Drug Safety
https://doi.org/10.1007/s40264-021-01044-4

ORIGINAL RESEARCH ARTICLE

# Identification and Validation of Major Cardiovascular Events in the United Kingdom Data Sources Included in a Multi-database Post-authorization Safety Study of Prucalopride

**Ana Ruigómez[1] · Estel Plana[2] · Alicia Gilsenan[3] · Joan Fortuny[2] · Miguel Cainzos-Achirica[4,5] · Robert W. V. Flynn[6] · Thomas M. MacDonald[6] · Luis Garcia-Rodriguez[1] · Ryan Ziemiecki[3] · Elizabeth B. Andrews[3]**

Accepted: 11 January 2021


## Abstract

**Introduction** A multinational post-authorization safety study assessed cardiovascular safety in initiators of prucalopride for chronic constipation compared with a matched cohort of polyethylene glycol 3350 initiators. The primary safety outcome was major adverse cardiovascular events (MACE), a composite of hospitalization for acute myocardial infarction, stroke, or in-hospital cardiovascular death. We report the validation process for MACE endpoints in United Kingdom (UK) data sources: Clinical Practice Research Datalink (CPRD GOLD), The Health Improvement Network (THIN), and the Information Services Division (ISD) Scotland.

**Methods** Modified electronic algorithms from prior research identified potential MACE cases. Validation followed a common protocol, adapted for each database, with all information anonymized: (1) direct confirmation via linkage to hospital records (CPRD GOLD); (2) requests for additional clinical information through questionnaires (CPRD GOLD), free-text (THIN), or abstraction of hospital records (ISD); (3) manual review of electronic records of potential events retrieved by the algorithm (CPRD GOLD/THIN); and (4) event adjudication by three clinicians, blinded to exposure, for all remaining events.

**Results** Electronic algorithms identified 260 potential MACE cases: 38 confirmed via linkage to hospital records (CPRD GOLD), 56 ruled out as non-cardiovascular death cases (THIN), and three unavailable for review (ISD), leaving 163 potential cases. After manual review with additional information (steps 2 and 3), 45 were considered noncases (CPRD GOLD/THIN). Upon final adjudication (step 4), remaining potential events were adjudicated as definite ($n = 62$), probable ($n = 10$), possible ($n = 13$), or noncases ($n = 33$).

**Conclusions** Given the limitations of relying solely on computer algorithms to identify cardiovascular outcomes, validation with clinical review is essential to guide interpretation.

## 1 Introduction

Prucalopride (Resolor) is a 5-hydroxytryptamine receptor type 4 (5-HT4) agonist approved for the treatment of chronic constipation [1]. On regulatory request, a post-authorization safety study (PASS) was conducted in 2016–2017 to assess the cardiovascular safety of prucalopride by comparing the occurrence of cardiovascular events among initiators of this drug and among a matched comparator cohort of polyethylene glycol 3350 (PEG) initiators [2, 3].

The study was conducted in five European data sources: the Clinical Practice Research Datalink (CPRD GOLD), the Health Improvement Network (THIN), and the Information Services Division (ISD) Scotland in the United Kingdom (UK); the German Pharmacoepidemiological Research Database (GePaRD); and the Swedish National Registries

✉ Alicia Gilsenan
agilsenan@rti.org

1 Centro Español de Investigación Farmacoepidemiológica (CEIFE), Madrid, Spain

2 RTI Health Solutions, Barcelona, Spain

3 RTI Health Solutions, 3040 Cornwallis Rd, Research Triangle Park, NC 27709, USA

4 Division of Cardiovascular Prevention and Wellness, Department of Cardiology, Houston Methodist DeBakey Heart and Vascular Center, Houston, TX, USA

5 Center for Outcomes Research, Houston Methodist, Houston, TX, USA

6 University of Dundee, Dundee, Scotland

**Key Points**

In this population-based, multinational post-authorization safety study to evaluate cardiovascular safety in initiators of prucalopride versus a matched cohort of polyethylene glycol 3350 (PEG) initiators, major adverse cardiovascular events (MACE) endpoints were identified using electronic algorithms.

Validation of MACE endpoints followed a common protocol, adapted for each of three United Kingdom (UK) data sources: Clinical Practice Research Datalink (CPRD GOLD), The Health Improvement Network (THIN), and Information Services Division (ISD) Scotland.

Validation occurred through direct confirmation via linkage to hospital records, requests for additional clinical information, manual review of potential events retrieved by electronic algorithms, and adjudication of potential events by clinicians who were blinded to exposure.

Of 260 potential MACE events identified by the electronic algorithms (108 from CPRD GOLD, 79 from THIN, and 73 from ISD), 100 were considered actual events after identification of confirmed events, review, and the adjudication process. Given the limitations of electronic algorithms to identify cardiovascular outcomes, validation with clinical review is essential.

(SNR) [4]. The primary endpoint of the PASS was occurrence of a MACE, defined as first occurrence of any of the following endpoints: hospitalization for nonfatal acute myocardial infarction (AMI), hospitalization for nonfatal stroke, or in-hospital cardiovascular death.

Potential clinical endpoints during the study period were identified using electronic algorithms in the five databases. Although the validity of MACE endpoint definitions has been established in databases from both Europe and the United States [5–8], validation of study endpoints by clinical experts was planned a priori as part of the prucalopride PASS in the three databases for which this was expected to be feasible during the study period: CPRD GOLD, THIN, and ISD Scotland.

Here, we describe the methodology and results of the event identification and validation for MACE in the three UK data sources conducted as part of the prucalopride multidatabase PASS. The PASS aimed to provide standardized estimates of incidence rates and incidence rate ratios of MACE among initiators of prucalopride compared with PEG initiators, and methods and results for the PASS are reported elsewhere [2, 3].

## 2 Methods

### 2.1 PASS Setting, Data Sources, and Study Design

As noted previously, details on the research methods and study results of the prucalopride PASS have been reported in detail elsewhere [2, 3]. Briefly, this was a population-based, observational, retrospective cohort study of adult patients with chronic constipation initiating either prucalopride or PEG (matched on a 1:5 ratio), from the UK, Sweden, and Germany. Cohorts of new users were followed from date of first prescription of study drug until first MACE (i.e., hospitalization due to AMI or stroke or in-hospital cardiovascular death), death due to other reasons, end of data collection, or end of study period. The study compared the risk of MACE in prucalopride users versus PEG users, pooling the information from five large health care databases: CPRD GOLD, THIN, ISD Scotland, GePaRD, and SNR.

