


# Exploring the Potential Routine Use of Electronic Healthcare Record Data to Strengthen Early Signal Assessment in UK Medicines Regulation: Proof-of-Concept Study

Katherine Donegan<sup>1</sup>  · Rebecca Owen<sup>1</sup> · Helena Bird<sup>1</sup> · Brian Burch<sup>1</sup> · Alex Smith<sup>1</sup> · Phil Tregunno<sup>1</sup>

Published online: 3 May 2018  
© Springer International Publishing AG, part of Springer Nature 2018

## Abstract

**Introduction** Electronic healthcare record (EHR) databases are used within pharmacoepidemiology studies to confirm or refute safety signals arising from spontaneous adverse event reports. However, there has been limited routine use of such data earlier in the signal management process, to help rapidly contextualise signals and strengthen preliminary assessment or to inform decisions regarding action including the need for further studies. This study explores the value of EHR used in this way within a regulatory environment via an automated analysis platform.

**Methods** Safety signals raised at the UK Medicines and Healthcare products Regulatory Agency (MHRA) between July 2014 and June 2015 were individually reviewed by a multi-disciplinary team. They assessed the feasibility of identifying the exposure and event of interest using primary care data from the Clinical Practice Research Datalink (CPRD) within the Commonwealth Vigilance Workbench (CVW) Longitudinal Module platform, which was designed to facilitate routine descriptive analysis of signals using EHR. Three signals, where exposure and event could be well identified, were retrospectively analysed using the platform.

**Results** Of 69 unique new signals, 20 were for drugs prescribed predominately in secondary care or available without prescription, which would not be identified in

primary care. A further 17 were brand, formulation, or dose-specific issues, were related to mortality, were relevant only to a subgroup of patients, or were drug interactions, and hence could not be reviewed using the platform given its limitations. Analyses of exposure and incidence of the adverse event could be produced using CPRD within the CMV Longitudinal Module for 32 (46%) signals. The case studies demonstrated that the data provided supporting evidence for confirming initial assessment of the signal and deciding upon the need for further action.

**Conclusions** CPRD can routinely provide useful early insights into clinical context when assessing a large proportion of safety signals within a regulatory environment provided that a flexible approach is adopted within the analysis platform.

## Key Points

Simple analyses of electronic healthcare record data can routinely provide useful insights into the clinical context when assessing the strength of new safety signals arising from spontaneous adverse event reports or elsewhere within a regulatory environment to strengthen decision making.

Further work is still required to refine the platform used and to establish how the signal management process should operate efficiently, and in line with good pharmacovigilance practices, to maximise the value of the additional available data.

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s40264-018-0675-x>) contains supplementary material, which is available to authorized users.

✉ Katherine Donegan  
katherine.donegan@mhra.gov.uk

<sup>1</sup> Vigilance and Risk Management of Medicines, MHRA, 151 Buckingham Palace Road, London SW1W 9SZ, UK

## 1 Introduction

The value of spontaneous reporting systems in pharmacovigilance is well established [1, 2]. Individual case reports of adverse events made by patients and healthcare professionals to regulatory agencies can provide valuable insights into unforeseen risks potentially associated with medicines and can be individually and collectively explored to try to assess causality. Given the large volume of such reports, disproportionality analyses are routinely used to try to detect safety signals where there may be a true association [3, 4]. Despite use of these algorithms, the assessment of causality in signals arising from spontaneous data is complicated by the limitations of such reporting schemes, which rely on reporter suspicion, are subject to varying levels of under-reporting, and which lack comparator data in unexposed patients meaning that the nature and impact of any biases in reporting is unclear [5–7].

Electronic healthcare record (EHR) databases hold extensive longitudinal data, including prescriptions, clinical diagnoses and test results, on an individual patient-level basis [8]. They have been used repeatedly in epidemiological studies to test the association of a risk of a certain adverse event with a specific medicine as well as to explore prescribing patterns and the incidence and prevalence of different medical conditions in the population [9–11]. Simple extracts of EHR data have also been used to support vaccine surveillance by enabling individual spontaneous case reports to be placed into the context of age- and gender-specific background event rates and a rapidly increasing level of exposure [12, 13]. There has also been some wider interest in routinely utilising longitudinal patient healthcare data in the earlier phases of the pharmacovigilance cycle; providing data to support strengthened assessment and management of signals following their initial detection within large spontaneous adverse event report databases [14], particularly through the US FDA Sentinel Initiative, which has successfully demonstrated the added value of integrating summaries of healthcare data with their existing Adverse Event Reporting System to support early regulatory decision making [15]. In addition, there has also been exploration of the use of EHR data for signal detection purposes alongside traditional signal detection methods with the aim of addressing some of the limitations of spontaneous adverse event report schemes [16].

