

Terrie Walker-Smith, Katherine Joyce, Rachel Maishman, Penny Lambert, Emma Hopkins, Lucy Culliford

On behalf of Bristol Trials Centre (BTC), University of Bristol

Introduction

Patient data routinely collected by health service teams are increasingly being used to reduce the burden of data collection in clinical studies but it is often difficult for researchers to check the accuracy of the data. In the OMACS study (Outcome Monitoring After Cardiac Surgery) we collect data from multiple sources for patients undergoing cardiac surgery at University Hospitals Bristol and Weston NHS Foundation Trust (UHBW). We have assessed the accuracy of two routine data sources for identifying operation type in patients consented to OMACS.

Methods

- Data on operation type were collected from 2 routine data sources (Figure 1) for OMACS participants who underwent surgery in 2020.
- We identified patients who underwent Coronary Artery Bypass Graft (CABG), Aortic Valve Replacement/Repair (AVR) or Mitral Valve Replacement/Repair (MVR) according to each data source.
- The procedures identified in the two sources were compared.
- Where discrepancies were found the medical notes were checked to identify the actual procedure performed.

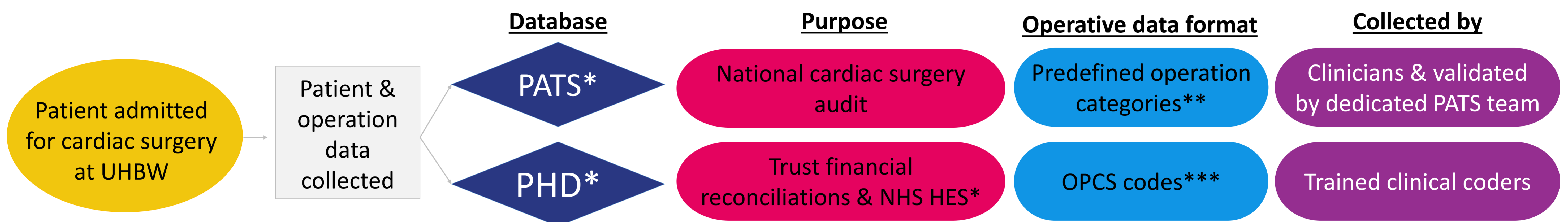


Figure 1: Summary of routine data sources used to identify operative procedure in the OMACS study

*PATS = Patient Administration and Tracing System, PHD = Patient History Database, HES = Hospital Episode Statistics
 ** Procedures categorised as coronary artery bypass graft, valve, major aortic or other cardiac procedure
 ***Office of Population Censuses and Surveys - Nationally standardised procedure codes