Event validation for the primary study endpoint was performed only in the UK data sources. Validation via medical record review was not performed in the SNR or GePaRD for this study. However, clinical data from the SNR and GePaRD have been previously validated and determined to be of good quality [9–12]. The characteristics of these three databases (CPRD GOLD, THIN, and ISD Scotland) have been reported previously [3]. Briefly, CPRD GOLD is a database derived from electronic medical records from general practices and includes linkage to Hospital Episode Statistics (HES) for inpatient data and to Office of National Statistics (ONS) data for cause-of-death data for some participating practices. THIN is a primary care database with no link to ONS or HES, but with access to anonymized free-text comments entered by treating general practitioners (GPs) at the time of this study. ISD Scotland is a data source derived from the linkage of routinely collated dispensing, hospitalization, and death certification data.

### 2.2 PASS Endpoint Definitions

The primary endpoint of the prucalopride PASS was first occurrence of a MACE within the study period (2010–2016), defined as the composite of hospitalization for nonfatal AMI [13, 14], hospitalization for nonfatal stroke (either ischemic or hemorrhagic) [15], and in-hospital cardiovascular death. For the latter, a very broad definition was used, aimed at ensuring that any potential in-hospital cardiovascular death would be detected as a potential study endpoint. This definition included deaths from AMI, stroke, heart failure, cardiovascular hemorrhage, sudden cardiac death, cardiovascular procedures (coronary revascularizations), and other cardiovascular deaths [14]. Individual components of MACE were also evaluated separately as secondary study endpoints.

## 2.3 Electronic Ascertainment of Potential Events

Potential study events were identified in all databases using modified versions of previously reported automated algorithms [16–20], which included diagnosis and procedure codes (International Classification of Diseases [ICD] codes and/or codes from local dictionaries). Supplementary Table S1 summarizes the codes used to define each study endpoint (see the electronic supplementary material). Compared with the published definitions, the algorithms underwent minor modifications after clinical review by study investigators to harmonize them across sources and to account for limitations in data availability.

The diagnosis codes were combined with additional qualifiers (e.g., hospital admission components) in operational definitions that also provided a preliminary, automated classification of the study endpoints as "confirmed," "definite," "probable," or "possible" cases, as well as "unknown cause of death" specifically for fatal events (Table 1).

In each data source, hospitalizations for nonfatal AMI and nonfatal stroke were identified by hospital discharge diagnosis codes in patients discharged alive from the hospital. In-hospital cardiovascular deaths were identified through primary cardiovascular discharge diagnoses or the underlying cause-of-death codes when these data were available (CPRD GOLD and THIN). Patients admitted to the hospital for a noncardiovascular cause, regardless of whether it resulted in a cardiovascular cause of death, were initially considered potential cases, and further information was collected, if available. However, if there was no clear cardiovascular cause of hospitalization, the event was not considered a relevant outcome for this study. In THIN, because linkage to HES/ONS was not possible, any (i.e., all-cause) in-hospital deaths were considered for review at this stage, and a preliminary additional manual plausibility evaluation of these death events was conducted before considering them potential cardiovascular death study endpoints.

## 2.4 Case Validation

After all potential study endpoints (including all occurrences during the study follow-up period, not just first occurrence) from the three UK data sources were identified electronically (and provided a preliminary classification), duplicate cases between THIN and CPRD GOLD were removed from the THIN data source (as described in Fortuny et al. [3]). The approach taken to ensure removal of duplicate patients involved (1) eliminating all the Scottish practices from the CPRD and THIN since they were also included in the ISD Scotland database and (2) excluding potentially duplicated practices in THIN and CPRD. Using a modified six-step algorithm based on patient-level demographic and pharmacy data [21], we identified patients thought to be the same individual in the CPRD and THIN without compromising patient or practice confidentiality. In case of duplication, CPRD practices were retained and their equivalents in THIN were removed.

Validation was conducted for all potential study endpoints. Validation occurred per the following four general steps: (1) identification of electronically confirmed cases and direct confirmation via linkage to hospital records (CPRD GOLD only); (2) request for additional clinical information through questionnaires (CPRD GOLD), free-text (THIN), or original hospital case records (ISD); (3) manual review by study investigators of the preliminary patient profiles (e.g., database listings of outpatient visits, procedures, and medications) of events retrieved by the algorithm (CPRD GOLD/THIN) to rule out noncases; and (4) event adjudication by the prucalopride PASS adjudication committee (three clinicians, all blinded to exposure). This common validation protocol had to be adapted to the resources and type of data available in each data source.

### 2.4.1 Case Validation in the CPRD GOLD

In CPRD GOLD (Fig. 1), potential cases were considered confirmed (with no additional validation needed) through hospital discharge diagnoses from HES data if any of the following occurred: (1) for nonfatal events, the primary hospital discharge diagnosis codes were consistent with the list of codes from the study protocol (Supplementary Table S1, see the electronic supplementary material); and (2) for in-hospital deaths, the death occurred in the hospital (identified in HES) and the underlying cause-of-death codes from ONS were consistent with the list of codes from the study protocol (see Supplementary Table S1).

For nonconfirmed potential cases, questionnaires were sent to the treating GPs. To complete the questionnaires, GPs were asked to review the medical records in their possession, including hospital discharge reports, autopsy reports, death certificates (if available to the GP), and free-text, and to complete a specific questionnaire designed for each endpoint. Review of free-text comments was also originally planned, but eventually these became unavailable during the study conduct due to data privacy reasons implemented throughout CPRD GOLD.