There is clear value in being able to place safety signals arising from case reports into the context of the size and characteristics of the treated population at an early stage to inform their initial assessment and EHR data could be a valuable tool for this. In addition, EHR could provide data on further potential cases and be used to explore the

temporality between exposure and events in the wider population and the prevalence of confounding factors. These data could then be used to help inform early decision-making regarding the need for further action and prioritisation. However, more studies are needed to establish if, when and how such data can provide useful insights that will increase efficiency and/or scientific robustness earlier in signal management and within an EU regulatory environment.

In this paper, we describe a proof-of-concept study conducted to further explore the value of EHR data to support routine pharmacovigilance processes, including early signal assessment and management as well as related decision making, within the UK Medicines and Healthcare products Regulatory Agency (MHRA). This was done through use of an automated software platform that provides simple routine analyses of EHR data designed specifically to explore individual drug and adverse event combinations.

## 2 Methods

### 2.1 The Yellow Card Scheme and Signal Detection at the UK Medicines and Healthcare products Regulatory Agency (MHRA)

In the UK, the Yellow Card Scheme (<https://yellowcard.mhra.gov.uk/>), run by the MHRA, enables healthcare professionals, patients and their carers to report any adverse events which they suspect may be associated with the use of a medicine. The MHRA also receives any reports of suspected adverse drug reactions made to pharmaceutical companies, who are legally required to submit such reports to regulatory agencies. Reports are entered on to the Yellow Card database with adverse events coded using MedDRA<sup>®</sup>. Disproportionality analyses, using the empirical Bayes geometric mean (EBGM) [3] and rule-based approaches, with pre-defined signalling threshold criteria are run on the whole database on a weekly basis and a preliminary assessment is made of all resulting signals by pharmacovigilance assessors. Drug–event combinations (DEC) judged to raise a possible safety concern based on at least one of the criteria are reviewed by an assessor and discussed at a weekly signal meeting that has further scientific and medical assessor input. Any potential safety concern that is felt to require further consideration regarding the need for action is then formally identified as a signal, according to the definition within the EU Good Pharmacovigilance Practice (GVP) guidelines [17]. It is taken to the weekly Signal Management Review Meeting (SMRM), along with signals raised from other sources such as those identified from epidemiological studies or

randomised clinical trials. In the SMRM, the need for, and form of, any next steps or further action is decided upon alongside a consideration of prioritisation and timelines. Actions that can be considered at this stage include requesting further information or data from Marketing Authorisation Holders, initiating further review, research, or strengthened monitoring, the introduction of amendment of risk minimisation measures, or communications directed to healthcare professionals or patients, for example. Where SMRM considers a signal to be validated [18], it will be transmitted to the EU network for consideration and agreement of actions by the Pharmacovigilance Risk Assessment Committee.

If, after thorough assessment, a causal relationship between the drug and the adverse event is at least a reasonable possibility, then it is included within Section 4.8 *Undesirable effects* of the Summary of Product Characteristics (SmPC) [19], the legal document that forms the basis of information for healthcare professionals regarding how to use the product, that must be approved as part of the marketing authorisation and continuously updated as new data emerge.

## 2.2 The Clinical Practice Research Datalink

The UK Clinical Practice Research Datalink (CPRD; <https://www.cprd.com>) holds a large database of longitudinal primary care EHR that now includes over 14 million patient records from over 700 general practices in the UK. It holds comprehensive records of all prescriptions issued and vaccines administered in primary care along with some patient demographics, clinical diagnoses, referrals to secondary care, and test results [20]. All clinical diagnoses are coded using Read Codes. The database has been used extensively in pharmacoepidemiology research, which has supported the MHRA's regulatory decision making. Use of the CPRD primary care data for this study was approved by the CPRD Independent Scientific Advisory Committee (Protocol number: 17\_119) [21]. The data build used in this study was released in July 2015.

## 2.3 Commonwealth Vigilance Workbench Longitudinal Module

For this study, CPRD primary care data were analysed using the Commonwealth Vigilance Workbench (CVW) Longitudinal Module which was engineered by Commonwealth Informatics, Inc. and is a reimplement and extension of the *vigiTrace* research software developed by the Uppsala Monitoring Centre. *vigiTrace*, and the Temporal Pattern Discovery analytic framework on which it is based, have been described in detail previously [14, 22, 23]. In brief, and in the context of this study, the

platform provided a set of routine analyses of the CPRD patient, practice, clinical and therapy datasets. The analyses available included summaries of the characteristics of the whole CPRD population and of patients with new episodes of prescribing of the specified drug substance and/or the adverse event of interest including, for example, the proportion of men versus women and the age breakdown. A new episode of prescribing for a specific substance is defined to start on the day of a prescription for that substance when there are no prescriptions for the same substance in the preceding 390 days in that patient's medical history.

