

# BMJ Open Identification of antithrombotic drugs related to total joint replacement using anonymised free-text notes: a search strategy in the Clinical Practice Research Datalink

Johannes TH Nielen,<sup>1,2</sup> Bart J F van den Bemt,<sup>3,4</sup> Annelies Boonen,<sup>5</sup> Pieter C Dagnelie,<sup>6</sup> Pieter J Emans,<sup>7</sup> Nicole Veldhorst,<sup>8</sup> Arief Lalmohamed,<sup>9</sup> Tjeerd-Pieter van Staa,<sup>1</sup> Frank de Vries<sup>1,8</sup>

**To cite:** Nielen JTH, van den Bemt BJF, Boonen A, *et al.* Identification of antithrombotic drugs related to total joint replacement using anonymised free-text notes: a search strategy in the Clinical Practice Research Datalink. *BMJ Open* 2015;**5**:e009017. doi:10.1136/bmjopen-2015-009017

► Prepublication history and additional material is available. To view please visit the journal (<http://dx.doi.org/10.1136/bmjopen-2015-009017>).

Received 9 June 2015  
Revised 9 November 2015  
Accepted 10 November 2015



CrossMark

For numbered affiliations see end of article.

**Correspondence to**  
Johannes T.H Nielen;  
[j.t.h.nielen@uu.nl](mailto:j.t.h.nielen@uu.nl)

## ABSTRACT

**Objectives:** We aimed to design and test a method to extract information on antithrombotic therapy from anonymised free-text notes in the Clinical Practice Research Datalink (CPRD).

**Setting:** General practice database representative of the UK.

**Participants:** All patients undergoing total hip replacement (THR, n=25 898) or total knee replacement (TKR, n=22 231) between January 2008 and October 2012 were included. Antithrombotic drug use related to THR or TKR was identified using anonymised free text and prescription data.

**Primary and secondary outcome measures:** Internal validity of our newly designed method was determined by calculating positive predictive values (PPVs) of hits for predefined keywords in a random sample of anonymised free-text notes. In order to determine potential detection bias, total joint replacement (TJR) patient characteristics were compared as per their status of exposure to antithrombotics.

**Results:** PPVs ranging between 97% and 99% for new oral anticoagulants (NOAC) or low-molecular weight heparins (LMWH) exposure related to TJR were obtained with our method. Our search strategy increased detection rates by 57%, yielding a total proportion of 18.5% of all THR and 18.6% of all TKR surgeries. Identified users of NOACs and LMWHs were largely similar with regards to age, sex, lifestyle, disease and drug history compared to patients without identified drug use.

**Conclusions:** We have developed a useful method to identify additional exposure to NOACs or LMWHs with TJR surgery.

## INTRODUCTION

Osteoarthritis (OA) is the most common musculoskeletal condition in older people in

## Strengths and limitations of this study

- This is the first study reporting an effective method to identify antithrombotic drug use in anonymised free text in a large population-based database.
- Our method can be implemented at relatively low costs.
- We were unable to determine the specificity and sensitivity of our method.

the UK.<sup>1</sup> Based on 7-year consultation prevalence in British general practice, an estimated 8.75 million people have been treated for OA. This constitutes one-third people aged 45 years and over in the UK.<sup>1</sup> Total joint replacements (TJR), such as total hip replacement (THR) or total knee replacement (TKR), substantially improve quality of life in these patients.<sup>2</sup> However, risk of potentially fatal venous thromboembolic events, such as deep-vein thrombosis (DVT), is increased up to 14-fold following these surgical procedures when no drug for thromboprophylaxis is given.<sup>3</sup> Intensive antithrombotic treatment up to 35 days is, therefore, recommended and has resulted in a reduced risk of asymptomatic DVT in a meta-analysis of randomised trials (RR=0.51, 95% CI (0.45 to 0.59), p<0.001).<sup>4 5</sup>

There has been much debate about the risk/benefit ratio of antithrombotic agents following THR or TKR surgery.<sup>6–8</sup> Low-molecular weight heparins (LMWHs) have gained popularity in the past decade, but can only be administered subcutaneously. The recently introduced orally administered direct thrombin inhibitors and direct factor Xa inhibitors (eg, dabigatran and

rivaroxaban) seem to combine both advantages of vitamin-K antagonists (VKAs) and LMWHs. Their benefit-risk profile has been suggested to outperform VKAs in patients with atrial fibrillation.<sup>9</sup> To further understand the benefit/risk profile of these antithrombotic drugs in real life use, data from large databases, such as the Clinical Practice Research Datalink (CPRD), could contribute largely. However, some antithrombotic therapy are 'red-listed' and therefore, predominantly dispensed during inpatient hospitalisations and the period postdischarge. Consequently, patients do not have to visit a public pharmacy to repeat their prescription. Unfortunately, in-hospital pharmacy data is often lacking in general practitioner (GP) databases. Therefore, there are limitations in using these databases to conduct drug utilisation or drug effect studies. However, Martinez *et al*<sup>10</sup> have found a feasible approach to overcome some of these restrictions. They identified a total of 754 rivaroxaban users by combining information from prescription files and electronic search of anonymised free text in the CPRD between October 2008 and January 2012. Approximately 66% of these users were identified by anonymised free text only, thereby tripling the information on exposure to rivaroxaban. Unfortunately, details of the electronic search method used were lacking in this study and it focused only on one drug (rivaroxaban).