Results

Data availability

536 OMACS participants were operated on in 2020.

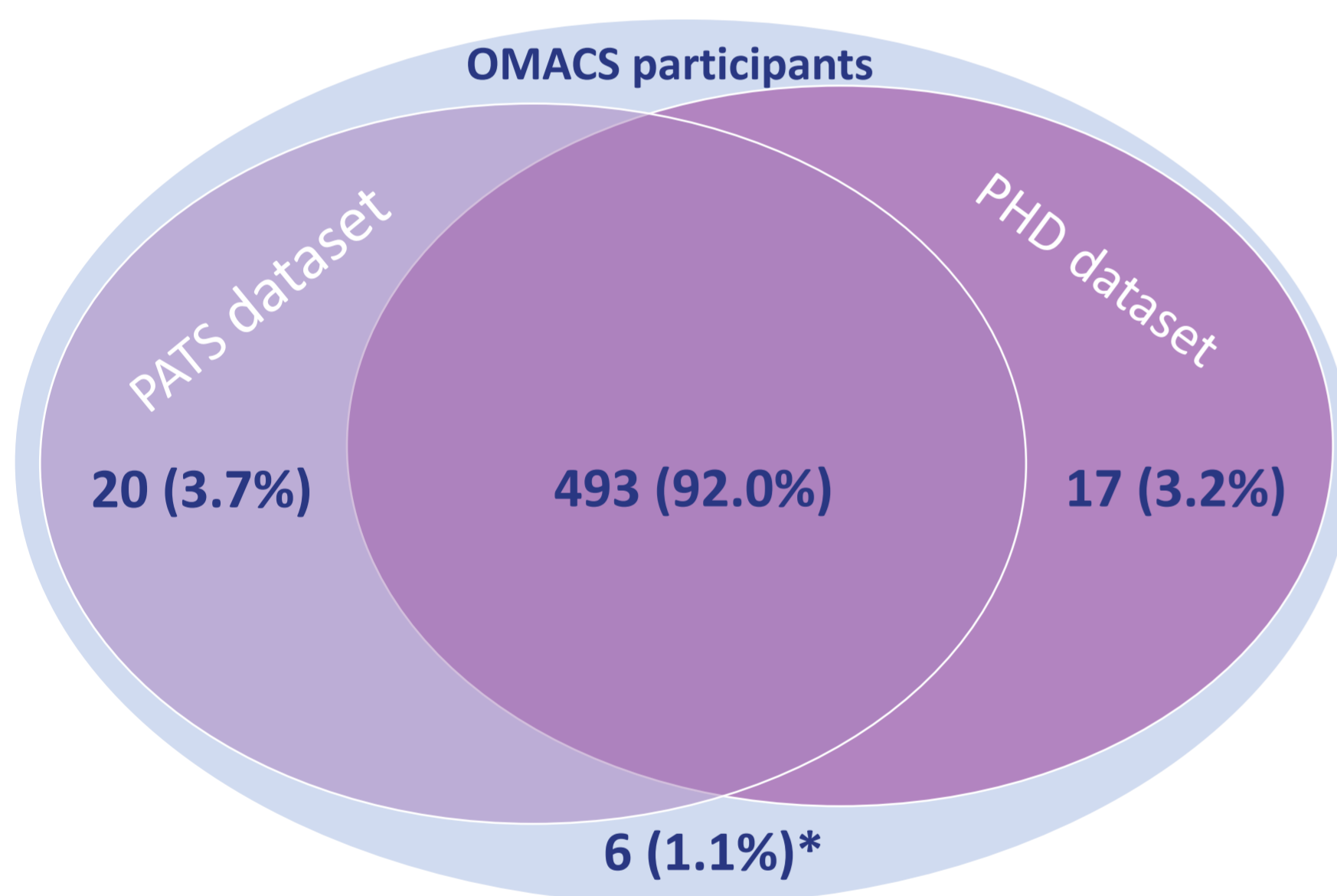


Figure 2: Number of OMACS participants with operation data identified in the two routine data sources. Records were linked by unique hospital number and operation date.

*6 consented participants did not have operation data identified in either routine data source.

- 461/493 (94%) participants with operation data available in both PATS and PHD had one of the three procedures of interest recorded in at least one of the routine data sources.
- The data sources agreed with respect to the procedure recorded for 433/461 (94%) participants.

Data accuracy

Medical notes were checked for the 28 participants whose operation type differed between the two data sources to identify the accurate data source (Figure 3).