Chronographs, a graphical approach to comparing the observed rate of recording of a pre-specified adverse event, in 30-day intervals relative to the start of new prescribing episodes of the drug of interest, to the expected rate, were also provided [22]. The expected rate is calculated using the number of recorded adverse events in all patients relative to the start of new prescribing episodes for all drug substances recorded in the CPRD. These figures can be used to start to explore the association of adverse events with starting a new treatment episode, with the strength of the association for each monthly period assessed using the Information Component (IC; the logarithm of the shrunk observed-to-expected ratio) and associated 95% confidence intervals. If the IC is  $> 0$  then the adverse event occurs more commonly with the drug substance of interest than overall in the database. Examples of these chronographs can be found in the results section of this paper. Specifically, the top panel shows the IC along with associated 95% confidence intervals while the bottom panel shows the absolute observed and expected numbers of adverse events. The observed first event is defined as the first record of that adverse event within an individual patient's medical record. These statistics are presented on a 30-day basis on the horizontal axis centred on the start of a new prescription episode defined as time 0.

In addition, the  $ICA$  statistic was also calculated to examine temporal associations in a calibrated self-controlled cohort approach. This statistic is computed as the ratio of the observed-to-expected ratio, as described above, in a surveillance period to that in a control period, subject to statistical shrinkage. Specifically, the observed-to-expected ratios in the first 30 days after the new prescription and day 31–360 inclusive, are compared with four pre-exposure control periods: the day of the new prescription, 30 days prior to the new prescription, 360–31 days prior to the new prescription, and a longer historical control period ranging from 3 years prior to 1 year prior to the new prescription.

The facility to explore summaries of concomitant medications and prior diagnoses in patients with a record of the specified adverse event following a prescription of the

specified drug substance was also available, as was the option to explore individual patient records.

## 2.4 Study Plan

This pilot study was devised based upon the current signal detection and management processes of the MHRA. An initial retrospective screening of all DECAs raised as new signals at the weekly SMRM between July 2014 and June 2015 inclusive was conducted by the team of pharmacovigilance scientists who manage the SMRM. Signals related to events already recognised and listed in Section 4.8 of the SmPC at the time of their presentation at the SMRM where the signal suggested a potential change in the risk (e.g. in terms of absolute or relative risk, duration, severity, or outcome following the event) were excluded as this proof of concept study used only one release of the CPRD database and as such could not compare analyses from different time points. Other signals that raised the need for alignment across different product information for the same drug substance were excluded as these related to recognised reactions rather than new signals and were administrative changes to ensure the provision of consistent information.

Further signals were also excluded as, due to limitations of the CPRD data and the CMV Longitudinal Module platform, it was clear that either the drug or event of interest would not be identifiable or analysable within the study. The additional signals excluded were (i) signals that were brand, formulation, or dose-specific related issues, as this data is either not available or inconsistently recorded in the CPRD and analyses conducted by the CVW Longitudinal Module platform were only at the drug substance level; (ii) signals that were for products prescribed exclusively in secondary care or widely available without a prescription for which prescription data would not be available in a primary care database; (iii) signals where the outcome was fatal as analyses within the platform identified events via Read codes, which would likely not be used to record death in the CPRD; (iv) signals that were only relevant for a subgroup of patients such as children, as the analyses conducted in the platform only worked across the whole patient population; and (v) signals related to drug interactions as the platform only analysed combinations of a single drug substance with a single event.

All remaining unique DECAs identified were then taken forward to a further review by a wider panel of pharmacovigilance scientists (including those who conducted the initial screening) and a pharmacoepidemiologist, experienced in the assessments of safety signals and the use and interpretation of analyses of primary care EHR data, respectively, using the CPRD data within the CVW Longitudinal Module platform. The purpose of this review was

firstly to evaluate the extent of prescribing data available for the relevant drug substances within the CPRD. Secondly, it was to consider the likelihood of valid event recording in primary care records, available to a specific-enough level using individual Read codes to be relevant for the signal in question, and hence retrospectively assess the feasibility of using the CPRD data within the software platform available to produce analyses robust enough to potentially help inform the decision making that occurred at the SMRM meeting when the signal was raised had they been available. This review was supported by data from IQVIA (previously QuintilesIMS) MIDAS<sup>TM</sup>, which captures the volume of drugs dispensed against a prescription in UK retail and hospital pharmacies, and that sold without a prescription (over-the-counter) in retail pharmacies. The retail dispensing data are based on volumes of products dispensed against a prescription in retail pharmacies and wholesaler sell-in data to dispensing doctors (a proxy for dispensed product). The hospital dispensing data covers usage/consumption levels of medicinal products within hospitals (irrespective of their source of supply). Over-the-counter sales data are estimated by excluding the data for stock dispensed against a prescription from data on the volume stock ordered via wholesalers by pharmacies, doctors and hospitals. This data was used to estimate the proportion of items (e.g. boxes of tablets) that were dispensed against a prescription made within a hospital setting, and which would hence be missing from a primary care database.