In order to assess safety and efficacy of the new oral anticoagulants (NOACs), new methods are needed to efficiently extract the additional information from anonymised free-text notes. In some fields this has already been done. For example, Shah designed an algorithm that identifies dosage instructions for a variety of drugs in anonymised free text. Their algorithm converted unstructured data from these notes to structured, usable data. They subsequently validated their structured data by manually checking a random sample of these records. Their algorithm identified correct dosage information in 99% of the analysed anonymised free text.<sup>11</sup> Since in-hospital pharmacy data are not included in most primary care databases, we aimed to design and test a method to extract additional information on antithrombotic therapy in patients undergoing TJR surgery from anonymised free-text notes in the CPRD.

## METHODS

### Data source

This study was conducted using CPRD GOLD, formally known as the General Practice Research Database (GPRD). The CPRD is a GP database currently containing approximately 8% of the UK population. GPs play a key role in the UK healthcare system, as they are responsible for primary healthcare and specialist referrals. Consequently, medical information from GPs, specialist referrals and hospitalisation data are recorded in the CPRD.<sup>12</sup> Most data are recorded using CPRD's therapy files, but some data, such as hospital discharge

summaries, are collected as anonymised free text. CPRD has been used to study outcomes after TJR surgery.<sup>13 14</sup>

### Study population

The study population comprised all patients who underwent a primary THR or TKR surgery from January 2008 to October 2012. This time frame was chosen because NOACs have been registered for antithrombotic therapy after TKR and THR in Europe since 2008 (dabigatran, 18 March 2008; rivaroxaban, 13 September 2008).