Using the information included in automated patient profiles and in available GP questionnaires, a preliminary adjudication/classification was conducted by trained epidemiologists so that obvious noncases could be excluded from the subsequent adjudication review. The investigators were blinded to exposure status, and the classifications and criteria used were the same as the ones used by the electronic algorithms for their preliminary classification, plus "noncase" (i.e., potential cases not meeting any of the prior definitions). Potential cases classified as noncases by one

**Table 1** Operational definitions used in the electronic algorithms for case ascertainment and used as guidelines by clinical reviewers for event adjudication purposes

| Hospitalization for nonfatal AMI | Definite | Probable | Possible definition 1 | Possible definition 2 |
|---|---|---|---|---|
| (A) Main codes (see Supplementary Table S1 in the ESM) | | | | |
| Code for AMI | Yes | Yes | Yes | No |
| Hospitalization within 30 days of AMI | Yes | Yes | Yes | Yes |
| Death within 30 days of hospitalization | No | No | No | No |
| (B) Additional clinical information (diagnosis and procedural codes + other sources of information) | | | | |
| Chest pain<br>Abnormal cardiac enzymes<br>ECG with AMI signs<br>Imaging test, abnormal<br>Thrombolytic therapy<br>Coronary revascularization | ≥ 2 items recorded 30 days before or after the index AMI code | 1 item recorded 30 days before or after the index AMI code | None recorded | Yes, coronary chest pain<br>≥ 1 item recorded 30 days before or after the index coronary chest pain code |

| Hospitalization for nonfatal stroke | Definite | Probable | Possible |
|---|---|---|---|
| (A) Main codes (see Supplementary Table S1 in the ESM) | | | |
| Code for stroke | Yes | Yes | Yes |
| Hospitalization within 30 days of stroke | Yes | Yes | Yes |
| Death within 30 days of hospitalization | No | No | No |
| (B) Additional clinical information (diagnosis and procedural codes + other sources of information) | | | |
| Referral to neurologist<br>Acute treatment for stroke<br>Residual damage from stroke<br>Brain imaging test, abnormal<br>Neurological physiotherapy | ≥ 2 items recorded 30 days before or after the index stroke code | 1 item recorded 30 days before or after the index stroke code | None recorded |

| In-hospital cardiovascular death | Definite | Possible | Unknown |
|---|---|---|---|
| (A) Main codes (see Supplementary Table S1 in the ESM) | | | |
| Code for death with date | Yes | Yes | Yes |
| Evidence of hospitalization within 30 days of death code | Yes | Yes | Yes |
| (B) Additional clinical information (diagnosis and procedural codes + other sources of information) | | | |
| Cardiovascular cause of death specifically reported | Yes | No | No |
| Evidence of AMI, stroke, sudden cardiac death, heart failure, cardiovascular hemorrhage, revascularization procedure, or other cardiovascular event within 30 days of death code | Irrelevant | Yes | No |
| Evidence of a non-cardiovascular condition potentially leading to death within 30 days of death code | Irrelevant | No | No |

The electronic algorithms provided a preliminary classification of all potential cases using diagnosis and procedural codes (A and B). Clinical reviewers conducting case adjudication used this information (A and B) together with additional sources of information (B), such as reports of electrocardiographic findings or of circulating troponin levels described in general practitioner questionnaires or free-text comments, among other sources. Possible cases of hospitalization for nonfatal AMI, hospitalization for nonfatal stroke, or in-hospital cardiovascular death not meeting any of the clinical definitions presented were classified as "noncases"

*AMI* acute myocardial infarction, *ECG* electrocardiogram, *ESM* electronic supplementary material

epidemiologist were reviewed by a second one and, when in agreement, were not considered as a case in the study. The remaining potential cases were reviewed by the study adjudication committee.

#### 2.4.2 Case Validation in THIN

In THIN (Fig. 2), linkage to HES or ONS was not feasible for this study because of the limited number of linked practices and the long lag of data updates. Therefore, there were no electronically confirmed cases in this data source. As

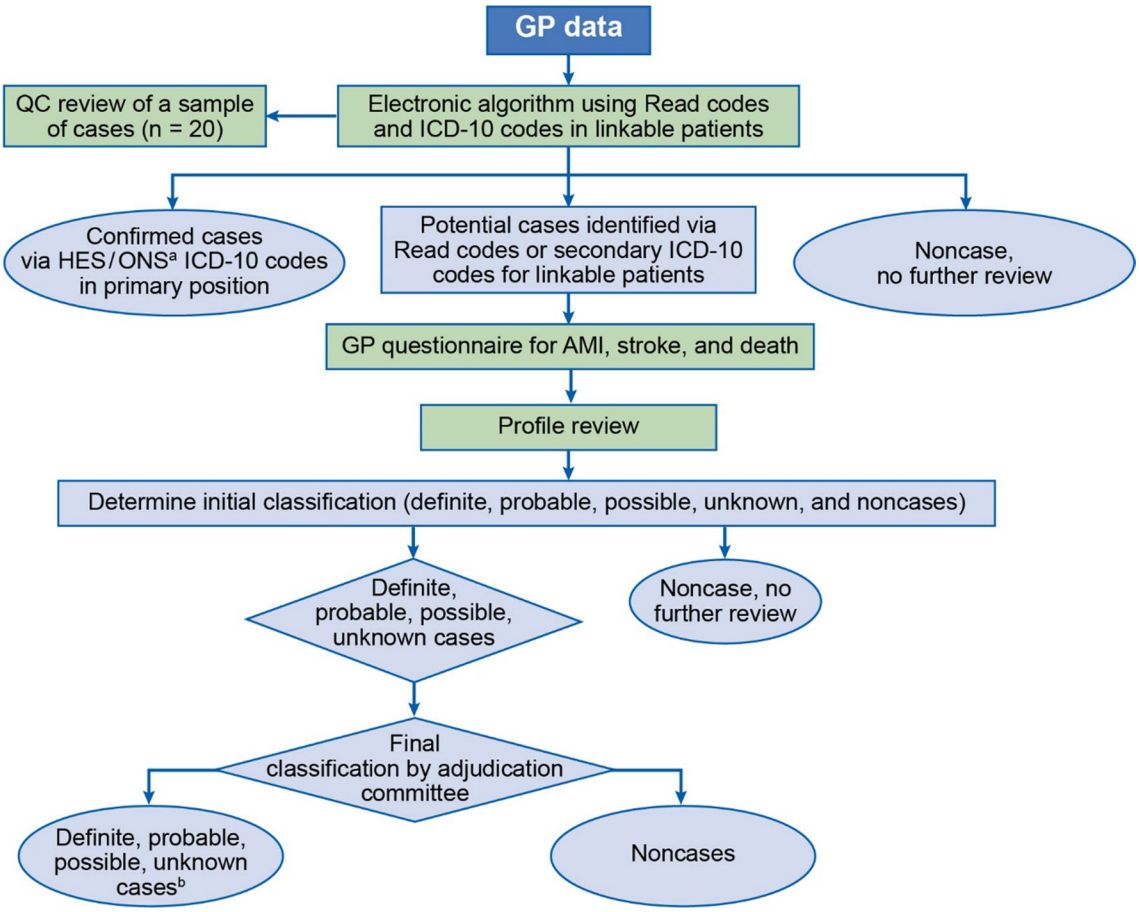


**Fig. 1** Validation process in CPRD GOLD. *AMI* acute myocardial infarction, *CPRD* Clinical Practice Research Datalink, *GP* general practitioner, *HES* Hospital Episode Statistics, *ICD-10* International Statistical Classification of Diseases and Related Health Problems, 10th Revision, *ONS* Office of National Statistics, *QC* quality control. [a]Approximately 50% of patients are linkable to HES and ONS. [b]Definite cases were reassigned as "confirmed" and unknown cases were reassigned as "noncases"; the unknown case classification applied only to cardiovascular deaths