Following assessment of the signals at this stage of the review, the decisions taken for each signal at the time of its first discussion at SMRM were extracted from meeting documentation and minutes.

Finally, a series of three more detailed case studies was conducted to explore the scientific value of the data in helping to confirm or refute the previous regulatory decision taken when the signal was discussed at SMRM. The three case studies were chosen for in-depth examination using CPRD data within the platform based on findings from the initial review because either both the exposure and event could be very clearly identified, meaning that the fuller potential scientific value of the analyses provide by the software could be explored with minimal impact from some of the limitations of using EHR data from primary care, or they were particularly difficult signals to assess potential causality using spontaneous data alone, meaning there was a larger potential added value in using EHR data. The choice of signals for further exploration was made by the full review team based upon experience; no specific criteria were used.

### 3 Results

#### 3.1 Potential Extent of Use of Primary Care Data for Strengthening Regulatory Signal Detection: Initial Screening of Possible Signals

A total of 83 unique signals were discussed at the SMRM during the study period, 14 of which were immediately excluded from further consideration in this study at the first review stage as they related to known issues already listed in the SmPC, but where there was inconsistent product information or where there was a signal of a potential change in the magnitude of a known risk. A summary of the remaining DECs excluded from analysis with the CVW Longitudinal Module according to the pre-specified exclusion criteria is presented in Fig. 1.

Of the 16 DECs that were excluded for further exploration with the CVW due to limitations in the set-up of the platform, six were excluded because they were either brand, formulation, or dose-specific related issues, three because the outcome was fatal, three because they were only relevant for a subgroup of patients such as children, and five because they related to drug interactions.

The remaining 32 DECs were taken forward to the second stage of the review, with a wider panel of pharmacovigilance scientists and a pharmacoepidemiologist, using the CPRD data and CVW platform—46% of those originally identified as new signals. Three of these were for drug substances subject to EU additional monitoring.

#### 3.2 Identification of Exposures and Adverse Events: Further Review Using the CPRD and CVW Longitudinal Module Platform

For each of the 32 DECs identified for further exploration, the ability to identify both the exposure and the event from the CPRD data analysed within the CMV Longitudinal Module was considered. Prescriptions for all drug substances involved in the 32 signals could be identified using

the CPRD data within the platform. Despite the initial screening that excluded signals for products known to be exclusively or predominantly prescribed in settings other than primary care, all drug substances were likely to have some level of prescribing in secondary care, although it was estimated, using data from IQVIA MIDAS™, that for only 6 (19%) were over 10% of prescriptions likely to be made outside primary care in the UK.

Five (16%) of the signals related to adverse events that were captured by one of the custom group Read code lists already defined within the CMV Longitudinal Module, as they are often seen reported as suspected adverse events associated with prescribed medicines [24]. The remaining adverse events could be identified by one or more Read codes at the pre-defined level and their sub-codes. Eighteen (56%) of the signals included adverse events that would likely present and be diagnosed within primary care, while 11 (34%) would likely present to a GP in the first instance but diagnosis would be confirmed within secondary care.

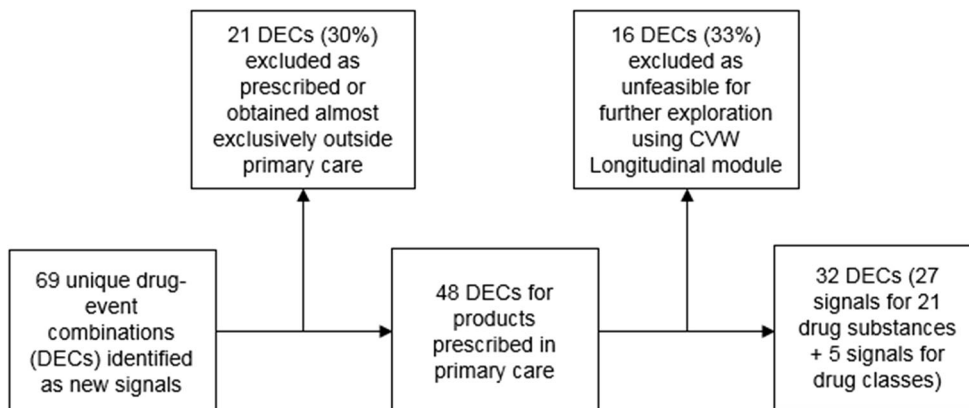
Of the 32 signals, when first raised at SMRM, 11 (34%) were taken forward into the European network or referred to the relevant lead Member State for either further review or action or were already being considered at an EU-wide level. Three (9%) were taken forward nationally for immediate action while further data was sought to support the review or for ongoing assessment for 11 (34%). The remainder (7, 22%) were not taken forward at that stage as the evidence was either inconclusive or suggestive of a non-causal association or no further regulatory action was considered necessary at that time.