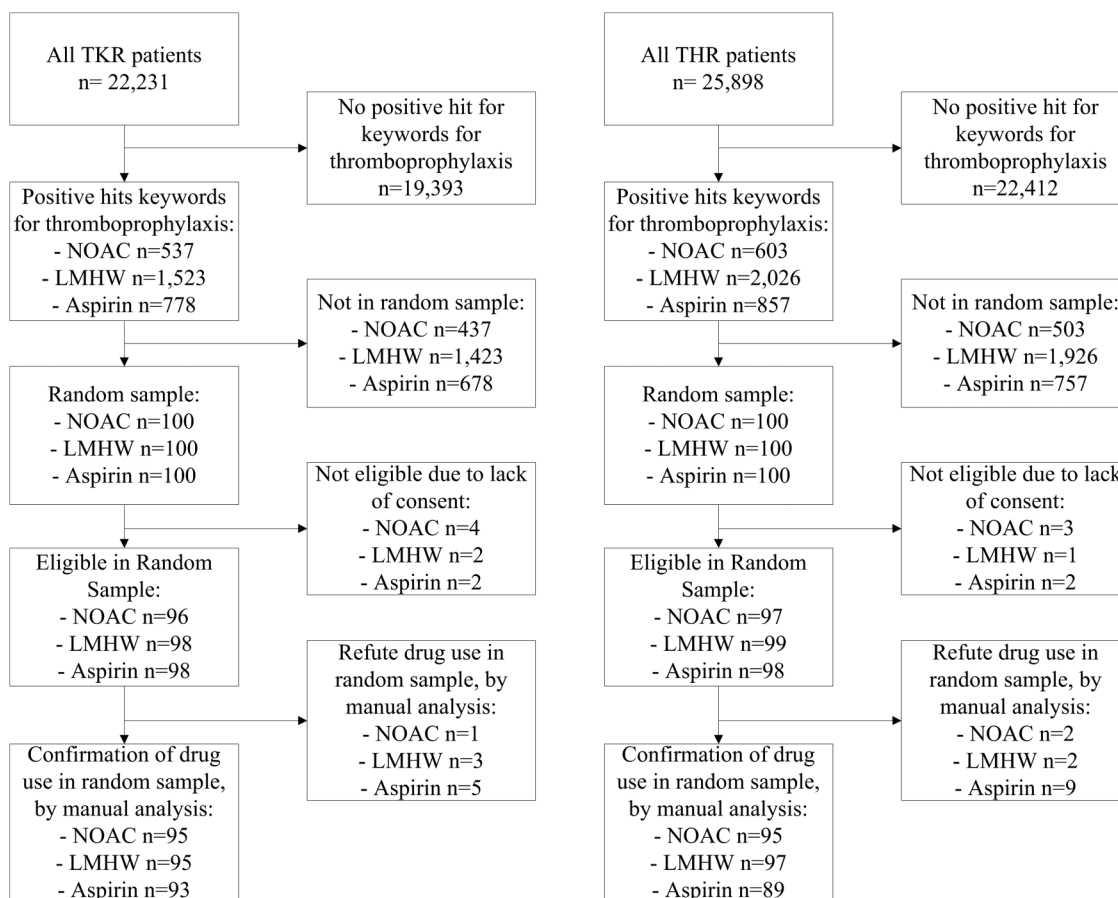
### Selection of comparison groups

We determined the exposure to antithrombotics using CPRD product codes and analysis of anonymised free text. Use of product codes of NOACs (dabigatran and rivaroxaban), LMWHs (bemiparin, certoparin, reviparin, enoxaparin, tinzaparin, dalteparin), and aspirin during the 10 days before and 10 days after surgery were included. Based on the annual reports of the National Joint Registry (NJR), VKAs were used for thromboprophylaxis in <1% of the TJR surgeries. They were, therefore, not included in the analyses. The procedure of the anonymised free-text analysis (figure 1) was based on methods used in a previous study that identified rivaroxaban use and a study that restructured dosage instructions.<sup>10 11</sup> For the anonymised free-text analysis, we first ran our search strategy on the study population to determine the use of NOACs, LMWHs and aspirin. This search strategy searched for specific keywords in the anonymised free-text notes of TJR patients. Keywords included abbreviations, mis-spelled variations, and actual names of both the generic and the brand products (see online supplementary appendix table S1). Anonymised free-text notes between 10 days before and 10 days after date of surgery were included in the analysis. This time window was based on the mean duration of hospital stays after TJRs, as reported in the National Registry of Hospitalisations of England (2008–2013).<sup>15</sup>

Patients with a positive hit for these keywords were further analysed to determine the validity of this positive response. From both THR and TKR patients, we randomly selected 100 patients with at least one positive NOAC hit, 100 patients with at least one positive LMWH hit and 100 patients with at least one positive aspirin hit. This resulted in a total of 300 THR patients and 300 TKR patients. Additionally, anonymised free-text notes of these randomly selected patients were analysed. The additional analysis consisted of anonymising of the free-text surrounding the positive hit in each record. By anonymising 20 words before and after the positive hit, we either confirmed or refuted the drug use. In the following examples drug use was confirmed (1) or refuted (2), respectively:

1. ...still on dabigatran...
2. ...aspirin was stopped in ...

Within the anonymised free text, we manually searched for negating or confirming terms. Two pharmacists (JTHN and FdV) then independently decided



**Figure 1** Study flowchart. TKR, total knee replacement; THR, total hip replacement; NOAC, new oral anticoagulant; LMWH, low-molecular weight heparin.

whether a patient had been exposed or unexposed to antithrombotic drugs. In case of disagreement, the participant was discussed until an agreement was reached.

Subsequently we determined exposure to antithrombotics in the entire study population by combining drug use based on a positive response in the anonymised free-text analysis and drug use based on product codes (see online supplementary appendix table S2). Concomitant use of NOACs and aspirin were categorised as NOAC use. Concomitant use of LMWHs and aspirin were categorised as LMWH use. This classification was chosen because it is more likely that in these cases the exposure to NOACs or LMWHs is associated with TJR. Patients without identified drug use were categorised as unknown users. Exposure to a NOAC and a LMWH is highly unlikely. Therefore, concomitant use of these drugs was also classified as unknown use.

### Data analysis

Positive predictive values (PPVs) of the positive hits by free-text search were calculated by dividing the number of positive hits, confirmed by manual analysis of surrounding free text, by the total number of positive hits. PPVs can range between 0% and 100%, where 100% corresponds with perfect predictive value of positive hits.

We were unable to calculate false or true negatives because we did not analyse the free-text notes without a positive hit. As a result, we could not determine the specificity and sensitivity of our method. In order to assess the additional detection rate by using our free-text search, we determined the distribution of cases within a drug category according to the identification method. Cases were categorised according to type of surgery and the previously described user categories. To gain insight for the external validity of the method, we first compared the proportion of NOAC and LMWH use in the CPRD to the proportion reported by the National Joint Registry (NJR)<sup>13</sup> by means of the  $\chi^2$  statistic for independent samples ( $p < 0.05$ ). Second, we compared the characteristics of the different user groups (NOACs, LMWHs and aspirin) to patients without identified drug use by means of the  $\chi^2$  or t test statistic ( $p < 0.05$ ). We used SAS 9.3 software for statistical testing and randomisation. This study protocol was approved by the Independent Scientific Advisory Committee (ISAC), protocol number: 14\_026R.

### RESULTS

A total of 22 231 TKR patients and 25 898 THR patients were included in our study. From all patients (n=6324)

with a positive hit for antithrombotic drug use, 600 patients were randomly selected for anonymisation of the free-text surrounding a keyword. However, 14 patients did not give consent for this analysis. The remaining 586 patients were further analysed to determine actual drug use (figure 1). On an average, TKR patients were younger than THR patients. They also had a higher BMI and used more drugs, such as statins, non-steroidal anti-inflammatory drugs (NSAIDs), and antihypertensives (table 1).

PPVs of drugs used for thromboprophylaxis related to TJR ranged between 97% and 99% for NOAC and LMWH use, and between 97% and 99% for aspirin use. Overall, 96.2% of the hits were true positives and 3.8% of the hits were false positives. By combining exposure according to positive anonymised free text hits with exposure according to product codes identification of drug use was increased by 57% compared to product codes only. Ultimately, drug use was identified in a total proportion of 18.5% of all THR and 18.6% of all TKR surgeries. Antithrombotic drug use was determined by free-text analysis only in 65–70% of the LMWHs group and >80% of the NOACs group; which threefold to fivefold increased the detection rates of these drugs compared to detection by product codes. In the aspirin group, drug use was mostly (~90%) determined by coded prescription data (figure 2 or see online supplementary appendix table S3). Use of NOACs was higher with our method in CPRD as compared to the NJR reports in both THR and TKR patients when ratio of NOAC and LMWH use was compared. NOAC/LMWH ratio was only statistically significantly different in TKR patients in 2009 and 2010. All other groups were not statistically significantly different (table 2). Given the lower PPV and the substantial difference of baseline characteristics, and the limited role of thromboprophylaxis as related to TJR, we did not further focus on comparing our aspirin data with those from the NJR.

