

# Linking Cancer Registry data with CPRD

## **NIHR STATISTICS GROUP Routine Data Meeting**

April 23 2018  
Jacqueline Birks

Oxford Biomedical Research Centre  
Research Design Service South Central

# Linking databases



- CPRD - Clinical Practice Research Datalink
  - Anonymised database of the primary care records of 5 million current UK patients
- NCRAS - National Cancer Registration and Analysis Service
  - Record of all registrable tumours diagnosed or treated in England of which NCRAS has been notified

# CPRD



- Coded data from 3 main GP electronic health record systems
- Data on patient demographics, registration, care events, medical diagnoses, referrals, prescriptions issued in PC, tests, immunisations, lifestyle

# CPRD

---

- To request data
  - protocols are submitted to ISAC (Independent Scientific Advisory Committee)
  - Include request for data linkage

# Research Project

- Independent validation of a machine learning algorithm to identify patients with undiagnosed colorectal cancer
  - Using full blood counts (FBC) collected for any reason which may reveal signs of anaemia and iron deficiency (haemoglobin, red blood cell distribution width, mean cell volume)

# Independent validation using UK routine primary care patient data

- Population of patients from Clinical Practice Research Datalink (CPRD) registered between 2000-2015
- Primary endpoint first diagnosis of colorectal cancer in the primary care record (Read code)
- Inclusion criteria
  - At least one FBC
  - Aged at least 40 years by latest date of inclusion
  - At least 2 years of follow up after registering with a practice
- Linked to cancer registry

# Data Linkage



- English practices who have consented
  - 58% of all practices
  - 75% of English practices
- Do not underestimate the time it takes to receive the linked data

# National Cancer Registry

---

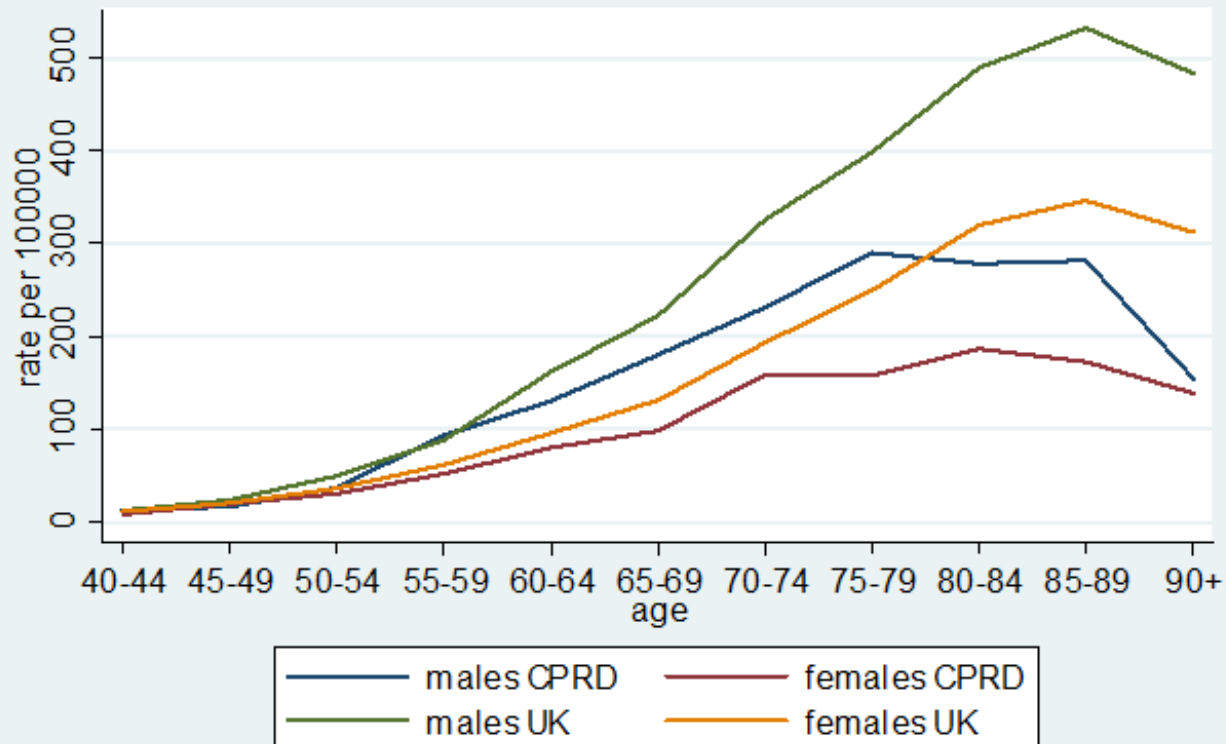
- Data retrieved for each tumour
  - Patient id
  - Date of diagnosis
  - Site coded (ICD-10)
  - Cancer site (eg liver, bladder)
  - Dukes' Stage (where applicable)
  - ethnicity

There may be multiple diagnoses for a patient



### Incidence of colorectal cancer in 2012 by age

CPRD data and UK data



# Missing data and errors

- 6256/21642 (29%) diagnoses in Cancer Registry missing from linked CPRD patients data
- 2631 diagnoses of colorectal cancer in CPRD linked patients not matched in Cancer Registry
  - 947 diagnosed after Cancer Registry cut off date
  - Possible errors, 723 had a different cancer diagnosis in Cancer Registry
  - 961 no explanation

# CPRD – Cancer Registry data

## Diagnosis of colorectal cancer for 1893641 patients linked to Cancer Registry

	No Cancer registry diagnosis	Cancer registry Diagnosis	total
No CPRD diagnosis	1869368	6256	1875624
CPRD diagnosis	2631	15386	18017
total	187999	21642	1893641

# Conclusions

---

- Cancer Registry data are essential for an accurate identification of cancer cases
  - 29% of cancer diagnoses may not be found in CPRD data
- Ascertain exactly which practices are linked to Cancer Registry