discussed previously, a preliminary plausibility evaluation (manual review of individual patient profiles) of all death events identified was conducted in THIN; death events for which there were conclusive records of an alternative, noncardiovascular etiology were removed from further validation steps and considered "noncases."

GP questionnaires were not used in THIN. Instead, for all potential cases, any free-text comments (i.e., notes made by physicians regarding the patient encounter, for up to 6 months before and after the qualifying diagnosis code or information on the cause of death) were requested and manually reviewed. As in CPRD GOLD, blinded patient profile review by trained epidemiologists was conducted, allowing for the exclusion of additional obvious noncases. The remaining potential cases of MACE were sent to adjudication.

#### 2.4.3 Case Validation in ISD Scotland

In ISD Scotland (Fig. 3), all events were automatically considered potential, and no cases were initially considered confirmed. Medical record abstraction was conducted for all potential cases using chart abstraction forms completed by a team comprising a research nurse and a study doctor. All potential cases underwent review by the study adjudication committee, except for three for which clinical data could not be obtained.

### 2.5 Event Adjudication and Final Case Classifications

An adjudication committee of three clinical experts from the research partner institutions involved in validation blinded to the study exposure reviewed all nonconfirmed, nonexcluded potential cases from the three data sources and determined their final status using all clinical information available.

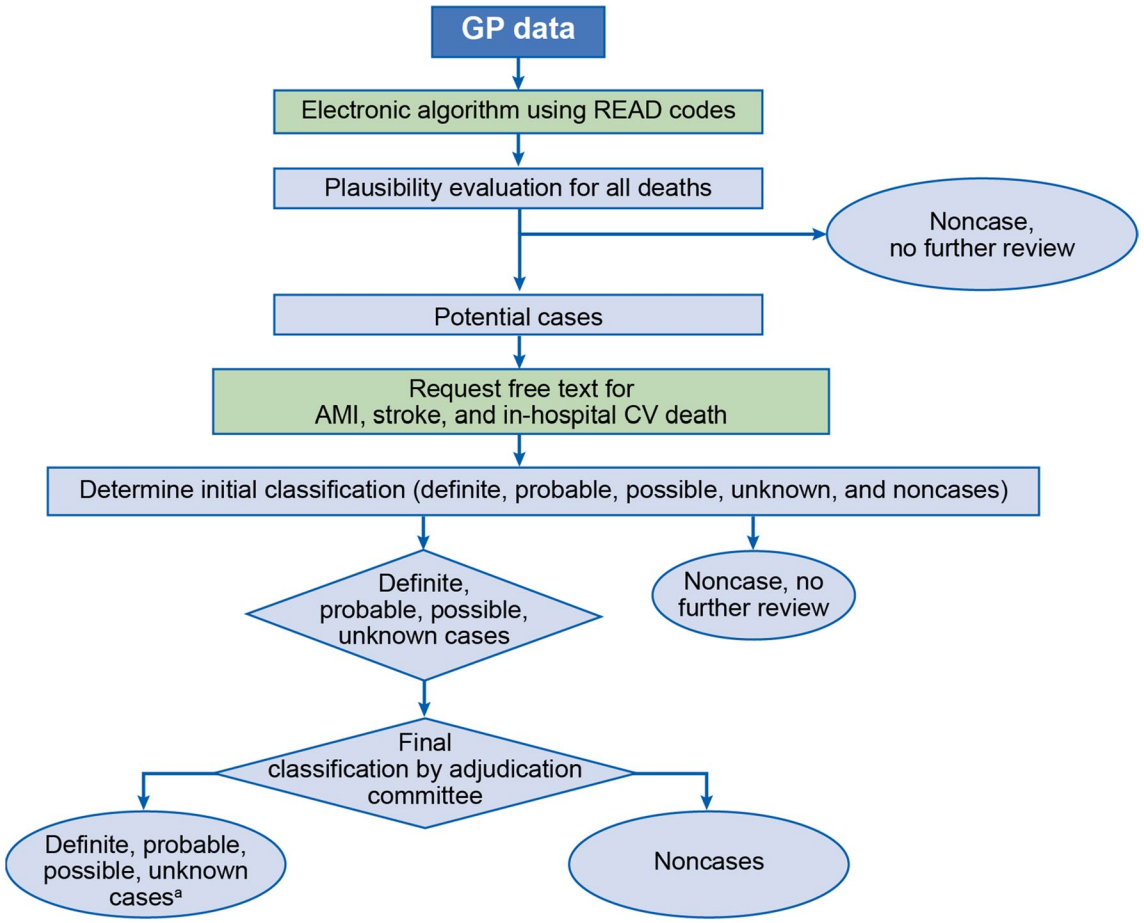


**Fig. 2** Validation process in THIN. *AMI* acute myocardial infarction, *CV* cardiovascular, *GP* general practitioner, *THIN* The Health Improvement Network. [a]Definite cases were reassigned as "confirmed" and unknown cases were reassigned as "noncases"; unknown case classification applied only to cardiovascular deaths.