Table 3 shows that TKR and THR patients using NOACs or LMWHs were largely similar with regard to distribution by age, sex, comorbidities and drug use as compared to patients with unknown drug use. Aspirin users, however, were different as compared to unknown users particularly with regard to a history of ischaemic heart disease. Compared to unknown use, prescription of antithrombotic drugs was higher in the highest socioeconomic status (SES) quintile and lower in the lowest SES quintile (except for THR patients using NOACs). Compared to unknown users, all other groups showed a different distribution across the four major regions. Moreover, only one NOAC prescription was identified in Northern Ireland. Furthermore, THR patients using NOACs or LMWHs had a higher BMI as compared to unknown users.

## DISCUSSION

We have developed a useful method to identify exposure to NOACs or LMWHs related to TJR surgery. Our search

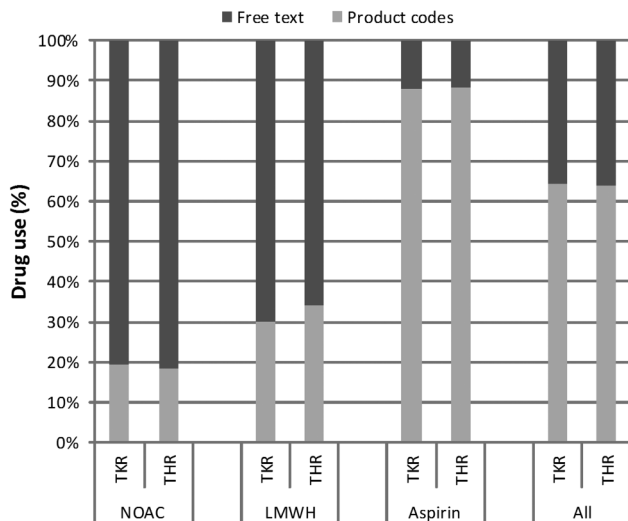
**Table 1** Baseline characteristics of TJR cases at date of surgery from January 2008 to October 2012

	TKR (n=22 231) n (%)	THR (n=25 898) n (%)
Females	12 576 (56.6)	16 036 (61.9)
Age (years, (SD))	68.9 (9.9)	70.1 (12.3)
18–49 years	637 (2.9)	1491 (5.8)
50–69 years	10 665 (48.0)	10 036 (38.8)
≥70 years	10 929 (49.2)	14 371 (55.5)
Socioeconomic status		
Low	3519 (15.8)	4214 (16.3)
Low-medium	3581 (16.1)	4158 (16.1)
Medium	2810 (12.6)	3143 (12.1)
Medium-high	2170 (9.8)	2295 (8.9)
High	1481 (6.7)	1561 (6.0)
Missing	8670 (39.0)	10 527 (40.6)
Region		
England	17 371 (78.1)	19 746 (76.2)
Northern-Ireland	480 (2.2)	795 (3.1)
Scotland	2016 (9.1)	2944 (11.4)
Wales	2364 (10.6)	2413 (9.3)
BMI, most recent before TJR		
BMI (kg/m <sup>2</sup> , mean(SD))	30.0 (5.4)	27.4 (5.2)
History of comorbidity ever before TJR		
Angina pectoris	2068 (9.3)	2214 (8.5)
AMI	871 (3.9)	1190 (4.6)
Haemorrhagic stroke	110 (0.5)	177 (0.7)
Ischaemic stroke	174 (0.8)	324 (1.3)
Heart failure	581 (2.6)	885 (3.4)
Cancer	1352 (6.1)	1838 (7.1)
IBD*	320 (1.4)	315 (1.2)
Varicose veins	3744 (16.8)	3957 (15.3)
GI-ulcer	1068 (4.8)	1174 (4.5)
Atrial fibrillation	985 (4.4)	1441 (5.6)
VTE	1074 (4.8)	1183 (4.6)
History of drug use within 6 months before TJR		
Statins	8471 (38.1)	8293 (32.0)
Systemic glucocorticoids	1227 (5.5)	1413 (5.5)
Any NSAID	8708 (39.2)	9027 (34.9)
COX-2 selective	685 (3.1)	663 (2.6)
NSAIDs		
Non-selective	8148 (36.7)	8506 (32.8)
NSAIDs		
Antibiotics	6764 (30.4)	7587 (29.3)
Antihypertensives	12 948 (58.2)	13 553 (52.3)

\*IBD: Crohn's disease and ulcerative colitis.

