Identification and validation of real-world androgen-deprivation therapy use among prostate cancer survivors in electronic health records in England

Shamsudeen Mohammed[1], Ajay Aggarwal[1], Harriet Forbes[1], Garth Funston[2], Mieke Van Hemelrijck[3], Alexander Lyon[4], Bernard Rachet[1], Richard M Martin[5], Krishnan Bhaskaran[1], Helena Carreira[1]

[1]London School of Hygiene and Tropical Medicine, London, UK; [2]Queen Mary University of London, London, UK; [3]King's College London, London, UK; [4]Royal Brompton Hospital, London, UK; [5]University of Bristol, Bristol, UK

1. Background & Aim

- Prostate cancer is the most common male cancer in the UK. Androgen deprivation therapy (ADT) lowers testosterone to control prostate cancer growth but has serious adverse effects.
- Risk prediction and safety monitoring rely on accurate exposure data, yet ADT data in England are fragmented across multiple electronic health record (EHR) systems, risking incomplete capture and misclassification.
- We aimed to develop an algorithm to identify ADT exposure in linked population-based EHR datasets in England.

2. Methods

- **Linked Data Sources:** Clinical Practice Research Datalink (CPRD) Aurum/GOLD, Hospital Episode Statistics Admitted Patient Care (HES APC), National Cancer Registry (NCRAS), and the Systemic Anti-Cancer Therapy (SACT) dataset.
- **Study Population:** Men with an incident prostate cancer diagnosis in NCRAS between 1999 and 2018, with no prior invasive cancer.
- **ADT Identification:** Clinician-validated codelists were used to identify surgical (orchidectomy) and pharmacological ADT across linked sources. Events were extracted, harmonised, and completeness validated using several methods.

3. Results

- Linked data were available for 138,119 men with incident prostate cancer between 1999 and 2018.
- After deduplication and administrative censoring (Fig. 1), 90,110 (65.2%) received ADT within the study period.

