

# Presence of Codes for Indication for Use in Clinical Practice Research Datalink Aurum: An Assessment of Benign Prostatic Hyperplasia Treatments

Rebecca Persson <sup>1</sup>, Katrina Wilcox Hagberg <sup>1</sup>, Catherine Vasilakis-Scaramozza<sup>1</sup>, Eleanor Yelland<sup>2</sup>, Tim Williams <sup>2</sup>, Puja Myles <sup>2</sup>, Susan S Jick <sup>1,3</sup>

<sup>1</sup>Boston Collaborative Drug Surveillance Program, Lexington, MA, USA; <sup>2</sup>Clinical Practice Research Datalink, Medicines and Healthcare products Regulatory Agency, London, UK; <sup>3</sup>Boston University School of Public Health, Boston, MA, USA

Correspondence: Rebecca Persson, Boston Collaborative Drug Surveillance Program, 11 Muzzey Street, Lexington, MA, 02421, USA, Email [rpersson@bcdsp.org](mailto:rpersson@bcdsp.org)

**Background:** Assessments of strengths and limitations of new data sources are critical for making decisions about suitability for specific research questions. For some studies, it is necessary to capture a drug's indication for use.

**Objective:** To assess the presence of indications for prescription use in Clinical Practice Research Datalink (CPRD) Aurum (January 1988–June 2021) by describing the proportion of men in CPRD Aurum who had a recorded indication for use of prescriptions for 5-alpha reductase inhibitors (5-ARI), alpha blockers (AB), or tadalafil, which have multiple indications.

**Methods:** From a random sample of 154 practices of CPRD Aurum data, we selected 85,597 male patients with a prescription for a 5-ARI, an AB, or tadalafil. Among these patients, we described presence of codes indicating whether the patient had benign prostatic hyperplasia, hypertension, erectile dysfunction, or alopecia using three indication definitions: narrow (specific diagnoses recorded within one year before and up to 90 days after the prescription), broad (specific diagnoses or supporting clinical codes in the time period described above), and widest (diagnoses or supporting codes recorded at any time before the prescription and up to 90 days after the prescription).

**Results:** Using the narrow indication definition limited to diagnoses only, 39,861 (46.6%) patients' records contained an indication for use. The broad definitions, which additionally included supporting codes, captured indications for 62,912 (73.5%) patients and the widest definition, which additionally included supporting codes and all available data before the first prescription date, captured indications for 71,478 (83.5%) patients. Indications were present more often for prescriptions in 2005 and later (85.9%).

**Conclusion:** The findings of this assessment suggest that CPRD Aurum can be used for studies that require information on treatment indications for BPH and potentially for treatments of other chronic diseases managed in the primary care setting.

**Keywords:** Clinical Practice Research Datalink, CPRD Aurum, data element presence, pharmacoepidemiology, indication for use, prostatic hyperplasia

## Introduction

The Clinical Practice Research Datalink (CPRD) Aurum, a database of general practice electronic health records from the United Kingdom (UK), has been available to researchers since 2018.<sup>1</sup> Prior studies of CPRD Aurum have described the quality and completeness of recording of diagnoses (pulmonary embolism, myocardial infarction) compared with an external reference and of the internal consistency of recorded diagnoses with other clinical codes within CPRD Aurum (type 2 diabetes, hypercholesterolemia, and anemia). In each of these assessments, the correctness of diagnoses recorded in CPRD Aurum was high while completeness varied by condition.<sup>2–4</sup>

For some research questions, particularly in pharmacoepidemiology, it is necessary to capture a drug's indication for use to properly define a study population or exposure cohort. For example, a study of patients treated for benign prostatic

hyperplasia (BPH) may include patients with prescriptions for 5-alpha reductase inhibitors (5-ARIs), alpha blockers (ABs), and tadalafil (phosphodiesterase-5 (PDE-5) inhibitor). In addition to BPH, many of these drugs are also indicated for other conditions: hypertension (ABs), erectile dysfunction (ED) (tadalafil), and alopecia (1 mg finasteride, a 5-ARI). While it may be preferable to address confounding by indication by limiting such a hypothetical study to patients with a BPH diagnosis as the indication for use, it is not known if this is feasible in CPRD Aurum. The objective of this study was to assess how often information on indication for use is captured in CPRD Aurum, an assessment of “element presence”,<sup>5</sup> by describing the proportion of men in CPRD Aurum who had a recorded indication for use of their prescriptions for a 5-ARI, AB, or tadalafil.