AMI, acute myocardial infarction; BMI, body mass index; GI, gastrointestinal; IBD, inflammatory bowel disease; NSAID, non-steroidal anti-inflammatory drug; THR, total hip replacement; TJR, total joint replacement; TKR, total knee replacement; VTE, venous thromboembolic event.

method identified NOAC and LMWH use with PPVs between 97% and 99%. Aspirin use yielded PPVs ranging from 91% to 95%. When combining our anonymised free text method with a traditional method using CPRD product codes, the identification of drug use increased by 57% on average. Moreover, the identification of NOAC use showed a fivefold increase and the



**Figure 2** Drug use according to identification method. TKR, total knee replacement; THR, total hip replacement; NOAC, new oral anticoagulant; LMWH, low-molecular weight heparin.

exposure to LMWHs a more than threefold increase as compared to identification by product codes. Users of NOACs or LMWHs were largely similar to patients with unknown thromboprophylaxis related to TJR surgery with regards to age, sex, comorbidities and use of other drugs. As expected, aspirin users were different and had, for example, more often a history of ischaemic heart disease.

To our knowledge, anonymised free text has only been used once to identify antithrombotic drug use in the CPRD.<sup>10</sup> However, details of the electronic search method used in this study are unclear and results were restricted to rivaroxaban use only. In this paper, we present an effective and highly efficient approach to additionally identify exposure to multiple antithrombotic drug classes, using a predefined set of keywords. Positive hits using the selected keywords in anonymised free-text notes were highly associated with actual use, especially the use of NOAC and LMWH. Thus, we believe that our free-text search method is a valid method for the additional identification of drug use in the CPRD. When

combining our anonymised free text method with a traditional method using CPRD product codes, identification of drug use is increased by 57% compared to the identification based on product codes. This increase of NOAC identification is somewhat higher than the threefold increase of rivaroxaban use presented by Martinez *et al.*<sup>10</sup> This difference could be caused by an overestimation in our method due to false positive responses. However, previously described PPVs indicate this can only account for 1–3% overestimation. It could also be due to the fact that we determined both rivaroxaban and dabigatran use, whereas Martinez and colleagues only measured rivaroxaban use. An alternative explanation is that Martinez *et al.* underestimated the drug utilisation. Our method could be more sensitive due to the wider variety of keywords used in our electronic search.

While we found a method to increase the detection rate of drug use, the question that arises is whether there is a reason why drug use was identified in these specific patients (ie, are these patients different compared to the patients without identified drug use, and could this difference be the reason why drug use was identified in the former group and not in the latter). In order to determine whether we were dealing with a deviating group of patients and to confirm external validity, we first compared use of NOACs and LMWHs with annual NJR reports. We did not include aspirin use in this analysis because of expected differences based on the fact that aspirin is indicated for various other cardiovascular diseases. The ratio of NOAC and LMWH use in our CPRD analysis appeared to be different compared to the NJR reports in TKR patients in 2009 and 2010 only. This could be due to the fact that NOAC prescription is considered to be new and therefore more likely to be mentioned in the free-text notes as compared to LMWH use. To further investigate this, we also compared comorbidities and history of drug use of NOACs, LMWHs or aspirin among patients with unknown drug exposure status.

Only minor differences in characteristics were found between NOAC or LMWH users and patients with an unknown exposure status associated with TJR surgery (table 3). This suggests that these groups are comparable

**Table 2** Difference in NOAC/LMWH ratio found in the CPRD and the NJR

Year	Source	TKR				THR			
		NOAC (%)	LMWH (%)	Ratio	p Value	NOAC (%)	LMWH (%)	Ratio	p Value
2008	CPRD	3.7	96.3	0.04	0.56	3.1	96.9	0.03	0.29
	NJR	2.3	97.7	0.02		1.0	99.0	0.01	
2009	CPRD	22.7	77.4	0.29	0.01	17.5	82.5	0.21	0.10
	NJR	9.5	90.6	0.10		9.5	90.5	0.10	
2010	CPRD	37.4	62.6	0.60	0.02	33.2	66.8	0.50	0.11
	NJR	22.7	77.3	0.29		22.9	77.1	0.30	
2011	CPRD	36.5	63.6	0.57	0.07	30.7	69.3	0.44	0.39
	NJR	24.6	75.4	0.33		25.2	74.8	0.34	

CPRD, Clinical Practice Research Datalink; LMWH, low-molecular weight heparins; NJR, National Joint Registry; NOAC, new oral anticoagulants; THR, total hip replacement; TKR, total knee replacement.

**Table 3** Characteristics of user groups identified by free-text analysis and product codes after total joint replacement