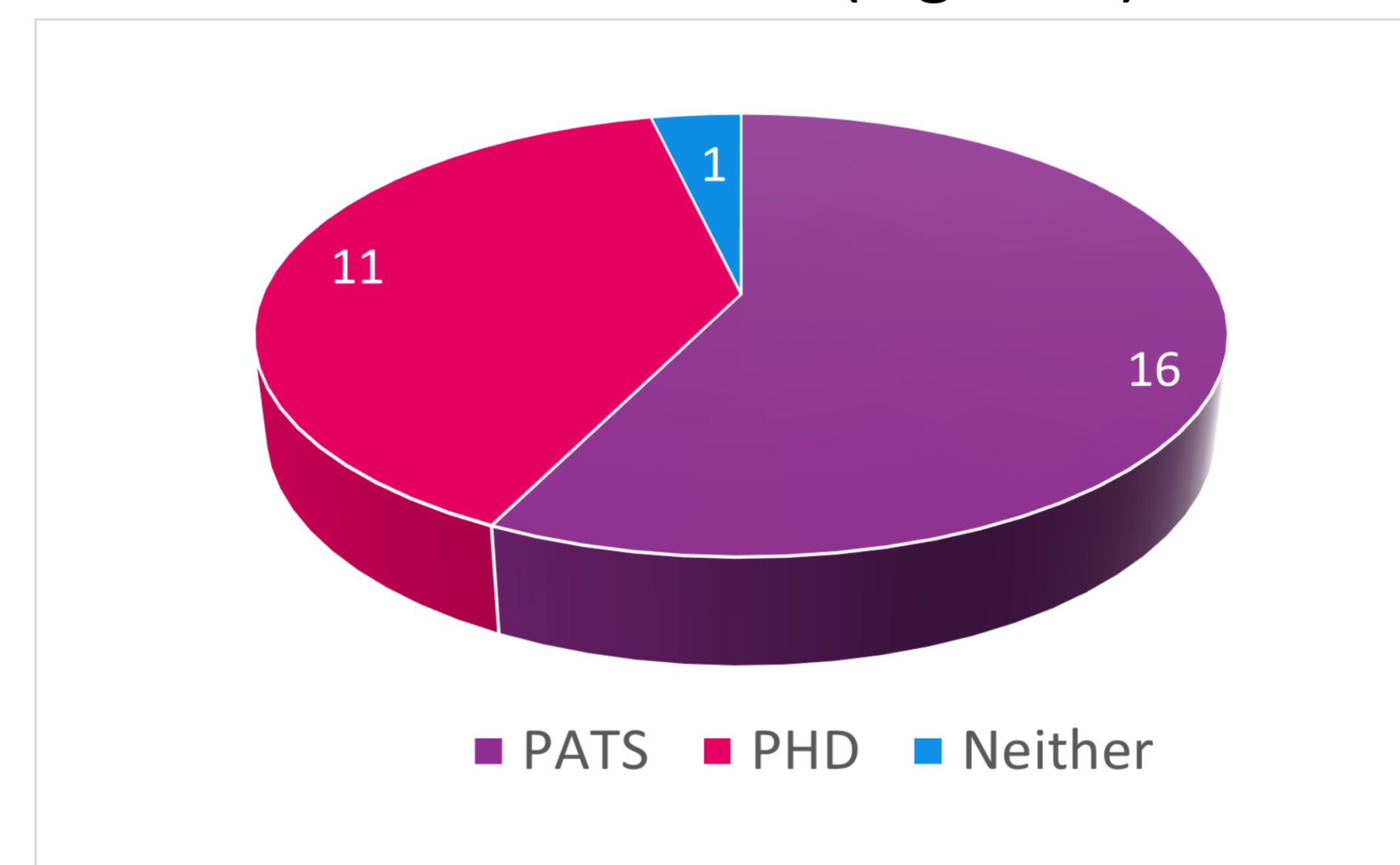


Figure 3: Data source with accurate operative procedure data in participants with discrepant data between PATS and PHD.

- Both data sources were found to contain errors, and there was one participant for whom neither source was correct.
- Therefore, neither data source was identified as being the 'gold standard' for identifying operation type.

Discussion

- Operation data was largely consistent between data sources, however, despite data cleaning and auditing, neither source was 100% accurate.
- We only studied the 3 most common cardiac procedures and data may be less accurate for rarer procedures.
- Levels of discrepant or missing data may be acceptable for some uses, e.g. analysing registry data but not for others, e.g. an end point of a randomised controlled trial.

We recommend trialists risk assess the use of routine data for key data points and where possible validate sources to determine accuracy. We also recommend including clinicians or staff who input the data when designing studies using routine data.

Some other key considerations for using routine data:

	Pros	Cons
Data Accuracy/format	Data has been audited and is released in a standardised format.	Data may not be 100% accurate/ complete and may require validation, if possible.
Resourcing	Reduces data collection burden on site teams	Data processing requires time from central team
Timing	Data are released at regular intervals which is easy to manage.	Initial access to data may be delayed and not available in real-time