## Methods

### Data Resources

CPRD Aurum is provided by CPRD, a research service jointly supported by the Medicines and Healthcare products Regulatory Agency and the National Institute for Health Research (NIHR), as part of the UK Department of Health and Social Care. CPRD Aurum is a large, prospectively collected, population-based, anonymized medical record database which, at the time of data extraction for this study (June 2021), encompassed data on 1491 National Health Service (NHS) practices and 40 million patients.<sup>6</sup> Medical records include demographic information, prescription details, clinical events, referrals, hospital admissions, laboratory results, and other patient details, such as smoking, alcohol consumption, and height and weight as captured by general practitioners (GPs). CPRD Aurum includes electronic patient data starting in 1988 and contains data captured using the EMIS<sup>®</sup> patient management software as well as data that has been migrated into EMIS from other platforms (eg, Vision<sup>®</sup>, SystmOne<sup>®</sup>).<sup>7</sup> GPs in the UK function as the gatekeepers for all NHS care. Thus, providers of hospital and secondary care are required to send information on patient encounters to the GP. Thus, primary care records are expected to contain diagnoses made by specialists and consultants,<sup>8</sup> however some diagnoses may be captured in free-text fields which are unavailable to researchers. Furthermore, except for some speciality treatments, drugs are typically prescribed by GPs regardless of whether the GP or a consultant initiated treatment. In CPRD Aurum, diagnoses and other non-prescription data are coded using a combination of SNOMED CT (UK edition), Read Version 2, and local EMIS Web<sup>®</sup> software-specific codes that have been cross-mapped by CPRD to a single diagnostic code (“MedCode”). Prescriptions are coded using the Dictionary of Medicines and Devices (dm+d) codes which are a subset of the SNOMED CT terminology and are assigned a “ProdCode” by CPRD.<sup>9–11</sup>

### Study Population

From CPRD Aurum release June 2021, we randomly selected a subset of data containing 154 practices. (These 154 practices were similar to all practices in the full database with respect to mean age, proportion of male patients, mean length of record and geographic region. There was a smaller proportion [3%] of practices in the sample of practices that had migrated from Vision [ie, CPRD GOLD] than in CPRD Aurum overall [14%.]) From this subset, we selected male patients with a first prescription for any treatment for BPH (“study drugs”): 5-ARIs (dutasteride and finasteride), ABs (alfuzosin, doxazosin, indoramin, prazosin, terazosin, tamsulosin, tamsulosin + solifenacin), 5-ARI+AB combination treatment (dutasteride + tamsulosin) or tadalafil recorded on 1 January 1988 or later.

We assigned the index date to the date of the first study drug in the patient record. We also estimated a start date and end date of each patient’s active CPRD Aurum electronic record using the available registration, prescription, and clinical data ([Supplement A](#)). From this population, we excluded patients who were less than age 20 or older than age 100 on the index date, whose index date was after 31 December 2020 (end of study period) or after the patient’s estimated record end date, or whose records contained less than one year of data between the start of their electronic medical record and the index date (potential prevalent prescription). We also excluded patients with more than one class of study drug on the index date (with the exception of 5-ARI+AB combinations), those with more than one study drug in the same class on the index date and those with a study drug recorded on an invalid date (ie, dates that are missing).

## Assignment of Indications for Use

We assigned indications for use by drug type using three progressively broad definitions according to recorded diagnoses and condition-specific symptoms, referrals and other clinical codes recorded around the time of the index date (Table 1). The “narrow” definition included condition-specific diagnoses recorded within 365 days before the index date (first study drug prescription) through 90 days after the index date. This time period was used to capture information recorded by the GP early on in the patient’s clinical course as well as after the GP had received notice of a diagnosis from a specialist. The “broad” definition included condition-specific diagnoses or presence of symptoms and other clinical codes in the

**Table 1** Narrow, Broad and Widest Definitions for Assigning Indications for Use by Study Drug Class

Indication	Study Drug Class	Definition	
		Codes	Time Window
Alopecia narrow	5-ARI (finasteride only)*	Diagnosis code for alopecia, hair loss or balding	365 days before through 90 days after index date
Alopecia broad		Alopecia narrow definition or code for hair transplant, alopecia severity score or referral	365 days before through 90 days after index date
Alopecia widest		Same as alopecia broad	Any time before through 90 days after index date
BPH narrow	5-ARIs (all) 5-ARIs+ABs (all) ABs (all) Tadalafil (2012 and later)†	Diagnosis code for BPH, prostatism or prostatitis	365 days before through 90 days after index date
BPH broad		BPH narrow definition or BPH symptoms (eg, micturition, enuresis), or urology referral/visits†	365 days before through 90 days after index date; urology codes within 30 days before or after index date
BPH widest		Same as BPH broad	Any time before through 90 days after index date; urology codes within 30 days before or after index date
ED narrow	Tadalafil	Diagnosis code for ED or impotence	365 days before through 90 days after index date
ED broad		ED narrow definition or ED-related codes (eg, ED advice, ED clinic, procedures/devices), sexual or psychosexual dysfunction diagnosis codes	365 days before through 90 days after index date
ED widest		Same as ED broad	Any time before through 90 days after index date; urology codes within 30 days before or after index date
Hypertension narrow	AB monotherapy (except tamsulosin)	Diagnosis code for hypertension	365 days before through 90 days after index date
Hypertension broad		Hypertension narrow definition or elevated blood pressure diagnosis codes, SBP > 140 mmHg or DBP >90 mmHg within 30 days before index date	365 days before through 90 days after index date; high BP measurements within 30 days before index date
Hypertension widest		Same as hypertension broad	Any time before through 90 days after index date; high BP measurements within 30 days before index date

**Notes:** \*Alopecia indication was not limited to the 1mg finasteride dose. However, only 10% of patients with an indication for use of alopecia had a prescription for 5mg finasteride. †Urology codes were not used to differentiate the indication for use of tadalafil (BPH versus ED).

same time period defined above, or specialist visits within 30 days before or after the index date. The “widest” definition included all condition-specific diagnoses, symptoms and other clinical codes recorded at any time before the index date through 90 days after, or any specialist visits within 30 days before or after the index date. Note that although the structure of CPRD Aurum allows for prescriptions to be linked to diagnoses and symptoms through “problem” records,<sup>1</sup> these linkages were not present for any drug prescriptions assessed in this study. All indications for use were limited to those approved for each drug type. BPH as an indication for tadalafil use was limited to prescriptions recorded in 2012 or later (when the indication for tadalafil was expanded by the European Medicines Agency to include BPH). See [Table 1](#). Code lists are available in [Supplement B](#).