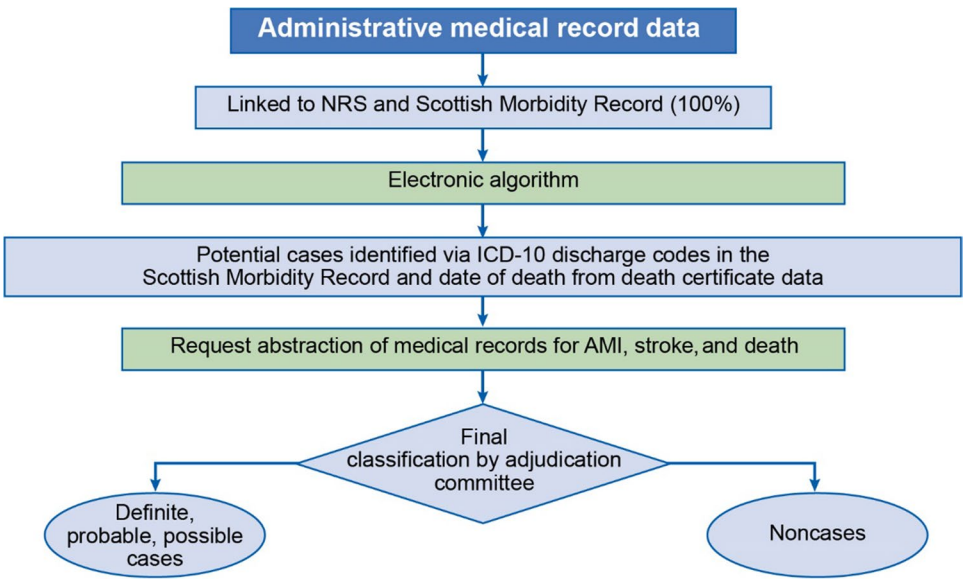


**Fig. 3** Validation process in ISD Scotland. If medical record abstraction (second green box) was not possible for a given patient, case status was assigned based on classification by the electronic algorithm. *AMI* acute myocardial infarction, *ICD-10* International Statistical Classification of Diseases and Related Health Problems, 10th Revision; *ISD* Information Services Division, *NRS* National Records of Scotland

Scorecards were used to document final classifications, as well as their rationale. Clinical reviewers attended a training session, in which they were provided detailed guidance on clinical definitions of study endpoints, on the use of scorecards, and on the levels of certainty to be used for classifying study endpoints [i.e., definite, probable, and possible case; noncase; and unknown cause of death (Table 1)].

Two reviewers independently evaluated each potential case and each provided a classification. In the event of disagreement, a third reviewer also evaluated the event, followed by discussion by the full adjudication committee until a final classification was reached by consensus. Only those potential cases with a final classification of "confirmed" or "definite" after validation were included in the PASS main analysis as MACE cases, whereas the remaining were considered noncases. In a sensitivity analysis, potential cases classified by adjudicators as "probable" (and for death events, as "possible") were also considered to be cases.

### 2.6 Statistical Analyses

We described the flow of the potential cases of the primary study endpoint, from initial electronic identification to final classification after adjudication review, overall and by data source. In CPRD GOLD, we also described the response rates among the GP questionnaires sent to valid practices (active practices accepting questionnaires).

### 2.7 Research Ethics

All relevant authorities reviewed and approved the study on ethical grounds in the UK: the Independent Scientific Advisory Committee for CPRD GOLD; the Scientific Review Committee for THIN; and the East of Scotland Research Ethics Services and the Public Benefit and Privacy Panel for Health and Social Care for ISD, who allowed such access for regulatory purposes for the first time. Any clinical information used for validation purposes was de-identified.

## 3 Results

### 3.1 Prucalopride PASS Population from the UK

The prucalopride PASS initially included 16,426 unique participants from the UK before matching and trimming: 5710 individuals from the CPRD GOLD, 3222 from THIN, and 7494 from ISD Scotland. The characteristics of each of these cohorts, overall and by treatment group, have been described elsewhere [2, 3].

### 3.2 Potential MACE Cases Identified by Electronic Algorithms

Figures 1, 2, 3, and 4 present the flow of the potential study endpoints, from initial identification to final classification, by database as well as overall. The electronic algorithms identified 260 potential events of the primary study endpoint in the three UK data sources: 108 identified in the CPRD GOLD, 79 in THIN, and 73 in ISD Scotland. These identified events could occur both within and beyond the risk window specified in the primary analyses.

Of the 260 potential MACE events identified by the electronic algorithms, 38 cases were considered confirmed via linkage to hospital records (CPRD GOLD only), 56 were ruled out as clear non–cardiovascular death cases (THIN) after review of patient records of all identified deaths, and three were not available for further review (ISD); the remaining 163 events were considered potential cases at this point. After manual review with additional information, 45 were considered noncases (CPRD GOLD and THIN). In the final adjudication of the 118 remaining potential MACE cases, 62 were adjudicated as definite, 10 as probable, 13 as possible, and 33 as noncases (Fig. 4). The GP questionnaire response rate in CPRD GOLD was 80%, free-text was available for all potential events from THIN, and all but three requested hospital case records from ISD were retrieved. This was the first observational study in Scotland in which access to hospital case records was granted.

### 3.3 Validation Flow in the CPRD GOLD

Of the 108 potential cases identified, 38 were automatically confirmed (see criteria for automatic confirmation above) and did not undergo further evaluation (Fig. 1, Table 3). A total of 50 questionnaires were sent to valid practices, with a response rate of 80%. Table 2 shows the response rate for GP questionnaires by type of study outcome. From patient profile review and questionnaire information, 43 were classified as noncases, and the remaining 27 potential cases were reviewed by the adjudication committee. Of these, ten were classified as definite MACE events.

### 3.4 Validation Flow in THIN

Of the 79 electronic identified cases (Fig. 2, Table 3), an initial plausibility evaluation was performed among all deaths before considering them as potential MACE cases. This resulted in ruling out 56 deaths as in-hospital cardiovascular deaths (which did not undergo further validation). Free-text comments were requested for validation purposes for the remaining 23 potential cases. Patient profile review with free-text yielded the identification of two additional noncases (originally considered as one potential AMI and

