The ADAPTT Study: a study using linked CPRD-HES data

Jessica Harris
University of Bristol

Background

‘Real world’ bleeding events in patients exposed to different regimens of dual antiplatelet therapy (DAPT) have a significant clinical and economic impact but have not been previously quantified.

Methods

Using linked Clinical Practice Research Datalink (CPRD) and Hospital Episode Statistics (HES) data we will assemble populations eligible for three "target trials" in patients who underwent:

- percutaneous coronary intervention (PCI)
- coronary artery bypass grafting (CABG)
- patients who are conservatively managed (medication only) after acute coronary syndrome (ACS).

Methods

Steps:

1. Identify the members of the target trials
2. Identify exposure categories (aspirin or DAPT)
3. Extract confounders from CPRD and HES data, as identified in literature review
4. Measure “minor bleeding” (CPRD data) and “major bleeding” (HES data) *primary outcome*
5. Identify secondary outcomes (e.g. mortality, MI, stroke)
6. Formally test treatment effect on time to first bleed

Step 1: Identify the members of the target trials

Data sources

We requested the following data from CPRD:

- Primary care plus: Full HES APC, HES Outpatient HES, HES A&E, ONS death registration, and Index of Deprivation (IMD) data.

The extraction population will comprise of all acceptable patients in CPRD who are eligible for linkage to HES, ONS Mortality and Index of Multiple deprivation data. CPRD primary care data will be derived from Jan 2018 static version of CPRD GOLD. Linked data will be provided from Set 15 data (HES APC: 01/04/1997- 31/07/2017; HES OP: 01/04/2003 -31/07/2017; HES A&E: 01/04/2007- 31/07/2017; ONS Mortality: 02/01/1998 -19/09/2017)

Data sources

We provided the ICD-10 diagnosis codes and OPCS procedure codes for each of PCI, CABG and ACS.

Definitions

PCI: Percutaneous coronary intervention identified by OPCS codes

K49	Transluminal balloon angioplasty of coronary artery
K50	Other therapeutic transluminal operations on coronary artery
K75	Percutaneous transluminal balloon angioplasty and insertion of stent into coronary artery

CABG: Coronary artery bypass grafting identified by OPCS codes

K40	Saphenous vein graft replacement of coronary artery
K41	Other autograft replacement of coronary artery
K42	Allograft replacement of coronary artery
K43	Prosthetic replacement of coronary artery
K44	Other replacement of coronary artery
K45	Connection of thoracic artery to coronary artery
K46	Other bypass of coronary artery

ACS: acute coronary syndrome identified by ICD10 codes

I20.0	Unstable angina
I21	Acute myocardial infarction
I22	Subsequent myocardial infarction
I24.9	Acute ischaemic heart disease, unspecified

Data sources

We provided the ICD-10 diagnosis codes and OPCS procedure codes for each of PCI, CABG and ACS.

Individuals with at least one mention of any of these between 01/04/2010 and 31/07/2016 were identified by CPRD, plus:

- Record that the data were “acceptable”
- Data were eligible for linkage with HES, ONS and LSOA data
- M/F recorded
- Age 18+
- At least 12 months data available prior to index date

Data received

Individuals with at least one mention of any of PCI / CABG / ACS between 01/04/2010 and 31/07/2016, plus:

- Record that the data were “acceptable”
- Data were eligible for linkage with HES, ONS and LSOA data
- M/F recorded
- Age 18+
- At least 12 months data available prior to index date

CPRD identified 41,119 patients that met these criteria:

- 24.7 million consultations
- 32.8 million prescriptions

Data files received: GP data

CPRD data files	*Description*
Patient data	*1 row per person, patient details*
Practice data	*1 row per GP practice*
Staff data	*Many rows per patient; data on practice staff member entering data*
Consultation data	*Many rows per patient; 1 row per consultation per GP practice*
Clinical data	*Many rows per patient; many rows per consultation per GP practice †*
Additional clinical data	*Many rows per patient; 1 row per additional data**
Referral data	*Many rows per patient; information on any referral**
Immunisation data	*Many rows per patient; information on any immunisations**
Test data	*Many rows per patient; information on any tests†**
Therapy data	*Many rows per patient; information on any prescriptions*; includes product codes, BNF codes*