## Assessment of the Presence of Indication for Use

We described the proportion of patients with a recorded indication for use among all patients and by study drug class (5-ARIs+/- ABs, AB monotherapy, or tadalafil) using progressively broad definitions of indication: narrow, broad and widest for each condition. We then described coding patterns by age and calendar time. We repeated each assessment within two sensitivity analysis populations: 1) among patients with no history of prostate cancer, prostatectomy or other procedures of the prostate, urethra, or bladder neck, as some study drugs are used off-label in these patients and 2) among patients with their first study drug prescription in 2005 or later after the implementation of Quality and Outcomes Framework (QOF) incentives.<sup>12</sup>

Lastly, we conducted a subgroup analysis among patients with 5-ARI prescriptions who were likely to have BPH as the indication for use. We selected patients with no alopecia codes who had a prescription for dutasteride or finasteride (5mg dose) monotherapy recorded in 2005 or later (after the implementation of data quality QOFs). Among this subpopulation of likely BPH patients, we compared whether patient characteristics and BPH disease severity varied by the definition used to capture the BPH indication. In this analysis, we categorized patients according to mutually exclusive categories using the most stringent definition met by their record: narrow, broad not including narrow, and widest exclusive of narrow or broad, or none.

## Results

The final study population included 85,597 patients with a first prescription for a study drug: 9095 (10.6%) received 5-ARIs (with or without ABs), 60,457 (70.6%) received AB monotherapy, and 16,045 (18.7%) received tadalafil ([Table 2](#), [Supplement C](#)). The median record length before the index date was approximately 10 years for all study drugs. There were differences in patient characteristics by study drug class. Men with prescriptions for 5-ARIs and ABs were older than those with tadalafil. 5-ARI use was similarly distributed across calendar time after 2000, AB use increased over time, and tadalafil use was highest in the years 2005 through 2014. “Unknown” BMI, smoking and race/ethnicity data were highest in 5-ARI±AB users. History of prostate cancer and diabetes was highest among users of tadalafil, while cardiovascular disease was highest among users of 5-ARIs ([Table 2](#)).

Among users of all study drugs, 39,861 (46.6%) had a recorded indication for use using the narrow definitions (condition-specific diagnoses recorded within one year before and up to 90 days after the index date). The proportion of patients with a recorded indication for use increased with broader definitions. The broad definition (diagnoses, symptoms and other clinical codes within one year before and up to 90 days after the index date) captured indications for use for 62,912 (73.5%) patients and the widest definition (diagnoses, symptoms and other clinical codes at any time before or up to 90 days after the index date) captured indications for use for 71,478 (83.5%) patients ([Figure 1](#)). Using the narrow definition, the proportion of patients with an indication for use was highest among tadalafil users. However, using the widest definition, the proportion was highest among AB only users ([Table 3](#)). Overall, the results were similar when patients with a history of prostate cancer, prostatectomy or other procedures of the prostate, urethra or bladder neck were excluded. ([Supplement D](#)) When the analysis was limited to patients with an index date in 2005 and later, 53,833 of 62,665 (85.9%) users of any study drug had a recorded indication for use using the widest definition ([Supplement E](#)).

The proportion of patients with a recorded indication for use was similar across most study drug categories stratified by age but was somewhat lower among 5-ARI±AB users younger than age 40. Capture of indication for use was higher in 2005 and later than in earlier years ([Table 4](#)). Results were similar in both sensitivity analyses ([Supplements F and G](#)).