#### 3.3 Case Studies

Background information for the three case studies is described in Table 1.

The identification of each DEC, along with the summary of the CPRD data, is presented below. Full code lists for the adverse events identified in the three case studies can be found in the Electronic Supplementary Material.

**Fig. 1** Flow chart showing impact of exclusion criteria on number of DECs included in study. CVW Commonwealth Vigilance Workbench



**Table 1** Background information for the three case studies

Signal (drug substance and MedDRA® preferred term)	UK authorised indication of use for the drug	No. of Yellow Card case reports reported at the time of the signal (EBGM) <sup>a</sup>	Comments	Action previously taken following review by the SMRM
Donepezil and salivary hypersecretion	A cholinesterase inhibitor indicated for the symptomatic treatment of mild to moderately severe Alzheimer's dementia	12 (5.1)	Salivation was listed in Section 4.9 of the SmPC in relation to the symptoms of overdose and is also listed in Section 4.8 for rivastigmine, another acetylcholinesterase inhibitor	Updates to the product information for donepezil, to include salivary hypersecretion in Section 4.8 of the SmPC, were implemented
Mirtazapine and alopecia	An atypical antidepressant used primarily for the treatment of episodes of major depression	19 (0.74)	Two of the cases demonstrated positive re-challenge and alopecia is listed for some other antidepressants	The decision was that no further regulatory action was warranted given the small number of cases, the extensive use of mirtazapine and the length of time the product had already been available on the market
Finasteride and cognitive disorder	A 5 $\alpha$ -reductase inhibitor indicated for use in benign prostatic hyperplasia and male pattern baldness	13 (8.30)	The signal was first raised from Yellow Cards in 2012. A number of papers have been published exploring the risk of cognitive impairment with finasteride, but no definitive link has been established. Decreased libido and depressed mood are already listed in Section 4.8 of the SmPC for finasteride	The evidence was deemed inconclusive and therefore the decision was made that no further regulatory action was warranted at this stage, although the signal remained under review

EBGM empirical Bayes geometric mean, SmPC Summary of Product Characteristics, SMRM Signal Management Review Meeting

<sup>a</sup>At the point within the study period when the signal was discussed at SMRM

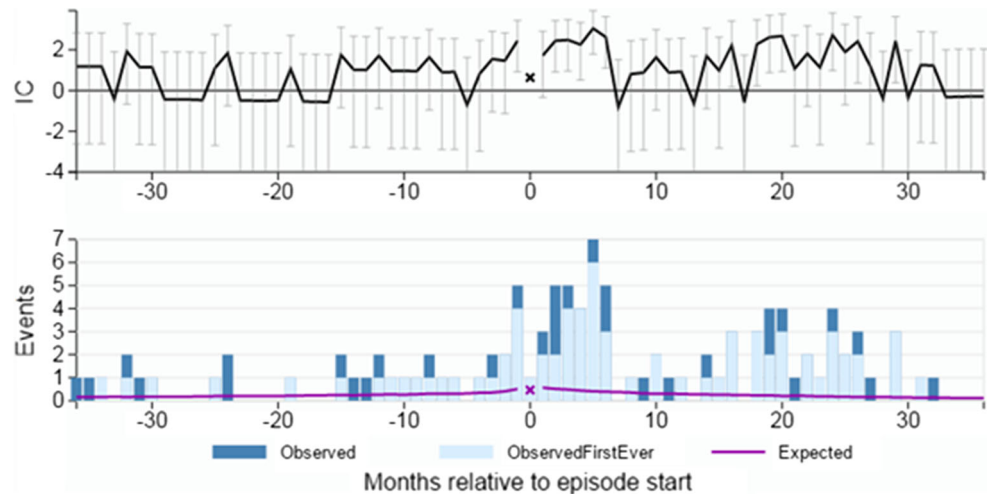
### 3.3.1 Donepezil and Salivary Hypersecretion

In the CPRD, approximately 26,000 episodes of prescribing with donepezil could be identified by the CVW Longitudinal Module platform. Patients prescribed donepezil were more commonly female (66.4%) and were predominantly older than 70 years, which is reflective of the population with Alzheimer's. There was a steady increase in the number of prescriptions identified over the data period covered by the CPRD data. Salivary hypersecretion was identified using the "1925.00: Excessive salivation" Read code with approximately 5150 patients with this Read code identified. The number of cases of excessive salivation also increased with age after an early peak in young children. A total of 300 cases were identified in children aged 3–4 years. This was followed by a low of 15 cases in children aged 8 years of age before a steady increase to a peak of 145 cases in adults aged 80 years.