*† includes medical terms (Read codes); * link with consultation/clinical data*

HES data: structure of episodes, spells and CIPs

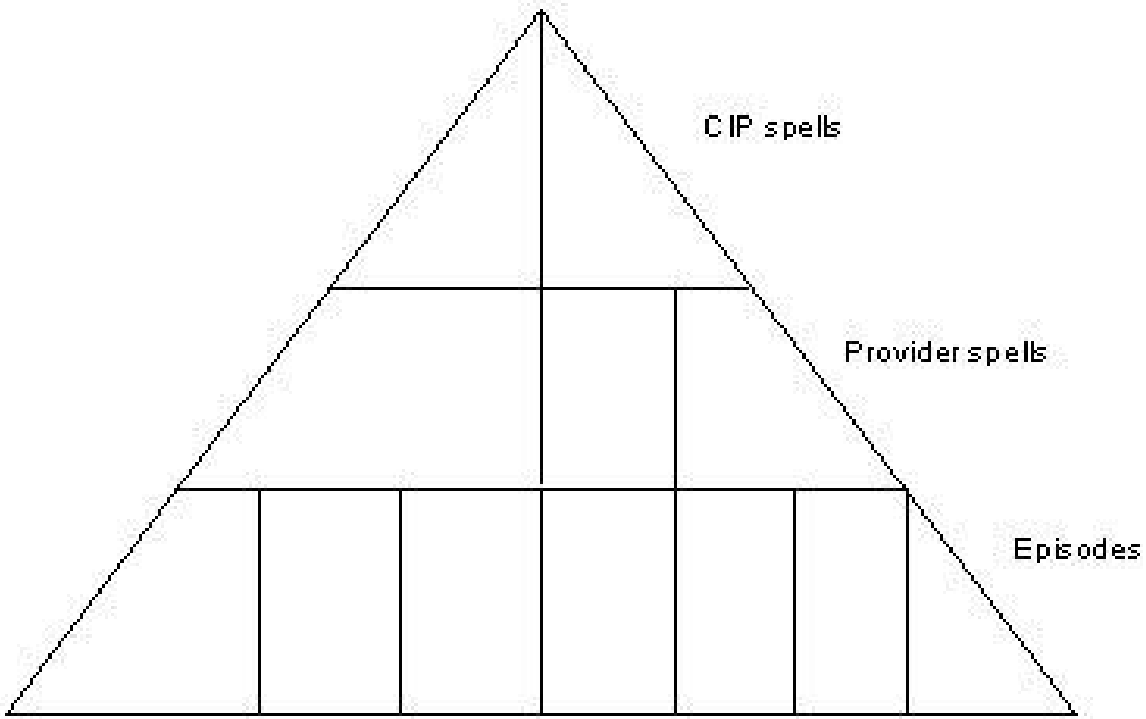


Figure 3. Demonstration of how CIP Spells, Provider Spells and Episodes relate to each other.

Data files received: Hospital admissions data

HES data files	*Description*
Patient data	*1 row per person, patient details*
Hospital data*	*1 row per unique spell including dates, etc.*
Episode data*	*1 row per episode including dates, etc.*
Diagnosis data*	*ICD-10 diagnoses codes; datasets by i) episode; ii) hospitalisation. Further dataset of primary diagnoses by hospitalisation*
Procedures data*	*1 row per episode including OPCS procedure codes*
ACP data*	*1 row per augmented care period*
Critical care data*	*1 row per critical care period*
Maternity data*	*Maternity data*
HRG data*	*Health resource group data; 1 row per episode*

** Many rows per patient*

Other data files received

A&E	Outpatient	Other
A&E patient data	Outpatient patient data	Mortality data
A&E attendance data	Outpatient appointment data	IMD data
A&E diagnosis data	Outpatient clinical data	
A&E investigation data	Outpatient pathway data	
A&E pathway data		
A&E HRG data		
A&E treatment data		

Step 2: Identify exposure categories (aspirin or DAPT)