**Table 2** Patient and CPRD Aurum Record Characteristics at Index Date by Study Drug Class on Index Date

Characteristic at Index Date	All Study Drugs N=85597 (%)	5-ARI ± AB* N=9095 (%)	AB Only N=60457 (%)	Tadalafil N=16045 (%)
Age				
20–39	3233 (3.8)	459 (5.0)	1439 (2.4)	1335 (8.3)
40–59	24,061 (28.1)	1057 (11.6)	15,229 (25.2)	7775 (48.5)
60–79	47,296 (55.3)	5310 (58.4)	35,241 (58.3)	6745 (42.0)
80–99	11,007 (12.9)	2269 (24.9)	8548 (14.1)	190 (1.2)
Median (IQR)	66 (57–74)	72 (64–79)	67 (58–75)	58 (49–65)
Calendar Year				
1988–1994	1263 (1.5)	348 (3.8)	915 (1.5)	0 (0.0)
1995–1999	6247 (7.3)	1084 (11.9)	5163 (8.5)	0 (0.0)
2000–2004	15,422 (18.0)	1704 (18.7)	11,468 (19.0)	2250 (14.0)
2005–2009	20,232 (23.6)	2063 (22.7)	11,980 (19.8)	6189 (38.6)
2010–2014	21,319 (24.9)	1956 (21.5)	14,052 (23.2)	5311 (33.1)
2015–2020	21,114 (24.7)	1940 (21.3)	16,879 (27.9)	2295 (14.3)
BMI				
<18.5	488 (0.6)	74 (0.8)	347 (0.6)	67 (0.4)
18.5 to <25	18,340 (21.4)	2366 (26.0)	12,937 (21.4)	3037 (18.9)
25 to <30	30,211 (35.3)	3063 (33.7)	21,374 (35.4)	5774 (36.0)
≥30	20,068 (23.4)	1407 (15.5)	14,381 (23.8)	4280 (26.7)
Unknown	16,490 (19.3)	2185 (24.0)	11,418 (18.9)	2887 (18.0)
Smoking				
Current	14,993 (17.5)	1288 (14.2)	9877 (16.3)	3828 (23.9)
Former smoker	37,536 (43.9)	4077 (44.8)	27,134 (44.9)	6325 (39.4)
Non-smoker	28,765 (33.6)	2917 (32.1)	20,293 (33.6)	5555 (34.6)
Unknown	4303 (5.0)	813 (8.9)	3153 (5.2)	337 (2.1)
Ethnicity				
White/White British	57,149 (66.8)	5569 (61.2)	40,417 (66.9)	11,163 (69.6)
Asian/British Asian	2955 (3.5)	203 (2.2)	1981 (3.3)	771 (4.8)
Black/African/Caribbean/Black British	1199 (1.4)	87 (1.0)	828 (1.4)	284 (1.8)
Multiple	1212 (1.4)	95 (1.0)	817 (1.4)	300 (1.9)
Other specified ethnicity	85 (0.1)	11 (0.1)	57 (0.1)	17 (0.1)
Other unspecified ethnicity	41 (0.0)	7 (0.1)	22 (0.0)	12 (0.1)
Refused/Not available/Not recorded	22,956 (26.8)	3123 (34.3)	16,335 (27.0)	3498 (21.8)

(Continued)

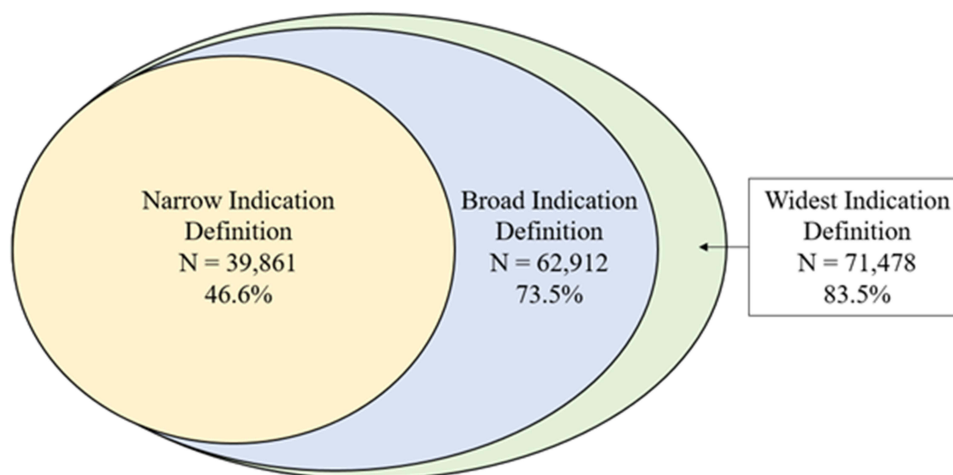
**Table 2** (Continued).

Characteristic at Index Date	All Study Drugs N=85597 (%)	5-ARI ± AB* N=9095 (%)	AB Only N=60457 (%)	Tadalafil N=16045 (%)
Length of active record before index date [median (IQR), yr]	10 (5–16)	9 (5–16)	10 (5–16)	10 (5–15)
Comorbidities on or before index date				
Prostate cancer	4037 (4.7)	241 (2.6)	2805 (4.6)	991 (6.2)
Prostatectomy	945 (1.1)	67 (0.7)	294 (0.5)	584 (3.6)
TURP and other procedures <sup>†</sup>	5244 (6.1)	1178 (13.0)	3560 (5.9)	506 (3.2)
Diabetes	67,368 (78.7)	6252 (68.7)	47,371 (78.4)	13,745 (85.7)
Cardiovascular disease	23,606 (27.6)	3339 (36.7)	17,661 (29.2)	2606 (16.2)

Notes: \*N= 6744 users of 5-ARI only and N= 2351 users of 5-ARI+AB. <sup>†</sup>Transurethral resection of the prostate and other procedures on prostate, urethra and bladder neck.

In the subgroup of patients with 5-ARI prescriptions and likely BPH (ie, a dutasteride or 5 mg finasteride prescription and no alopecia-related codes), patients with no recorded indication for use (ie, “none” in Table 5) were more likely to be less than 60 years old than patients with a recorded indication for use. These patients were also slightly more likely to be missing BMI, smoking, and race/ethnicity information. Patients without a coded indication were also less likely to have BPH-related complications coded in their CPRD Aurum records before or after the index date than patients with a recorded indication for use (Table 5).