In the first month following the first prescription in an episode of donepezil, three patients were identified as having a record of excessive salivation, two of which reflected a first recorded diagnosis of this event for that

patient (IC 1.71, 95% CI – 0.34 to 2.91). However, the Chronograph and IC showed a significantly higher than expected incidence of excessive salivation in each of the 2–6 months following the first-in-episode prescription for donepezil based on 26 records of excessive salivation in 23 patients identified in that time period (Fig. 2). There was also a particular increase in the identification of new events in that period (eight in the 1–6 months following the start of the episode compared with two in the 6 months preceding the episode). Further exploration of the individual patient-level data indicated that fewer than five patients with a diagnosis within 6 months appeared to have potentially stopped treatment before their recorded diagnosis based on their prescription records and their estimated episode of treatment. However, all of the remaining patients appeared to be on treatment at the time of the recorded event and, following the diagnosis, had a further prescription within the same episode, suggesting that perhaps the adverse event alone was not serious enough to outweigh the benefits of treatment, was thought to be temporary, or was not considered associated with treatment. Given this, there is likely to be some under-recording

**Fig. 2** Chronograph\* for donepezil and excessive salivation (Read code: 1925.00). \*The top panel indicates the logarithm of the observed-to-expected ratio (shrunk towards the baseline of one when the observed or expected counts are low) denoted *IC* = information component, with associated 95% confidence intervals. The bottom panel displays the underlying absolute observed and expected counts



of the event and so it is unclear if the data can be reliably used to estimate absolute risks associated with exposure.

There was also a raised IC in the month prior to the first prescription (IC 2.45, 95% CI 0.92–3.44), which may show events associated with an initial prescription of donepezil made in secondary care before prescribing is transferred to general practice [25]. It is recognised that the occurrence of the event may also reflect a worsening in the underlying dementia leading to the prescribing of donepezil, or the recording of it may be due to an increase in the recording of milder events in conjunction with an increased engagement with the GP around the time of first prescription. However, the magnitude of the increase in close temporal association with the start of treatment is striking.

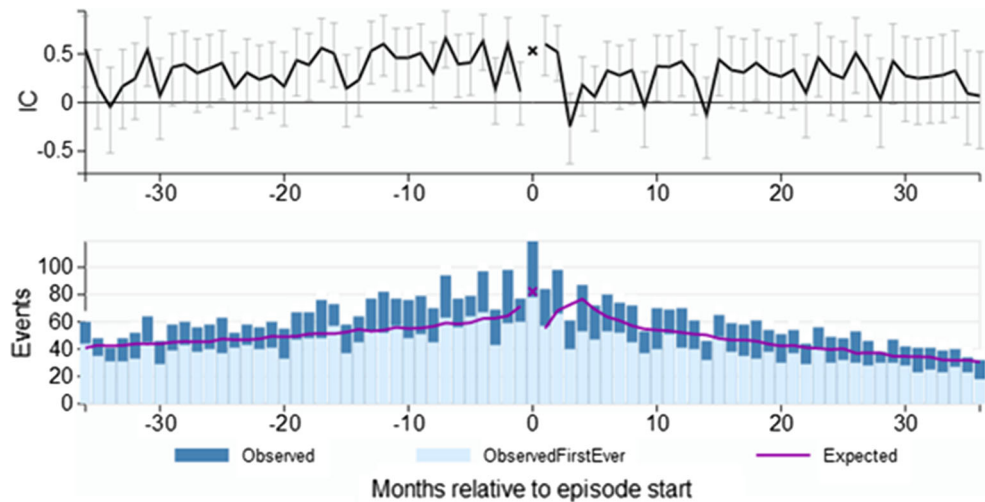
Using CPRD allowed for the identification of more events temporally associated with a first prescription for donepezil than through spontaneous reporting alone, although given the nature of the event it is likely that milder cases are not captured. There was some difficulty in identifying the correct Read code to use, although the main alternative “1926.00: Dribbling from the mouth” also showed a particularly raised IC in the month following the first prescription based on six new cases identified in that time frame with no increase seen before treatment. However, it is not clear if the dribbling observed is the result of excessive salivation or due to muscle weakness in the mouth or jaw. Overall, it was concluded that the data from CPRD, particularly the increased recording rate of first events observed in the 6 months following the start of a new prescribing episode, supports the regulatory action taken by the SMRM to request a variation to the licence for donepezil to include the risk of salivary hypersecretion as an identified adverse event.