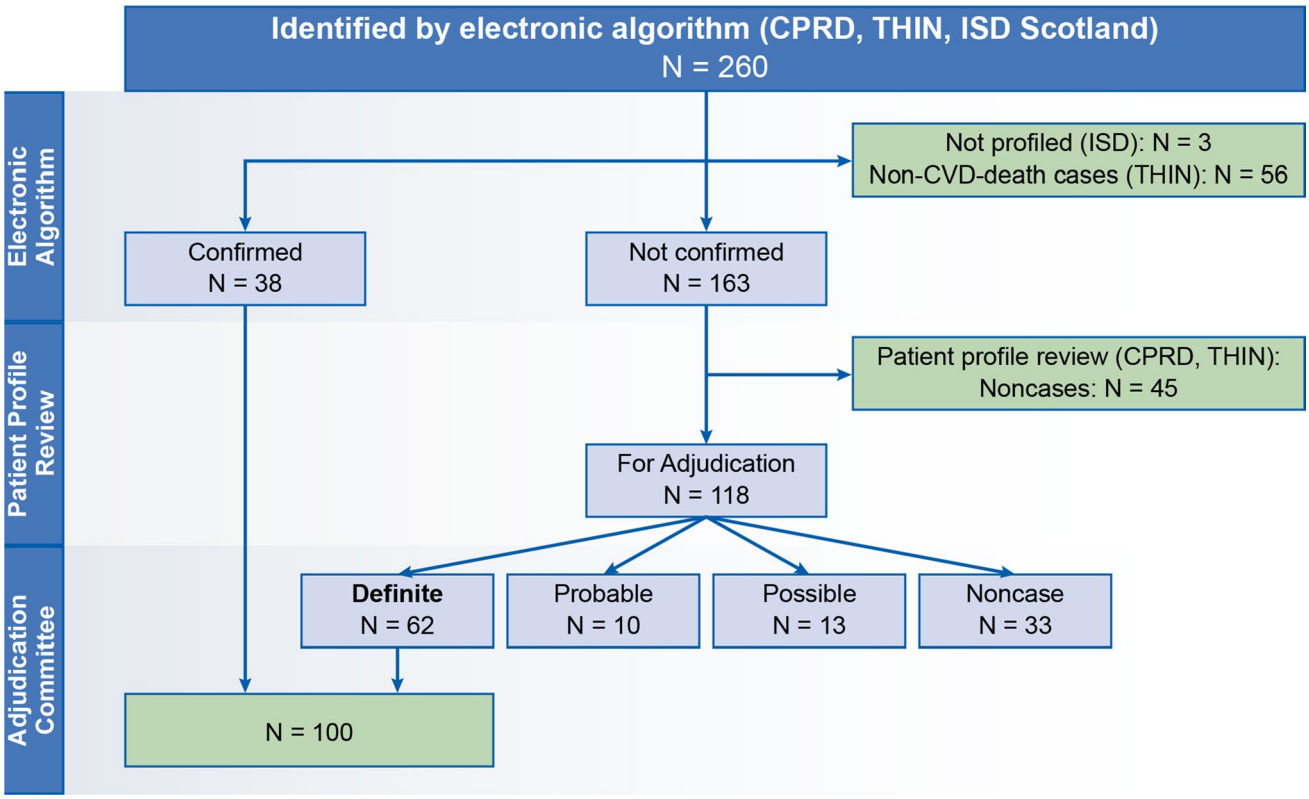


**Fig. 4** Validation flowchart, all UK data sources. *CPRD* Clinical Practice Research Datalink, *CVD* cardiovascular disease, *ISD* Information Services Division, *THIN* The Health Improvement Network, *UK* United Kingdom

**Table 2** Description of the general practitioner questionnaire process in CPRD GOLD

| Endpoint | CPRD GOLD | | | Total |
|---|---|---|---|---|
| | Hospitalization for nonfatal AMI | Hospitalization for nonfatal stroke | CVD death | |
| Number requested | 14 | 18 | 41 | 73 |
| Number sent to valid practices | 6 | 9 | 35 | 50 |
| Number completed | 6 | 6 | 28 | 40 |
| Percentage completed | 100% | 67% | 80% | 80% |

*AMI* acute myocardial infarction, *CPRD* Clinical Practice Research Datalink, *CVD* cardiovascular disease

one potential stroke), and the remaining 21 potential cases were sent to adjudication. Of these, 12 potential MACE cases were adjudicated as definite.

### 3.5 Validation Flow in ISD Scotland

Of the 73 potential cases identified in Scotland (Fig. 3, Table 3), medical chart abstraction could be performed for all except three potential cases (which did not undergo further validation). Thus, 70 potential cases from ISD Scotland were reviewed by the adjudication committee, 40 of which were classified as definite MACE cases.

## 4 Discussion

In the UK component of the prucalopride multidatabase PASS, which included 16,426 unique participants prior to matching and trimming and after de-duplication from three data sources, local adaptations were needed to successfully implement a common event ascertainment and validation protocol. This included patient profile review and use of GP questionnaires in the CPRD GOLD; preliminary plausibility evaluation, patient profile review, and review of free-text comments in THIN; and medical record abstraction in ISD Scotland. Of the 260 potential MACE events identified

**Table 3** Validation of study outcomes by data source

| | All UK data sources | CPRD GOLD | THIN | ISD Scotland |
|---|---|---|---|---|
| Total PEG/prucalopride new users | 16,426 | 5710 | 3222 | 7494 |
| ***Primary endpoint: MACE*** | | | | |
| Electronic algorithm identification | 260 | 108 | 79 | 73 |
| Excluded after initial plausibility evaluation (THIN)/not available (ISD Scotland) | 59 | – | 56 | 3 |
| Potential cases for validation | 201 | 108 | 23 | 70 |
| Confirmed/definite cases | 100 | 48 | 12 | 40 |
| ***Secondary endpoints: individual components of MACE*** | | | | |
| **Hospitalization for AMI** | | | | |
| Electronic algorithm identification | 68 | 28 | 11 | 29 |
| Potential cases for validation | 68 | 28 | 11 | 29 |
| Confirmed/definite cases | 42 | 17 | 8 | 17 |
| **Hospitalization for stroke** | | | | |
| Electronic algorithm identification | 65 | 30 | 7 | 28 |
| Case records not available (ISD Scotland) | 1 | – | – | 1 |
| Potential cases for validation | 64 | 30 | 7 | 27 |
| Confirmed/definite cases | 33 | 17 | 3 | 13 |
| **In-hospital cardiovascular death** | | | | |
| Electronic algorithm identification | 127 | 50 | 61 | 16 |
| Excluded after initial QC check (THIN)/not available (ISD Scotland) | 58 | – | 56 | 2 |
| Potential cases for validation | 69 | 50 | 5 | 14 |
| Confirmed/definite cases | 25 | 14 | 1 | 10 |

*AMI* acute myocardial infarction, *CPRD* Clinical Practice Research Datalink, *ISD* Information Services Division, *MACE* major adverse cardiovascular events, *PEG* polyethylene glycol 3350, *QC* quality control, *THIN* The Health Improvement Network, *UK* United Kingdom

by the electronic algorithms (108 from CPRD GOLD, 79 from THIN, and 73 from ISD), 100 were considered actual events after identification of confirmed events, review, and the adjudication process, with the CPRD GOLD being the largest contributor.