In this same subgroup analysis, patients who met the narrow BPH definition were somewhat younger and were less likely to have a history of prostate cancer, transurethral resection of the prostate (TURP) or other procedures of the urethra, prostate or bladder neck before the index date than patients who met either the broad (not including narrow definition) or widest (exclusive of broad or narrow) definitions. However, patients who met the narrow criteria were more likely to have had a prostate-specific antigen (PSA) test within the year before the index date. There were fewer differences in the period after the index date between patients who met various BPH definitions. The proportions of patients with urinary tract and kidney infections, prostate cancer and prostatectomy after the index date were similar across groups of patients whose records met the narrow, broad (not including narrow definition) and widest (exclusive of broad or narrow) definitions for BPH. TURP and other procedures after the index date were somewhat less likely in patients who met the widest BPH criteria (Table 5).

**Figure 1** Numbers of patients with recorded indications captured by various definitions (N = 85,597).

**Table 3** Recorded Indications for Use with Increasingly Broad Definitions, by Study Drug

Indication	All Study Drugs N=85597 (%)	5-ARI ± AB N=9095 (%)	AB Only N=60457 (%)	Tadalafil N=16045 (%)
Any indication				
Narrow	39,861 (46.6)	3685 (40.5)	25,564 (42.3)	10,612 (66.1)
Broad	62,912 (73.5)	6384 (70.2)	45,459 (75.2)	11,069 (69.0)
Widest	71,478 (83.5)	7162 (78.7)	51,508 (85.2)	12,808 (79.8)
BPH*				
Narrow	18,433 (21.5)	3288 (36.2)	15,070 (24.9)	75 (0.5)
Broad	39,105 (45.7)	5995 (65.9)	32,335 (53.5)	775 (4.8)
Widest	46,854 (54.7)	6783 (74.6)	38,468 (63.6)	1603 (10.0)
Hypertension *				
Narrow	11,367 (13.3)	0 (0.0)	11,367 (18.8)	0 (0.0)
Broad	15,715 (18.4)	0 (0.0)	15,715 (26.0)	0 (0.0)
Widest	19,099 (22.3)	0 (0.0)	19,099 (31.6)	0 (0.0)
ED*				
Narrow	10,580 (12.4)	0 (0.0)	0 (0.0)	10,580 (65.9)
Broad	10,745 (12.6)	0 (0.0)	0 (0.0)	10,745 (67.0)
Widest	12,442 (14.5)	0 (0.0)	0 (0.0)	12,442 (77.5)
Alopecia*				
Narrow	400 (0.5)	400 (4.4)	0 (0.0)	0 (0.0)
Broad	408 (0.5)	408 (4.5)	0 (0.0)	0 (0.0)
Widest	472 (0.6)	472 (5.2)	0 (0.0)	0 (0.0)

**Notes:** \*Patients may have diagnoses/clinical codes consistent with more than one indication for use. Among users of all study drugs and using the widest definitions: N = 6059 (7.1%) BPH+HTN, N = 1237 (1.4%) BPH+ED, and N = 93 (0.1%) BPH+alopecia.

## Discussion

Capturing the indication for use of prescribed treatments is important for many pharmacoepidemiological studies. In this assessment of CPRD Aurum data, the indications for use were present in the coded electronic record in 83.5% of male patients with prescriptions for BPH treatments. Capture was lower (46.6%) if the indication definition was limited to diagnoses only (excluding symptoms and other clinical codes) or was restricted to the time period around the first prescription. Coded indications for use were more likely to be present for prescriptions recorded in 2005 and later (ie, after the implementation of data quality QOFs).<sup>12</sup> These findings support the use of CPRD Aurum for studies that require information on treatment indications for BPH as well as hypertension, erectile dysfunction, alopecia and potentially other chronic diseases managed in the primary care setting.

We chose to assess BPH treatments for this evaluation as BPH is a chronic condition generally managed by a GP and because the drugs prescribed for BPH have multiple indications for use. In their role as the gatekeeper of clinical care in the NHS, it is expected that a patient's GP would record diagnoses and prescribe continuing treatments regardless of where the initial diagnosis of BPH, hypertension, ED or alopecia was made. As CPRD Aurum is a relatively new data source, it was not known whether GPs reliably record these details in a way that is accessible to researchers. In this assessment, an indication for use was captured for most (>80%) patients with prescriptions for the study drugs of interest.