### 3.3.2 Mirtazapine and Alopecia

Over 200,000 prescribing episodes of mirtazapine were identified in the CPRD using the CVW Longitudinal Module platform. Women were slightly more likely to be prescribed mirtazapine than men (59.6% of treated patients), with treatment almost exclusively in adults aged 18+ years (99.5%), with a declining absolute number of prescriptions in older patients generally reflective of the overall adult population in CPRD, although there was a small increase in the number of prescriptions identified in patients aged around 80 years. There has been an increase in the rate of prescribing in recent years with nearly 26,000 treatment episodes identified in approximately 5 million patients in 2013 compared with around 18,000 in 2009. The Read code used to search for cases of alopecia (M240.00: Alopecia) covers reports of hair loss, alopecia, alopecia areata, thinning of hair, baldness and alopecia unspecified as sub-codes and was considered, following review, to comprehensively identify cases. Alopecia was also widely recorded in the CPRD with ~150,000 patients with a record of the chosen Read code identified. The distribution of ages of patients with a record of alopecia is also widely reflective of the overall CPRD population, although following lower rates in children there is an increase in the rate of diagnosis in patients from their teenage years to their mid-20 s, which may reflect hormonal changes or periods of greater stress.

There were 84 records of alopecia observed in the month following a first-in-episode prescription of mirtazapine and a further 364 in the 2–6 months following. The chronograph and associated observed and expected number of events are shown in Fig. 3. While a significant excess of cases of alopecia was observed in the two months following the first prescription of mirtazapine compared with after the first prescription of other drugs, this was also seen

**Fig. 3** Chronograph for mirtazapine and alopecia (Read code: M24). \*The top panel indicates the logarithm of the observed-to-expected ratio (shrunk towards the baseline of one when the observed or expected counts are low) denoted *IC* = information component, with associated 95% confidence intervals. The bottom panel displays the underlying absolute observed and expected counts



across 9 of the 12 months in the year leading up to the first prescription, and on the day of the first mirtazapine prescription, which may reflect the mirtazapine-treated population as emotional stress is known to be associated with hair loss. It may also reflect alopecia associated with previously prescribed antidepressants. Of those who had a record of alopecia in the 30 days following the start of mirtazapine, six (7%) had a prescription of sertraline in the preceding 30 days, for example. While there may be prescribing of mirtazapine in secondary care prior to prescribing by the GP, particularly if it is a second-line treatment, this is unlikely to be of a duration sufficient to lead to the pattern seen here if there is a true association.

In the population of patients prescribed mirtazapine, there was no indication of an increase in the incidence of alopecia in the first month or the 2–6 months following first prescription compared with before treatment, as demonstrated by the *ICA* statistics (0.07, lower 95% CI bound  $-0.255$  and  $-0.373$ , lower 95% CI bound  $-0.525$ , respectively). It was concluded that the data from CPRD did not provide evidence of an association between mirtazapine and alopecia, which was in line with the decision taken by the SMRM at the time of the signal. However, the original decision was taken based on a small number of cases and the data from the CPRD provides valuable supporting evidence to increase confidence in the decision taken.

### 3.3.3 Finasteride and Cognitive Disorder

The CPRD data confirm that finasteride is prescribed almost exclusively to men (99.4%) aged  $> 50$  years in line with the primary indication. Approximately 67,000 prescription episodes were identified. It is contraindicated for use in women, children and adolescents. However, presumably mainly due to its indication for use in treating

male pattern hair loss, a low number of prescriptions are also seen in younger patients aged 20+ years. A steady increase in the number of episodes of treatment identified as starting per year is seen in the CPRD since 2000 with approximately 6000 new episodes of finasteride use identified in nearly 2.5 million men in the CPRD data specifically in 2014. It was very difficult to identify cognitive impairment using the level of Read code available. Two main options were identified, which were recorded with a similar frequency: “28E.00: Cognitive decline” and “Z7C1.00: Impaired cognition”. These were identified in  $\sim 5500$  and 7000 patients, respectively. The background distribution by gender and age for each of these codes were largely the same, that is, more common in women than men ( $\sim 60\%$ ) and first recorded mostly in older patients aged 70+ years. As previously discussed, the CVW Longitudinal Module operated all analyses over a single Read code, excluding a number of pre-defined custom groups for adverse events of special interest, so the two codes identified were considered separately.