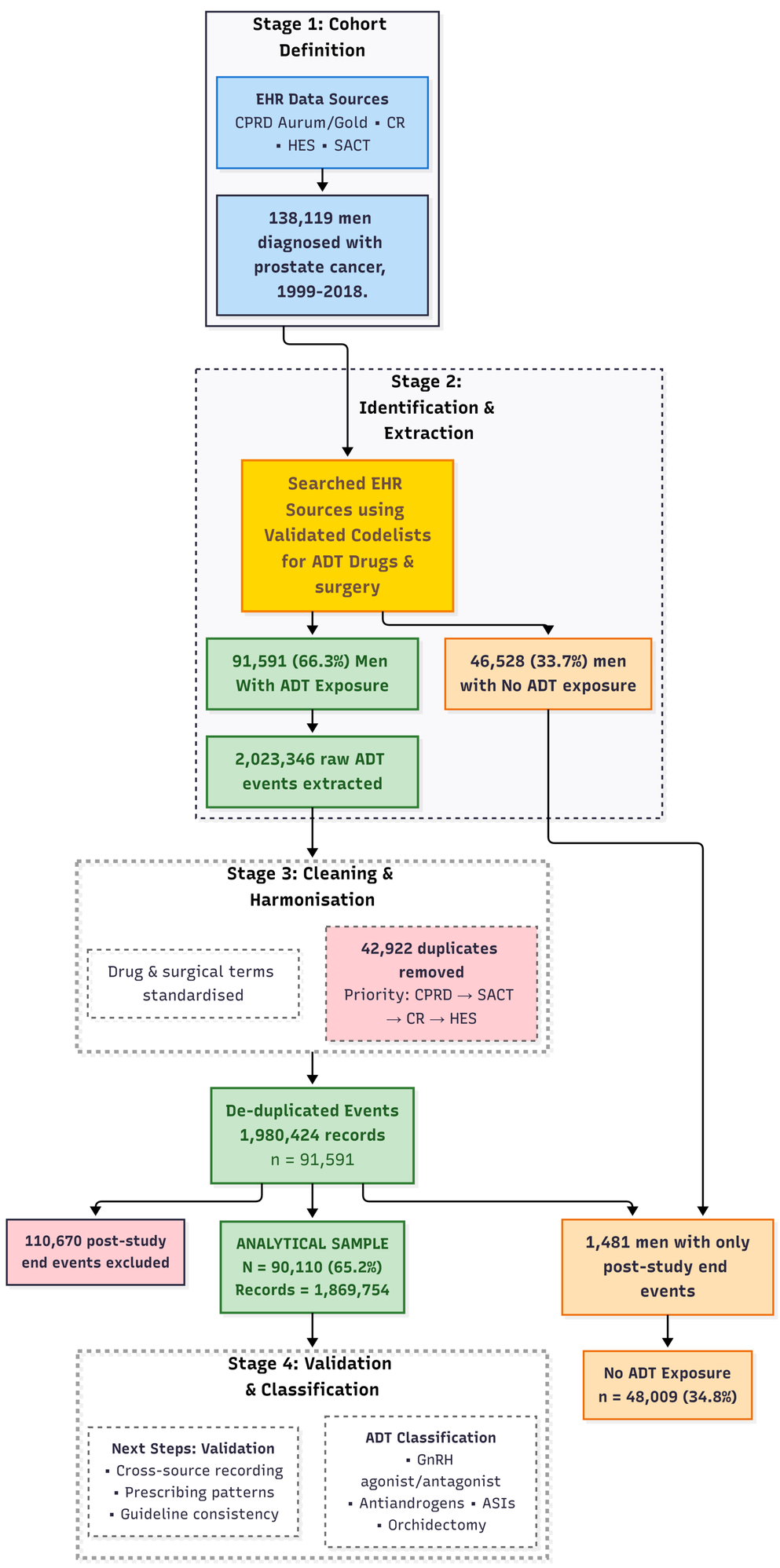

Fig 1. ADT identification and extraction in linked English population-based EHR datasets

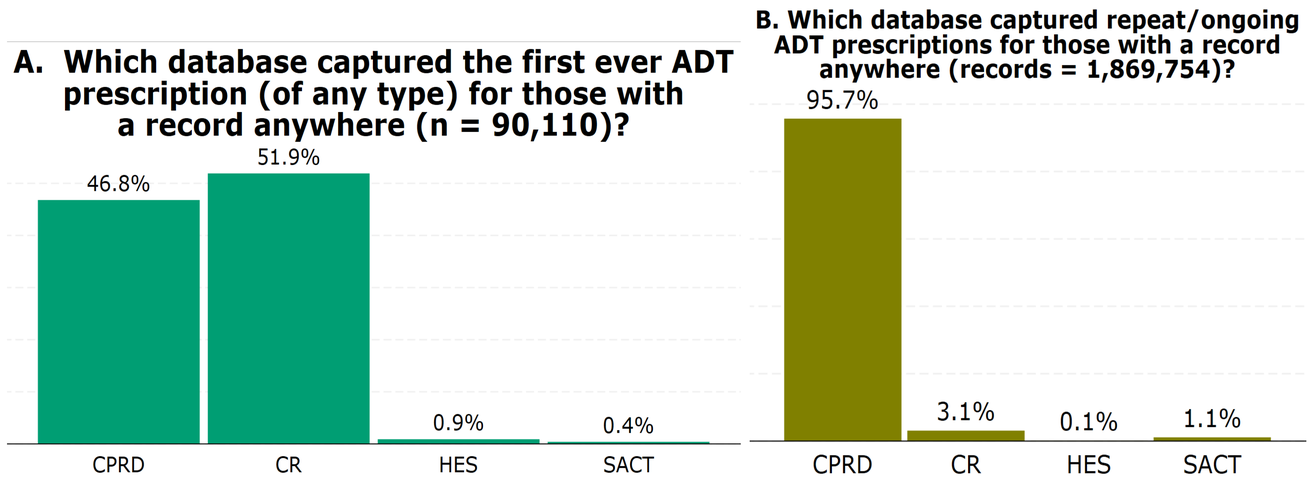

Fig 2. Recording of ADT initiation and repeat prescriptions in EHR databases

CPRD = Clinical Practice Research Datalink
HES APC = Hospital Episode Statistics Admitted Patient Care
CR = Cancer Registry
SACT = Systemic Anti-Cancer Therapy dataset (available from 2012)

Table 1. Validation of ADT recording completeness and clinical consistency

A. Completeness of ADT recording for those with a record anywhere	**N = 90,110**
% with ADT recorded in SACT and in CPRD	96.4% (n = 3,122)
% with ADT recorded in cancer registry (CR) and in CPRD	94.4% (n = 53,474)
% with ADT recorded in HES APC and in CPRD	85.5% (n = 1,614)
B. Clinical consistency of the harmonised ADT records from all sources	
% received ADT within 12 months of radiotherapy	85.5% (n = 20,290)
% of ADT use in metastatic prostate cancer	94.7% (n = 8,936)
% of ADT use in non-metastatic prostate cancer	50.2% (n = 34,046)
Median time interval between consecutive ADT prescriptions: **32 days (IQR: 21 to 84)**	

- Of the 90,110 who received ADT, 42.5% had records in only one data source
- Among the 90,110 men, 51.9% had their first ADT recorded in CR, and 46.8% in CPRD (Fig. 2A).
- CPRD was the main source of repeat/ongoing pharmacological ADT prescriptions (95.7%) after initiation (Fig. 2B).
- Most surgical ADT procedures (orchidectomy) were recorded in HES APC.
- Cross-source concordance was high (Tab. 1), and treatment patterns in the combined dataset were consistent with clinical practice and guidelines.
- The median interval between pharmacological ADT prescriptions was consistent with real-world ADT prescribing patterns (Tab. 1).

4. Conclusion

- Our algorithm has effectively integrated fragmented ADT information across English EHRs, providing more complete treatment data, consistent with real-world treatment patterns and clinical practices.
- This real-world ADT data will enable precision medicine studies in prostate cancer outcomes.

Email: shamsudeen.mohammed1@lshtm.ac.uk