**Table 4** Presence of Any Indication (Widest Definitions) by Study Drug Class, Age and Calendar Year

Characteristic at Index Date	n with Recorded Indication/N Total (%)			
	All Study Drugs	5-ARI±AB	AB Only	Tadalafil Only
All	71,478/85,597 (84%)	7162/9095 (79%)	51,508/60,457 (85%)	12,808/16,045 (80%)
Age				
20–39	2532/3233 (78%)	330/459 (72%)	1158/1439 (80%)	1044/1335 (78%)
40–59	20,157/24,061 (84%)	796/1057 (75%)	13,133/15,229 (86%)	6228/7775 (80%)
60–79	39,776/47,296 (84%)	4249/5310 (80%)	30,137/35,241 (86%)	5390/6745 (80%)
80–99	9013/11,007 (82%)	1787/2269 (79%)	7080/8548 (83%)	146/190 (77%)
Calendar Year				
1988–1994	911/1263 (72%)	226/348 (65%)	685/915 (75%)	NA
1995–1999	4432/6247 (71%)	642/1084 (59%)	3790/5163 (73%)	NA
2000–2004	12,302/15,422 (80%)	1222/1704 (72%)	9458/11,468 (82%)	1622/2250 (72%)
2005–2009	16,815/20,232 (83%)	1707/2063 (83%)	10,369/11,980 (87%)	4739/6189 (77%)
2010–2014	18,229/21,319 (86%)	1679/1956 (86%)	12,162/14,052 (87%)	4388/5311 (83%)
2015–2020	18,789/21,114 (89%)	1686/1940 (87%)	15,044/16,879 (89%)	2059/2295 (90%)

**Note:** Tadalafil was first marketed in UK in 2003.

**Abbreviation:** NA, not applicable.

**Table 5** Patient and CPRD Aurum Record Characteristics for Users of Dutasteride or 5mg Finasteride Monotherapy with and without Recorded BPH Indication with First Study Drug Prescription in 2005 or Later

Characteristic	All Users in Subgroup*	BPH Definition			
		Narrow	Broad Not Including Narrow	Widest Exclusive of Narrow or Broad	None <sup>†</sup>
		N=3341 (%)	N=1259 (%)	N=1258 (%)	N=292 (%)
Age at index date					
20–59	410 (12.2)	171 (13.6)	104 (8.3)	29 (9.9)	106 (19.9)
60–79	1904 (57.0)	798 (63.4)	659 (52.4)	176 (60.3)	271 (50.9)
80–99	1027 (30.7)	290 (23.0)	495 (39.3)	87 (29.8)	155 (29.1)
Median (IQR)	74 (66–81)	71 (65–79)	77 (69–83)	75 (67–82)	72 (62–81)
Calendar year of index date					
2005–2009	1324 (39.6)	554 (44.0)	414 (32.9)	126 (43.2)	230 (43.2)
2010–2014	1042 (31.2)	378 (30.0)	405 (32.2)	84 (28.8)	175 (32.9)
2015–2020	975 (29.2)	327 (26.0)	439 (34.9)	82 (28.1)	127 (23.9)
BMI at index date					
<18.5	26 (0.8)	8 (0.6)	9 (0.7)	2 (0.7)	7 (1.3)
18.5 to <25	927 (27.7)	325 (25.8)	359 (28.5)	87 (29.8)	156 (29.3)

(Continued)



Table 5 (Continued).

Characteristic	All Users in Subgroup*	BPH Definition			
		Narrow	Broad Not Including Narrow	Widest Exclusive of Narrow or Broad	None <sup>†</sup>
		N=3341 (%)	N=1259 (%)	N=1258 (%)	N=292 (%)
25 to <30	1262 (37.8)	490 (38.9)	485 (38.6)	103 (35.3)	184 (34.6)
≥30	631 (18.9)	245 (19.5)	228 (18.1)	61 (20.9)	97 (18.2)
Unknown	495 (14.8)	191 (15.2)	177 (14.1)	39 (13.4)	88 (16.5)
Smoking at index date					
Current	451 (13.5)	165 (13.1)	162 (12.9)	35 (12.0)	89 (16.7)
Former smoker	1720 (51.5)	643 (51.1)	672 (53.4)	151 (51.7)	254 (47.7)
Non-smoker	1143 (34.2)	443 (35.2)	416 (33.1)	103 (35.3)	181 (34.0)
Unknown	27 (0.8)	8 (0.6)	8 (0.6)	<5 (1.0)	8 (1.5)
Ethnicity					
White/White British	2450 (73.3)	944 (75.0)	943 (75.0)	208 (71.2)	355 (66.7)
Asian/British Asian	76 (2.3)	20 (1.6)	30 (2.4)	6 (2.1)	20 (3.8)
Black/African/Caribbean/Black British	36 (1.1)	6 (0.5)	15 (1.2)	6 (2.1)	9 (1.7)
Multiple	33 (1.0)	14 (1.1)	11 (0.9)	5 (1.7)	<5 (0.6)
Other ethnicity	9 (0.3)	<5 (0.2)	5 (0.4)	0 (0.0)	<5 (0.2)
Refused/Not available/Not recorded	737 (22.1)	272 (21.6)	254 (20.2)	67 (22.9)	144 (27.1)
Length of active record before index date (median (IQR))	13 (7–18)	13 (7–18)	14 (7–19)	13 (7–18)	12 (6–17)
Length of active record after index date (median (IQR))	5 (2–9)	6 (3–10)	4 (2–8)	5 (2–9)	5 (2–10)
<b>BPH complications and comorbidities on or before index date</b>					
Urinary tract infections	478 (14.3)	171 (13.6)	213 (16.9)	49 (16.8)	45 (8.5)
Kidney infections	20 (0.6)	5 (0.4)	12 (1.0)	0 (0.0)	<5 (0.6)
Bladder stones	39 (1.2)	10 (0.8)	20 (1.6)	5 (1.7)	<5 (0.8)
Prostate cancer	117 (3.5)	12 (1.0)	58 (4.6)	27 (9.2)	20 (3.8)
Prostatectomy	36 (1.1)	8 (0.6)	22 (1.7)	<5 (1.0)	<5 (0.6)
TURP and other procedures <sup>‡</sup>	537 (16.1)	152 (12.1)	266 (21.1)	78 (26.7)	41 (7.7)
Diabetes, any	2990 (89.5)	1128 (89.6)	1125 (89.4)	275 (94.2)	462 (86.8)
CVD, any	1378 (41.2)	482 (38.3)	588 (46.7)	124 (42.5)	184 (34.6)
PsA test in year before index date	1961 (58.7)	872 (69.3)	690 (54.8)	140 (47.9)	259 (48.7)
<b>BPH complications and comorbidities after index date</b>					
Urinary tract infections	466 (13.9)	181 (14.4)	184 (14.6)	46 (15.8)	55 (10.3)
Kidney infections	12 (0.4)	<5 (0.2)	6 (0.5)	<5 (0.7)	<5 (0.2)