	TKR (n=22 231) n (%)				THR (n=25 898) n (%)			
	NOAC n=450	LMWH n=1164	Aspirin n=2519	Unknown n=18 098	NOAC n=488	LMWH n=1534	Aspirin n=2770	Unknown n=21 106
Females	268 (59.6)	659 (56.6)	1281 (50.9)*	10 368 (57.3)	287 (58.8)	943 (61.5)	1602 (57.8)*	13 204 (62.6)
Age (years, (SD))	68.5 (9.0)	69.0 (9.5)	71.8 (8.7)*	68.5 (10.0)	68.2 (10.7)*	69.5 (11.4)	73.7 (10.9)*	69.7 (12.5)
Socioeconomic status								
Low	75 (16.7)*	169 (14.5)*	305 (12.1)*	2970 (16.4)	78 (16.0)	244 (15.9)*	344 (12.4)*	3548 (16.8)
Low-medium	81 (18.0)*	191 (16.4)*	381 (15.1)*	2928 (16.2)	81 (16.6)	263 (17.1)*	377 (13.6)*	3437 (16.3)
Medium	64 (14.2)*	179 (15.4)*	280 (11.1)*	2287 (12.6)	54 (11.1)	226 (14.7)*	316 (11.4)*	2547 (12.1)
Medium-high	29 (6.4)*	106 (9.1)*	232 (9.2)*	1803 (10.0)	42 (8.6)	150 (9.8)*	231 (8.3)*	1872 (8.9)
High	39 (8.7)*	114 (9.8)*	171 (6.8)*	1157 (6.4)	31 (6.4)	142 (9.3)*	160 (5.8)*	1228 (5.8)
Missing	162 (36.0)*	405 (34.8)*	1150 (45.7)*	6953 (38.4)	202 (41.4)	509 (33.2)*	1342 (48.4)*	8474 (40.1)
Region								
England	388 (86.2)*	963 (82.7)*	1735 (68.9)*	14 285 (78.9)	386 (79.1)*	1267 (82.6)*	1795 (64.8)*	16 298 (77.2)
Northern-Ireland	0 (0.0)*	43 (3.7)*	145 (5.8)*	292 (1.6)	1 (0.2)*	56 (3.7)*	242 (8.7)*	496 (2.4)
Scotland	42 (9.3)*	83 (7.1)*	313 (12.9)*	1565 (8.6)	60 (12.3)*	137 (8.9)*	405 (14.6)*	2342 (11.1)
Wales	20 (4.4)*	75 (6.4)*	326 (12.4)*	1956 (10.8)	41 (8.4)*	74 (4.8)*	328 (11.8)*	1970 (9.3)
BMI, most recent before TJR								
BMI (kg/m <sup>2</sup> , mean(SD))	30.0 (5.5)	30.2 (5.5)	29.9 (5.2)	30.0 (5.4)	28.4 (5.4)*	27.8 (5.3)*	27.6 (5.3)*	27.4 (5.2)
History of comorbidity ever before TJR								
Angina pectoris	33 (7.3)	110 (9.5)*	611 (24.3)*	1314 (7.3)	34 (7.0)	112 (7.3)	628 (22.7)*	1440 (6.8)
AMI	9 (2.0)	62 (5.3)*	269 (10.7)*	531 (2.9)	20 (4.1)	65 (4.2)	358 (12.9)*	747 (3.5)
Haemorrhagic stroke	6 (1.3)*	7 (0.6)	10 (0.4)	87 (0.5)	2 (0.4)	13 (0.8)	15 (0.5)	147 (0.7)
Ischaemic stroke	2 (0.4)*	6 (0.5)	56 (2.2)*	110 (0.6)	0 (0.0)*	21 (1.4)	80 (2.9)*	223 (1.1)
Heart failure	8 (1.8)	36 (3.1)	110 (4.4)*	427 (2.4)	10 (2.0)	53 (3.5)	167 (6.0)*	655 (3.1)
Cancer	29 (6.4)	77 (6.6)	176 (7.0)*	1070 (5.9)	32 (6.6)	111 (7.2)	204 (7.4)	1491 (7.1)
IBD†	8 (1.8)	22 (1.9)	44 (1.7)	246 (1.4)	11 (2.3)*	17 (1.1)	38 (1.4)	249 (1.2)
Varicose veins	75 (16.7)	215 (18.5)	449 (17.8)	3005 (16.6)	79 (16.2)	252 (16.4)	440 (15.9)	3186 (15.1)
GI-ulcer	23 (5.1)	75 (6.4)*	136 (5.4)	834 (4.6)	20 (4.1)	70 (4.6)	149 (5.4)*	935 (4.4)
Atrial fibrillation	9 (2.0)*	84 (7.2)*	137 (5.4)*	755 (4.2)	20 (4.1)	116 (7.6)*	237 (8.6)*	1068 (5.1)
VTE	26 (5.8)	96 (8.2)*	113 (4.5)	839 (4.6)	12 (2.5)*	119 (7.8)*	122 (4.4)	930 (4.4)
History of drug use within 6 months before TJR								
Statins	172 (38.2)	482 (41.4)*	1644 (65.3)*	6173 (34.1)	168 (34.4)*	515 (33.6)*	1639 (59.2)*	5971 (28.3)
Systemic glucocorticoids	25 (5.6)	76 (6.5)	124 (4.9)	1002 (5.5)	26 (5.3)	86 (5.6)	150 (5.4)	1151 (5.5)
Any NSAID	171 (38.0)	450 (38.7)	867 (34.4)*	7220 (39.9)	190 (38.9)	569 (37.1)	837 (30.2)*	7431 (35.2)
COX-2 selective	20 (4.4)	41 (3.5)	62 (2.5)	562 (3.1)	11 (2.3)	42 (2.7)	65 (2.3)	545 (2.6)
NSAIDs								
Non-selective NSAIDs	153 (34.0)	419 (36.0)	818 (32.5)*	6758 (37.3)	182 (37.3)	543 (35.4)	784 (28.3)*	6997 (33.2)
Antibiotics	145 (32.2)	394 (33.8)*	770 (30.6)	5455 (30.1)	133 (27.3)	481 (31.4)*	968 (34.9)*	6005 (28.5)
Antihypertensives	261 (58.0)	704 (60.5)*	1972 (78.3)*	10 011 (55.3)	259 (53.1)	803 (52.3)*	2036 (73.5)*	10 455 (49.5)

Percentages may not add up to 100% due to rounding.