Identify the exposure categories

In hospital, the majority of patients will be assigned to the following treatment regimens:

PCI:	aspirin and prasugrel
	aspirin and clopidogrel
	aspirin and ticagrelor
CABG:	aspirin / aspirin and clopidogrel
ACS:	aspirin / aspirin and clopidogrel

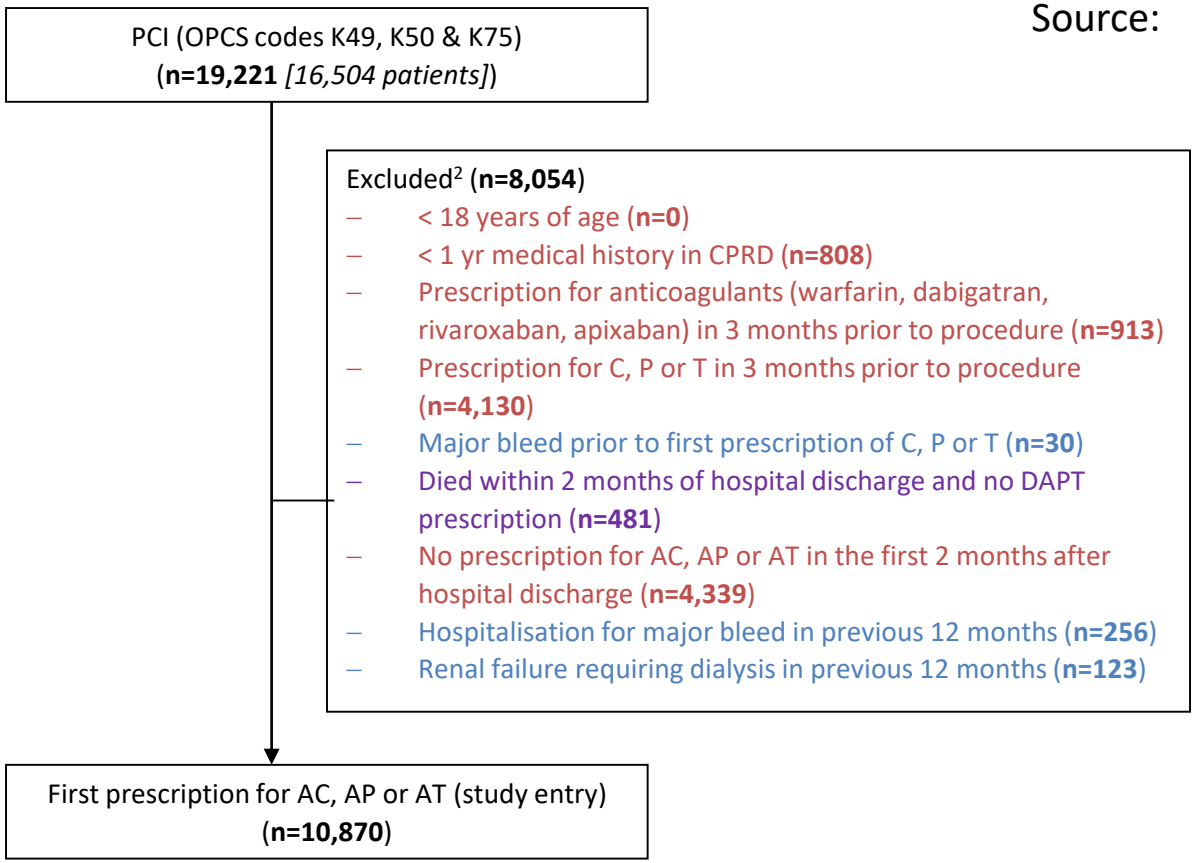
However treatment details are not recorded in HES so we are reviewing CPRD data for *prescriptions within two months of hospital discharge.*

Exposure: medication on discharge

- We have extracted all the Read codes for aspirin, clopidogrel, prasugrel and ticagrelor from the CPRD December 2017 code browser:
  - 329 codes for aspirin; 24 for clopidogrel; 4 for prasugrel; 5 for ticagrelor