It is important to note that the definition used to identify potential causes of in-hospital cardiovascular death was, per regulatory request, very broad. This likely inflated the number of initial potential cases identified by the algorithms. Researchers conducting this type of study need to leverage the sensitivity of their endpoint definitions (very sensitive definitions yielding very few "false negatives," but many "false positives") with the resources available for case validation purposes—which may be insufficient if the potential number of events identified is large. In the prucalopride PASS, because the study population was young and death rates were low during the study period, using a very sensitive definition generated a reasonable number of potential events to review. In addition, the preadjudication patient profile review resulted in the identification of many obvious noncases, which reduced markedly the burden on the adjudication committee. Furthermore, to ensure the capture of all MACE, all potential events identified were subject to validation, not only those captured during the primary event risk window.

Additional factors may have contributed to the difference in identification and case confirmation proportions, particularly for in-hospital cardiovascular death events. In each data source, some sources of clinical information were lacking: for the CPRD GOLD, free-text comments were not available for validation purposes, and the response rate of the GP questionnaires was slightly lower than anticipated. Thus, for some potential CPRD GOLD cases, the patient profile was the only source of information for adjudicators to review. Linkage to the HES and ONS was possible in CPRD GOLD for approximately 50% of the patients. In THIN, no GP questionnaires were obtained and no linkage to the HES or ONS was possible. However, free-text comments were available for manual review for all potential cases. These features may have limited the ability of clinical reviewers to adjudicate the study events, particularly regarding assignment of cause of death events, for which the information included in the patient profiles/medical records was often very limited. This resulted in many events classified as "unknown" cause of death, which were considered noncases for analysis purposes. In CPRD GOLD, in-hospital cardiovascular deaths

comprised the majority of potential events identified by the algorithms compared with the other data sources.

As mentioned previously, prior published MACE clinical definitions were used to create electronic algorithms to ascertain cases in health care databases [13–20]. A sensitivity analysis was performed that relaxed the required alignment with the algorithmic definitions used in the study. In this sensitivity analysis, where potential cases classified by adjudicators as "probable" (and for death events, as "possible") were also considered to be cases, the results were consistent with those from the main analysis. Key lessons learned from this event validation exercise include the following: (1) the need to perform validation of outcomes in order to substantiate the validity of the study outcomes used in the analyses using these data sources; (2) the need to locally adapt study protocols to the disparate types of clinical data available in each data source; (3) the importance of using data sources with detailed clinical information for validation purposes, particularly for studies including cardiovascular death events; (4) the value of conducting preliminary review of potential cases in order to rule out obvious noncases, as a means for reducing the burden on the adjudication committee; and (5) the importance of involving clinical expert reviewers in the study validation.

## 5 Conclusions

A common validation protocol, with local adaptations of search algorithms and validations steps based on the types of clinical information available in each data source, allowed for the validation of MACE endpoints in the prucalopride multidatabase PASS in three UK data sources. Given the limitations of relying solely on computer algorithms to identify cardiovascular outcomes, validation with clinical review is essential to guide interpretation.

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s40264-021-01044-4.

**Acknowledgements** Jennifer Bartsch from RTI Health Solutions contributed to the statistical programming. Abenah Harding from RTI Health Solutions provided quality checks and support for the development of the manuscript and assistance with responding to reviewer comments. Kate Lothman from RTI Health Solutions provided medical writing support. Helle Kieler contributed to the study as a primary investigator in Sweden. Alison McGinnis, a research nurse, toured Scottish hospitals to gather original hospital case records for the purposes of endpoint adjudication. The authors acknowledge the Electronic Data Research and Innovation Service (eDRIS) of ISD Scotland for providing the ISD data. Amy Rogers and Kerr Grieve provided assistance with validation. This study is based in part on data from the CPRD obtained under license from the UK Medicines and Healthcare products Regulatory Agency. The data are provided by patients and collected by the National Health Service (NHS) as part of their care and support. The interpretation and conclusions contained in this study are those of the authors alone. 

## Declarations

**Funding** This study was funded by Shire Pharmaceuticals GmbH, now part of the Takeda group of companies.

**Conflict of interest** This study was performed under a research contract between RTI Health Solutions and Shire Pharmaceuticals GmbH, now part of the Takeda group of companies. The content of the publication was developed independently from the sponsor. Alicia Gilsenan, Joan Fortuny, Estel Plana, Miguel Cainzos-Achirica, Ryan Ziemiecki, and Elizabeth B. Andrews are or were salaried employees of RTI Health Solutions at the time the study was conducted. Miguel Cainzos-Achirica now works at Houston Methodist DeBakey Heart and Vascular Center and has participated in a research study funded by a non-conditioned research grant from Vifor. Ana Ruigómez and Luis-Alberto Garcia-Rodriguez work at the Spanish Centre for Pharmacoepidemiologic Research (CEIFE Madrid, Spain), which has received research funding from RTI Health Solutions. Luis-Alberto Garcia Rodriguez also received honoraria for serving on advisory boards for Bayer AG. Robert W.V. Flynn is an employee of the University of Dundee and has received research grants from Novartis and GlaxoSmithKline. Professor Thomas M. MacDonald is an employee of the University of Dundee who received financial support under the research contract between RTI Health Solutions and Shire Pharmaceuticals GmbH, now part of the Takeda group of companies, to carry out the study. Thomas M. MacDonald also works with MEMO Research at the University of Dundee and received or has received research funding from Novartis, Amgen, Ipsen, GRK Teijin, and Menarini and has received honoraria for educational lectures from Takeda. He has been the Principal Investigator on trials paid for by Novartis, Ipsen, Teijin, GlaxoSmithKline, and Menarini. In the last 3 years, he has been paid consulting fees by Novartis.

**Ethics approval** All relevant authorities reviewed and approved the study on ethical grounds in the UK: the Independent Scientific Advisory Committee for CPRD GOLD; the Scientific Review Committee for THIN; and the East of Scotland Research Ethics Services and the Public Benefit and Privacy Panel for Health and Social Care for ISD, who allowed such access for regulatory purposes for the first time.