\*Statistically significant difference as compared to patients with unknown exposure with regard to chemical thromboprophylaxis ( $p < 0.05$ ). THR patients using NOACs, LMWHs, or aspirin were compared to THR patients with unknown exposure. TKR patients using NOACs, LMWHs, or aspirin were compared to TKR patients with unknown exposure.

†IBD: Crohn's disease and ulcerative colitis.

AMI, acute myocardial infarction; BMI, body mass index; GI, gastrointestinal; IBD, inflammatory bowel disease; NSAID, non-steroidal anti-inflammatory drug; THR, total hip replacement; TJR, total joint replacement; TKR, total knee replacement; VTE, venous thromboembolic event.

and that there is no apparent reason for a detection bias of NOAC or LMWH use in these specific patients. This is reassuring when long-term effects and side effects of the different drugs are compared in further research. As expected, aspirin users were very different compared to unknown users with regard to comorbidities and drug use (table 3). This is most likely due to the fact that aspirin is indicated for multiple other conditions such as angina pectoris, the prevention of myocardial infarction and other types of ischaemic heart disease. Therefore, our method seems to have limited usability for the identification of aspirin use in chemical thromboprophylaxis related to TJR. Differences in drug use according to region may be caused by use of regional guidelines. NICE guidance is predominantly designed for use in England and Wales. The Health and Social Care (HSC) and the Scottish Intercollegiate Guidelines Network (SIGN) are responsible for guidance in Northern Ireland (NI) and Scotland, respectively. However, linkage with NICE guidelines for implementation in NI has been available since 2006.

Our study had various strengths. To the best of our knowledge, this is the first study reporting an effective method to identify antithrombotic drug use from anonymised free text in a large population-based GP database in a peer-reviewed scientific journal. Moreover, our method can be implemented at relatively low costs, since actual anonymisation of free text is not required. This will then allow us to use the wealth of data in the world's largest primary care database to study, for example, the potential side effects of thromboprophylaxis associated with TJR. Our study also had limitations. We were practically unable to calculate false or true negatives. As a result, we were unable to determine the specificity and sensitivity of our method. In order to evaluate whether documentation of the exposure was differential between patients with a positive hit and patients without a hit, we applied two methods. First, we assessed distribution of NOAC and LMWH use from an external source, the National Joint Registration (NJR). Second, we assessed differences in baseline characteristics of the various exposure groups to the group without a hit. With these surrogate measurements we have generated information concerning the potential differential detection of drug use exposure. Another limitation was that we were able to identify thromboprophylaxis related to TJR for only 18% of our patients, whereas actual thromboprophylaxis is likely to be close to 100%. This is probably the result of under-reporting of thromboprophylaxis by either hospitals, the GP or simply because discharge letters were sent to GPs as scanned PDF files rather than searchable free text. Nevertheless, we were still able to substantially increase identification using anonymised free text compared to identification based on product codes only. In the future, this may be further enhanced by linking to other data sources such as the in-hospital prescribing data, or by making the existing linkage between the NJR and CPRD available to researchers without any restrictions.

In conclusion, we have developed a useful method to identify exposure to NOACs or LMWHs associated with TJR surgery. We can conclude that positive hits using the selected keywords in anonymised free-text notes are strongly associated with actual use and that by using this method, identification of drug use has increased five-fold. Furthermore, identified users of NOACs and LMWH users appear to be reasonably similar to patients without identified drug use (table 3). In contrast, aspirin users were very different as compared to patients without identified drug use, possibly due to the fact that aspirin is prescribed for various other health problems. Identification of drugs used for a specific postsurgery health ailment can be substantially enhanced by using our method; similar methods may be used for the identification of drugs prescribed for other diseases or hospital-specific medications such as biologicals and blood products. Our method is a useful tool to identify exposure to NOACs or LMWHs related to total TKR or THR surgery in the CPRD, and increases statistical power to evaluate potential side effects of these drugs in pharmacoepidemiological studies.

#### Author affiliations

<sup>1</sup>Division of Pharmacoepidemiology & Clinical Pharmacology, Utrecht University, Utrecht, The Netherlands

<sup>2</sup>Department of Epidemiology, CAPHRI School for Public Health and Primary Care, Maastricht University, Maastricht, The Netherlands

<sup>3</sup>Department of Pharmacy, Sint Maartensskliniek, Nijmegen, The Netherlands

<sup>4</sup>Department of Pharmacy, Radboud University Medical Center, Nijmegen, The Netherlands

<sup>5</sup>Department of Internal Medicine, Division of Rheumatology, Maastricht University Medical Centre, Maastricht, The Netherlands

<sup>6</sup>Department of Epidemiology, CAPHRI School for Public Health and Primary Care, and CARIM School for Cardiovascular Diseases, Maastricht University, Maastricht, The Netherlands

<sup>7</sup>Department of Orthopaedics, Maastricht University Medical Center, Maastricht, The Netherlands

<sup>8</sup>Department of Clinical Pharmacy and Toxicology, Maastricht University Medical Center, Maastricht, The Netherlands

<sup>9</sup>Department of Clinical Pharmacy, University Medical Center Utrecht, Utrecht, The Netherlands

**Contributors** JTHN, FdV, PCD, PJE, NV, BJFvdB, T-PvS and AB contributed to the conception and design of the study. T-PvS, AL, FdV and JTHN contributed to the acquisition of data. JTHN and AL contributed to analysis of data. All authors contributed to interpretation of data. JTHN and FdV drafted the manuscript, and all authors contributed intellectual content while revising the article. All authors gave final approval of the version to be submitted and any revised version.