(Continued)

Table 5 (Continued).

Characteristic	All Users in Subgroup*	BPH Definition			
		Narrow	Broad Not Including Narrow	Widest Exclusive of Narrow or Broad	None <sup>†</sup>
	N=3341 (%)	N=1259 (%)	N=1258 (%)	N=292 (%)	N=532 (%)
Bladder stones	31 (0.9)	12 (1.0)	10 (0.8)	5 (1.7)	<5 (0.8)
Prostate cancer (new)	175 (5.2)	61 (4.8)	69 (5.5)	11 (3.8)	34 (6.4)
Prostatectomy	30 (0.9)	10 (0.8)	10 (0.8)	<5 (1.4)	6 (1.1)
TURP and other procedures <sup>‡</sup>	383 (11.5)	146 (11.6)	161 (12.8)	28 (9.6)	48 (9.0)
Diabetes (new)	256 (7.7)	104 (8.3)	94 (7.5)	10 (3.4)	48 (9.0)
CVD (new)	417 (12.5)	172 (13.7)	142 (11.3)	38 (13.0)	65 (12.2)

Notes: \*All users of dutasteride or 5mg finasteride monotherapy on the index date excluding N = 47 patients with 5 mg finasteride and a recorded alopecia diagnosis. <sup>†</sup>None = Patients whose records met no BPH definitions. <sup>‡</sup>Transurethral resection of the prostate and other procedures on prostate, urethra and bladder neck.

Capture was lower (<50%) when only those diagnoses recorded close to the index date were considered. This is consistent with the structure of primary care electronic health records which, unlike in claims data, may, depending on individual GP recording patterns as well as incentives such as QOFs, only code the presence of a chronic condition at first diagnosis. Researchers using CPRD data would be advised to use all available data before an index or cohort entry date to improve capture of chronic conditions. Issues related to differences in record length may be addressed through matching or other study design choices.

Among patients with a dutasteride or 5mg finasteride prescription and no alopecia (likely BPH), patients without a coded indication for use were more likely to be younger than age 60 and to be missing race/ethnicity information. They were also less likely to have coded diagnoses for BPH-related complications. We cannot know from this assessment whether these patients had less severe disease than patients with a coded BPH indication or whether their GPs were less likely to code any diagnoses in a way that is available to researchers. Among patients whose records did include a coded indication for use, patients whose record met the narrow BPH definition (a diagnosis code close to the index date) were somewhat less likely to have prostate cancer, TURP or related procedures before index date than patients whose records met the broad definition not including narrow, or the widest exclusive of narrow or broad definitions. However, after the index date the proportion of patients with BPH-related complications was similar regardless of which BPH definition they met. Thus, it is likely that severity of disease is not directly associated with the BPH definition used to capture the patient.

It is possible that we misclassified the indication for use for some prescriptions. We reviewed all codes present in the year before and after the index date in a sample of patient records so it is unlikely that we missed important diagnosis or supporting codes, but it is possible that we misinterpreted some codes for symptoms, or referrals, that were associated with other conditions. It is also possible that for some patients, symptoms recorded well before prescription date may have resolved independently and were unrelated to the study prescription. Furthermore, we did not fully assess how these definitions performed among patients with records consistent with more than one indication for use or among patients with concomitant use of more than one class of study drugs. However, the goal of this study was to describe the presence of coded indications for use of BPH treatments, not to assess whether the coded diagnoses were “correct” compared to other internal or external data. Formal validation studies of each condition are still necessary. Prior studies have found correctness of diagnoses recorded in CPRD Aurum to be high for a range of conditions, though completeness/missingness of diagnosis recording in CPRD Aurum varied from ~50% to >90% depending on the condition, QOF reporting requirements, and the method of assessment.<sup>2-4</sup>

The findings of this study demonstrate that indications for use of BPH treatments are present in CPRD Aurum in a large proportion of patients. It is likely that GP recording practices are similar for treatments of other chronic conditions managed

in the primary care setting in the UK. Thus, these results suggest that CPRD Aurum may be useful for research in other clinical areas where capture of indication information is important. Researchers are advised to use symptoms and other supporting clinical codes in addition to diagnosis codes to identify the indication for use including all available data before the first prescription date but recognizing the potential for misclassification of treatment indication, especially where a drug has multiple indications for use. Researchers should also consider methods, such as matching by GP, to reduce biases related to differential recording of drug indication between GPs. Depending on the study objectives, validation of treatment indication through GP questionnaires may be necessary to achieve sufficiently accurate information.<sup>13</sup>