**Consent to participate** Not applicable.

**Consent for publication** Not applicable.

**Availability of data and material** For data obtained from the CPRD and THIN, data access is available and subject to a license agreement that contains terms and conditions of use set forth by UK Medicines and Healthcare Products Regulatory Agency. For data obtained from ISD, data access is available subject to the terms and conditions set forth by the NHS National Services Scotland.

**Code availability** Not applicable.

**Authors' contributions** AR, AG, JF, and MC contributed to the conception and planning of the work, analysis and interpretation of the data, and drafting the manuscript. EP contributed to the analysis and interpretation of the data and drafting of the manuscript. RF, TM, LG, RZ, and EA contributed to the analysis and interpretation of the data and critical revision of the manuscript. All authors read and approved the final version.

## References

1. EMA. Resolor (prucalopride). European Medicines Agency. 2016. http://www.ema.europa.eu/ema/index.jsp?curl=pages/medicines/human/medicines/001012/human_med_001030.jsp&mid=WC0b01ac058001d124. Accessed 12 Nov 2020.
2. Gilsenan A, Fortuny F, Cainzos-Achirica M, Cantero AF, Flynn RWV, Garcia-Rodriguez L, et al. Cardiovascular safety of prucalopride in patients with chronic constipation: a multinational population-based cohort study. Drug Saf. 2019;42:1179–90.
3. Fortuny F, Gilsenan A, Cainzos-Achirica M, Cantero AF, Flynn RWV, Garcia-Rodriguez L, et al. Study design and cohort comparability in a study of major cardiovascular events in new users of prucalopride versus polyethylene glycol 3350. Drug Saf. 2019;42:1167–77.
4. ENCePP. Cohort study of the relative incidence of major cardiovascular events among patients initiating prucalopride versus a matched comparator cohort (EUPAS9200). European Network of Centres for Pharmacoepidemiology and Pharmacovigilance. 2018. http://www.encepp.eu/encepp/viewResource.htm?id=22643 . Accessed 12 Nov 2020.
5. Varas-Lorenzo C, Castellsague J, Stang MR, Tomas L, Aguado J, Perez-Gutthann S. Positive predictive value of ICD-9 codes 410 and 411 in the identification of cases of acute coronary syndromes in the Saskatchewan Hospital automated database. Pharmacoepidemiol Drug Saf. 2008;17:842–52.
6. Arana A, Margulis AV, Varas-Lorenzo C, Bui CL, Gilsenan A, McQuay LJ, et al. Validation of cardiovascular outcomes and risk factors in the clinical practice research datalink in the United Kingdom. Pharmacoepidemiol Drug Saf. 2020. https://doi.org/10.1002/pds.5150.
7. Bush M, Sturmer T, Stearns SC, Simpson RJ Jr, Brookhart MA, Rosamond W, et al. Position matters: validation of medicare hospital claims for myocardial infarction against medical record review in the atherosclerosis risk in communities study. Pharmacoepidemiol Drug Saf. 2018;27:1085–91.
8. Chung CP, Murray KT, Stein CM, Hall K, Ray WA. A computer case definition for sudden cardiac death. Pharmacoepidemiol Drug Saf. 2010;19:563–72.
9. Ludvigsson JF, Andersson E, Ekbom A, Feychting M, Kim JL, Reuterwall C, et al. External review and validation of the Swedish national inpatient register. BMC Public Health. 2011;11:450.
10. Hammar N, Alfredsson L, Rosen M, Spetz CL, Kahan T, Ysberg AS. A national record linkage to study acute myocardial infarction incidence and case fatality in Sweden. Int J Epidemiol. 2001;30(Suppl 1):S30–4.
11. Schubert I, Ihle P, Koster I. Internal confirmation of diagnoses in routine statutory health insurance data: concept with examples and case definitions. Gesundheitswesen. 2010;72:316–22.
12. Peters F, Kreutzburg T, Kuchenbecker J, Marschall U, Remmel M, Dankhoff M, et al. Quality of care in surgical/interventional vascular medicine: what can routinely collected data from the insurance companies achieve? Gefässchirurgie. 2020. https://doi.org/10.1007/s00772-020-679-4.
13. Thygesen K, Alpert JS, Jaffe AS, Simoons ML, Chaitman BR, White HD, et al. Third universal definition of myocardial infarction. Circulation. 2012;126:2020–35.
14. Hicks KA, Hung HMJ, Mahaffey KW, Mehran R, Nissen SE, Stockbridge NL, et al. Standardized definitions for cardiovascular and stroke endpoint events in clinical trials. In: Clinical Data Interchange Standards Consortium (standardized data collection for cardiovascular trials initiative). 2014. Accessed 12 Nov 2020.
15. Sacco RL, Kasner SE, Broderick JP, Caplan LR, Connors JJ, Culebras A, et al. An updated definition of stroke for the 21st century: a statement for healthcare professionals from the american heart association/american stroke association. Stroke. 2013;44:2064–89.
16. Andersohn F, Schade R, Suissa S, Garbe E. Cyclooxygenase-2 selective nonsteroidal anti-inflammatory drugs and the risk of ischemic stroke: a nested case-control study. Stroke. 2006;37:1725–30.
17. Arana A, Varas C, Gonzalez-Perez A, Gutierrez L, Bjerrum L, Garcia-Rodriguez LA. Hormone therapy and cerebrovascular events: a population-based nested case-control study. Menopause. 2006;13:730–6.
18. Ray WA, Varas-Lorenzo C, Chung CP, Castellsague J, Murray KT, Stein CM, et al. Cardiovascular risks of nonsteroidal antiinflammatory drugs in patients after hospitalization for serious coronary heart disease. Circ Cardiovasc Qual Outcomes. 2009;2:155–63.
19. Roumie CL, Mitchel E, Gideon PS, Varas-Lorenzo C, Castellsague J, Griffin MR. Validation of ICD-9 codes with a high positive predictive value for incident strokes resulting in hospitalization using Medicaid health data. Pharmacoepidemiol Drug Saf. 2008;17:20–6.
20. Ruigómez A, Martín-Merino E, Rodríguez LA. Validation of ischemic cerebrovascular diagnoses in The Health Improvement Network (THIN). Pharmacoepidemiol Drug Saf. 2010;19:579–85.
21. Cai B, Xu W, Bortnichak E, Watson DJ. An algorithm to identify medical practices common to both the general practice research database and the health improvement network database. Pharmacoepidemiol Drug Saf. 2012;21:770–4.