- This list has been reviewed and some aspirin codes have been excluded, such as aspirin and caffeine tablets, and dispersible tablets. Final numbers:
  - 198 codes for aspirin; 24 for clopidogrel; 4 for prasugrel; 5 for ticagrelor

Identify the members of the target trials



Source: CPRD data
HES data
Other data

Continuous inpatient stays (CIPS) that include both PCI and CABG are considered CABG; CIPS that include both ACS and PCI are considered PCI; CIPS that include both ACS and CABG are considered CABG

Step 3: Extract confounders from CPRD and HES data

Extracting confounders

- Systematic review identified which confounders we are interesting in capturing, in order to adjust for these in the final analysis.

- Some will be extracted from CPRD data, some from HES data, some from both sources.

	Confounding domain	Confounders		Notes
1	Age	–	Age	Recorded in CPRD data: defined as age at PCI / CABG / ACS diagnosis
2	Gender	–	Gender	Recorded in CPRD data
3	Body mass index	–	Body mass index	Calculated using the most recent measurement of weight and height available within CPRD data from the whole period available. Applied acceptable ranges as previously published
4	Ethnic group	–	Ethnic group	Extracted from HES data
5	Smoking	–	Smoking	Used Read code list available online (124 Read codes) to identify ex-smoker / non-smoker / smoker from CPRD data, using most recent data recorded
6	Alcohol intake and recreational drug use	–	Alcohol intake	Not reviewed yet but alcohol consumption (units per week) is recorded (entity=5)

	Confounding domain	Confounders	Notes
7	Co-morbidities		Admission data with diagnosis code:
		– Hypertension	– I10 Essential (primary) hypertension
		– Hypercholesterolemia	– E78.0 Pure hypercholesterolaemia
		– Angina	– I20 Angina pectoris
		– Diabetes	– E10 – E14 Type 1 diabetes mellitus; Type 2 diabetes mellitus; Malnutrition-related diabetes mellitus; Other specified diabetes mellitus; Unspecified diabetes mellitus
		– Stroke/transient ischaemic attack	– I64 Stroke, not specified as haemorrhage or infarction
		– Heart failure	– I50 Heart failure
		– Atrial fibrillation	– I48 Atrial fibrillation and flutter
		– Renal disease	– N18 Chronic kidney disease
		– Liver disease	– K70 – K77 Alcoholic liver disease; Toxic liver disease; Hepatic failure, not elsewhere classified; Chronic hepatitis, not elsewhere classified; Fibrosis and cirrhosis of liver; Other inflammatory liver diseases; Other diseases of liver; Liver disorders in diseases classified elsewhere
		– Chronic obstructive pulmonary disease	– J44 Other chronic obstructive pulmonary disease
			NB: all these were reviewed within 1 year prior to PCI / CABG / ACS diagnosis using HES admissions data only. **Action: review CPRD data where appropriate (e.g. 25 hypertension Read codes available online)**

Etc…….

Step 4: Measure “minor bleeding” and “major bleeding” (HES data)

Primary outcome

- **Minor bleeding:** as identified in CPRD data using Read codes which represent GP appointments where minor bleeding has been recorded (e.g.nosebleeds)

- **Major bleeding:** as identified in HES admissions data using ICD10 diagnosis codes which represent admissions where bleeding has been recorded (e.g. gastrointestinal bleeding)

We are interested in the time until the first bleed, up until 12 months after the index procedure/diagnosis as this reflects the time period that a patient would be prescribed these medicines.

To do:
Step 5: Identify secondary outcomes
Step 6: Formally test treatment effect on time to first bleed

Thoughts

Really interesting to work with such large datasets but not easy to build analysis datasets:

- For identification of "eligible" participants, and identification of confounders, have to carefully extract the relevant data from different sources before linking these together.
- Often requires the use of code lists, can source these online or create from the CPRD code browser.
- In many cases, data extraction requires additional steps using lookup and text files to extract the data required.
- Sometimes guidelines are available (e.g. *BMI: Representativeness and optimal use of body mass index in the UK Clinical Practice Research Datalink (CPRD) Bhaskaran et al., 2013*)