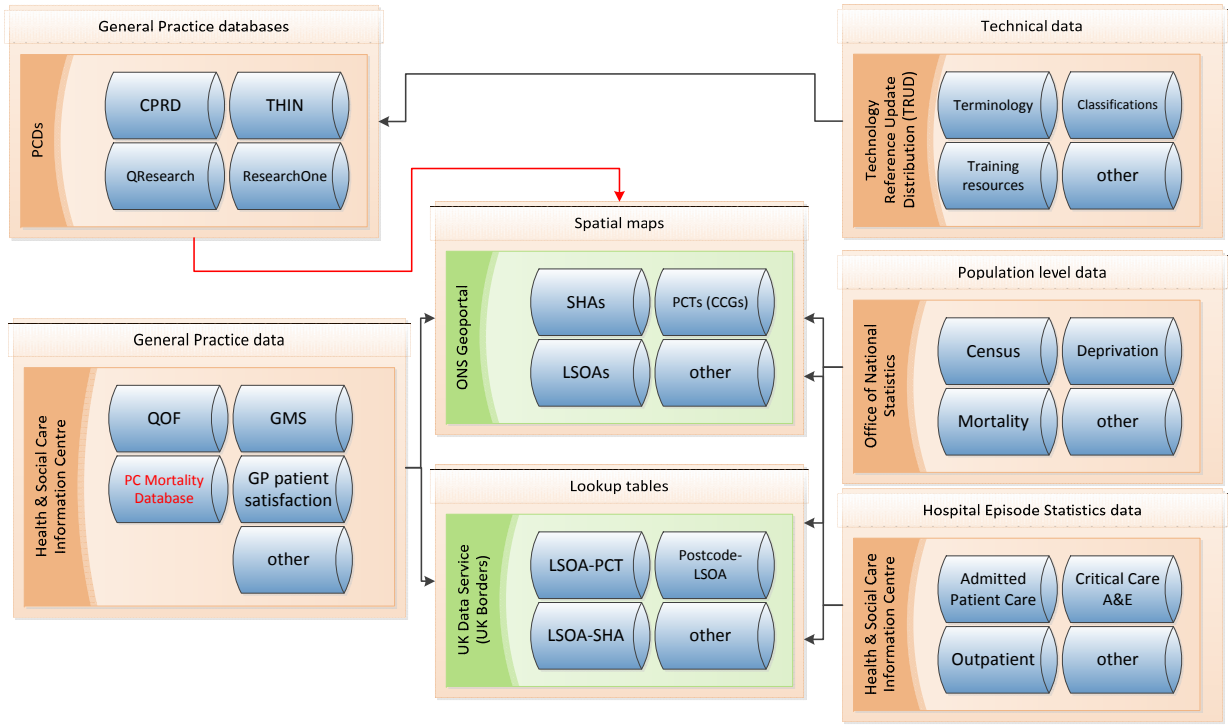
Spatial mapping

- After merging at a geographical level spatial coordinates are useful for plotting or accounting for spatial correlations in regression analyses
- ONS Geoportal holds various digital vector boundaries files (shapefiles) for 2001, 2011 and more recent geographies:
  - LSOAs
  - PCTs-CCGs
  - SHAs
  - Regions



Overview

Health Sciences related



- Comments and questions: e.kontopantelis@manchester.ac.uk