## Ethical Review and Copyright Statement

This study is based in part on data from the Clinical Practice Research Datalink obtained under license from the UK Medicines and Healthcare products Regulatory Agency. The data is provided by patients and collected by the NHS as part of their care and support. The interpretation and conclusions contained in this study are those of the authors alone. This study was approved by the Independent Scientific Advisory Committee (ISAC) for Medicines and Healthcare products Regulatory Agency (protocol no:18\_191A), and the protocol was made available to the journal reviewers upon request. Researchers can apply for a limited licence to access CPRD data for public health research, subject to individual research protocols meeting CPRD data governance requirements. More details including data specification, licence fees and applications process are available on the CPRD website (<https://www.cprd.com>).

## Funding

This study was unfunded.

## Disclosure

CPRD is jointly sponsored by the UK government's Medicines and Healthcare products Regulatory Agency and the National Institute for Health Research (NIHR). As a not-for-profit UK government body, CPRD seeks to recoup the cost of delivering its research services to academic, industry and government researchers through research user license fees. PM, EY and TW are employees of CPRD, the data custodians for CPRD Aurum. No funding was received by the Boston Collaborative Drug Surveillance Program (BCDSP) for the conduct of this study. The BCDSP receives industry funding to conduct research using CPRD data. The authors report no other conflicts of interest in this work.

## References

1. Wolf A, Dedman D, Campbell J, et al. Data resource profile: Clinical Practice Research Datalink (CPRD) Aurum. *Int J Epidemiol*. 2019;48:1740-1740g. doi:10.1093/ije/dyz034
2. Jick S, Hagberg KW, Persson R, et al. Quality and completeness of diagnoses recorded in the new CPRD Aurum Database: evaluation of pulmonary embolism. *Pharmacoepidemiol Drug Saf*. 2020;29(9):1134-1140. doi:10.1002/pds.4996
3. Persson R, Vasilakis-Scaramozza C, Hagberg KW, et al. CPRD Aurum database: assessment of data quality and completeness of three important comorbidities. *Pharmacoepidemiol Drug Saf*. 2020;29:1456-1464. doi:10.1002/pds.5135
4. Persson R, Sponholtz T, Vasilakis-Scaramozza C, et al. Quality and completeness of myocardial infarction recording in Clinical Practice Research Datalink Aurum. *Clin Epidemiol*. 2021;13:745-775. doi:10.2147/CLEP.S319245
5. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *J Am Med Informatics Assoc*. 2013;20(1):144-151. doi:10.1136/amiajnl-2011-000681
6. CPRD. CPRD Aurum 2021 dataset. *medRxiv*. 2021. doi:10.48329/pyc2-we97
7. Kontopantelis E, Stevens RJ, Helms PJ, Edwards D, Doran T, Ashcroft DM. Spatial distribution of clinical computer systems in primary care in England in 2016 and implications for primary care electronic medical record databases: a cross-sectional population study. *BMJ Open*. 2018;8:e020738. doi:10.1136/bmjopen-2017-020738
8. Herrett E, Gallagher AM, Bhaskaran K, et al. Data resource profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol*. 2015;44:827-836. doi:10.1093/ije/dyv098
9. NHS Digital. SNOMED CT. Available from: <https://digital.nhs.uk/services/terminology-and-classifications/snomed-ct>. Accessed April 28, 2022.
10. NHS Digital. Read codes. Available from: <https://digital.nhs.uk/services/terminology-and-classifications/read-codes>. Accessed April 28, 2022.
11. NHS Business Services Authority. Dictionary of medicines and devices (dm+d). Available from: <https://www.nhsbsa.nhs.uk/pharmacies-gp-practices-and-appliance-contractors/dictionary-medicines-and-devices-dmd>. Accessed April 28, 2022.
12. Taggar JS, Coleman T, Lewis S, Szatkowski L. The impact of the Quality and Outcomes Framework (QOF) on the recording of smoking targets in primary care medical records: cross-sectional analyses from The Health Improvement Network (THIN) database. *BMC Public Health*. 2012;12:329. doi:10.1186/1471-2458-12-329
13. CPRD. CRPD PROVE (PProviding Online Verification of EHR). Available from: <https://cprd.com/cprd-prove>. Accessed April 28, 2022.

Clinical Epidemiology

Dovepress

## Publish your work in this journal

Clinical Epidemiology is an international, peer-reviewed, open access, online journal focusing on disease and drug epidemiology, identification of risk factors and screening procedures to develop optimal preventative initiatives and programs. Specific topics include: diagnosis, prognosis, treatment, screening, prevention, risk factor modification, systematic reviews, risk & safety of medical interventions, epidemiology & biostatistical methods, and evaluation of guidelines, translational medicine, health policies & economic evaluations. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use.

Submit your manuscript here: <https://www.dovepress.com/clinical-epidemiology-journal>