**Funding** This study was internally funded by the Maastricht University Medical Center+.

**Competing interests** JTHN, AL, T-PvS and FdV are employed by the Division of Pharmacoepidemiology and Clinical Pharmacology at Utrecht Institute for Pharmaceutical Sciences, which has received unrestricted funding from the Netherlands Organisation for Health Research and Development (ZonMW), the Dutch Health Care Insurance Board (CVZ), the Royal Dutch Pharmacists Association (KNMP), the private-public funded Top Institute Pharma (<http://www.tipharma.nl>) (including co-funding from universities, government, and industry), the EU Innovative Medicines Initiative (IMI), the EU 7th Framework Program (FP7), and the Dutch Ministry of Health and Industry (including GlaxoSmithKline, Pfizer, and others). AL reports grants from Netherlands Organisation for Scientific Research (NWO), outside the submitted work. BJFvdB reports research grants to his department from Pfizer and Roche and

occasionally speaker's honoraria from Pfizer, Roche, Abbvie and MSD. AB reports research grants to her department from Amgen Abbvie, Pfizer and Merck, and occasionally speaker's honoraria from Pfizer, UCB and Sandoz. PJE reports research grants to his department from Stryker, Active implants, Carbylan Biosurgery, DSM Biomedical and Regentis, and occasionally speaker's honoraria from Biomet and Push braces. PCD has received unrestricted grants from NWO, EU and nutritional industry for research unrelated to this topic.

**Ethics approval** Independent Scientific Advisory Committee MHRA Database Research.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data sharing statement** No additional data are available.

**Open Access** This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

## REFERENCES

1. Arthritis Research UK. Osteoarthritis in general practice. 2013:1–36. <http://www.arthritisresearchuk.org/arthritis-information/data-and-statistics/osteoarthritis.aspx> (accessed 9 Feb 2015).
2. Learmonth ID, Young C, Rorabeck C. The operation of the century: total hip replacement. *Lancet* 2007;370:1508–19.
3. Lalmohamed A, Vestergaard P, Jansen PA, *et al*. Prolonged outpatient vitamin K antagonist use and risk of venous thromboembolism in patients undergoing total hip or knee replacement. *J Thromb Haemost* 2013;11:642–50.
4. Falck-Ytter Y, Francis CW, Johanson NA, *et al*. Prevention of VTE in orthopedic surgery patients: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest* 2012;141(2 Suppl):e278S–325S.
5. Zufferey P, Laporte S, Quenet S, *et al*. Optimal low-molecular-weight heparin regimen in major orthopedic surgery: a meta-analysis of randomised trials. *Thromb Haemost* 2003;90:654–61.
6. Lip GY, Tse HF, Lane DA. Atrial fibrillation. *Lancet* 2012;379:648–61.
7. Di Nisio M, Middeldorp S, Bueller HR. Direct thrombin inhibitors. *N Engl J Med* 2005;353:1028–40.
8. Gómez-Outes A, Terleira-Fernández AI, Suárez-Gea L, *et al*. Dabigatran, rivaroxaban, or apixaban versus enoxaparin for thromboprophylaxis after total hip or knee replacement: systematic review, meta-analysis, and indirect treatment comparisons. *BMJ* 2012;344:e3675.
9. Banerjee A, Lane DA, Torp-Pedersen D, *et al*. Net clinical benefit of new oral anticoagulants (dabigatran, rivaroxaban, apixaban) versus no treatment in a 'real world' atrial fibrillation population: a modelling analysis based on a nationwide cohort study. *Thromb Haemost* 2012;107:584–9.
10. Martinez C, Rietbrock S, Suzart-Woischnik K, *et al*. Drug use pattern of rivaroxaban in the UK. *Int J Cardiol* 2015;180:246–54.
11. Shah AD, Martinez C. An algorithm to derive a numerical daily dose from unstructured text dosage instructions. *Pharmacoepidemiol Drug Saf* 2006;15:161–6.
12. Williams T, van Staa T, Puri S, *et al*. Recent advances in the utility and use of the General Practice Research Database as an example of a UK Primary Care Data resource. *Ther Adv Drug Saf* 2012;3:89–99.
13. Lalmohamed A, MacGregor AJ, de Vries F, *et al*. Patterns of risk of cancer in patients with metal-on-metal hip replacements versus other bearing surface types: a record linkage study between a prospective joint registry and general practice electronic health records in England. *PLoS ONE* 2013;8:e65891.
14. Wallace G, Judge A, Prieto-Alhambra D, *et al*. The effect of body mass index on the risk of post-operative complications during the 6 months following total hip replacement or total knee replacement surgery. *Osteoarthritis Cartil* 2014;22:918–27.
15. Hospital Episode Statistics. Admitted Patient Care—England, 2008–2013. <http://www.hscic.gov.uk/> (accessed 9 Nov 2014).