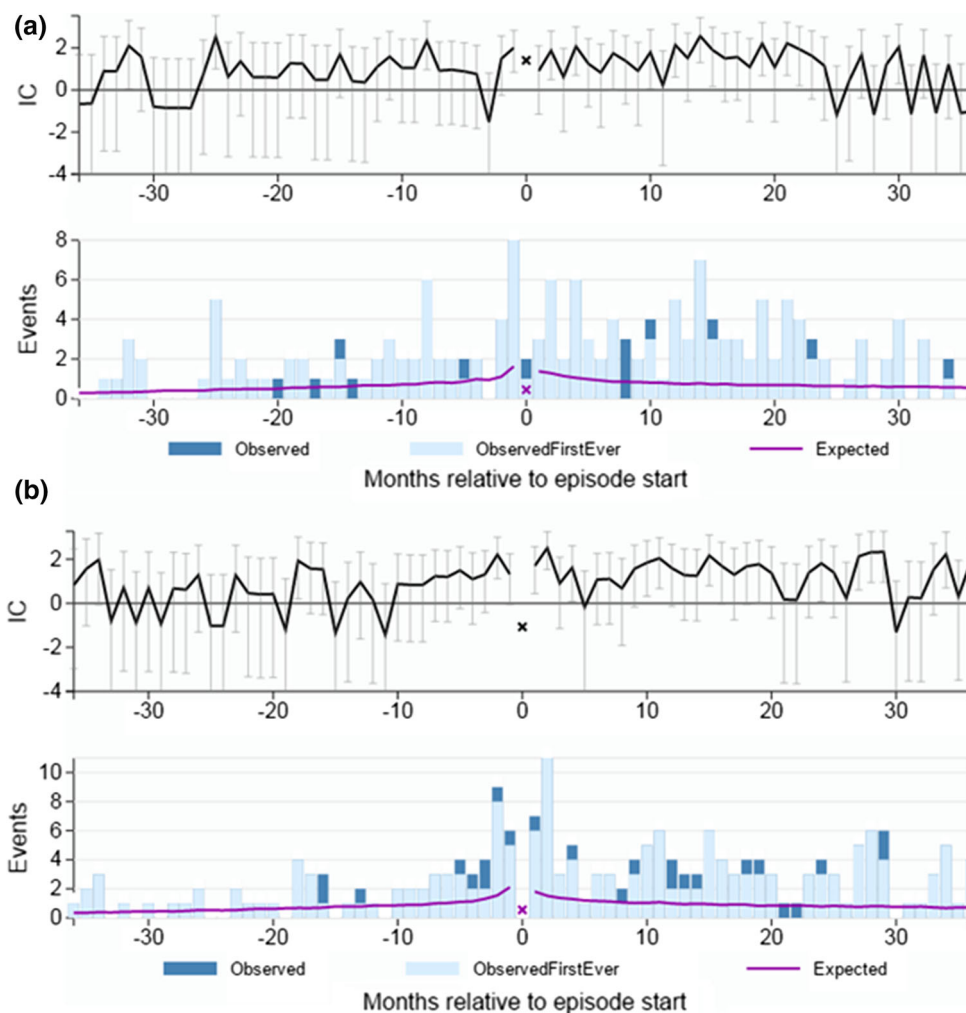
There were 22 cases of cognitive decline recorded within 6 months of a first-in-episode prescription of finasteride. For impaired cognition, 29 cases were identified. The *IC* statistics, as seen in the chronograph in Fig. 4a, b, suggest that there is some excess in recording of both codes following a prescription of finasteride compared with other drug substances, although given the small number of cases identified this is not consistently statistically significant. There is also some evidence of an increased recording of both codes in the 1–2 months before the first prescription, although again the absolute number of cases is small.

As discussed, there was no single Read code that could be simply used given the restrictions on the level of coding available. Further, other potential Read codes were identified including Eu05700: Mild cognitive disorder. Given



**Fig. 4 a** Chronograph for finasteride and cognitive decline (Read code: 28E.00). \*The top panel indicates the logarithm of the observed-to-expected ratio (shrunk towards the baseline of one when the observed or expected counts are low) denoted  $IC$  = information component, with associated 95% confidence intervals. The bottom panel displays the underlying absolute observed and expected counts.

**b** Chronograph for finasteride and impaired cognition (Read code: Z7C1.00). \*The top panel indicates the logarithm of the observed-to-expected ratio (shrunk towards the baseline of one when the observed or expected counts are low) denoted  $IC$  = information component, with associated 95% confidence intervals. The bottom panel displays the underlying absolute observed and expected counts



this difficulty in robustly identifying cases of cognitive decline, the strength of the signal is difficult to interpret using these data. While 12/13 of the case reports received via the Yellow Card scheme describe cases with onset within 1 month of starting finasteride, we expect an increased level of spontaneous reporting for events that occur within a very short period after starting a new treatment, with cases that develop more gradually less likely to be reported. However, the CPRD data as presented do provide some suggestion that there is an association between the prescribing of finasteride and cognitive disorder, albeit very weak and not at all sufficient to explore if any association is causal or coincidental given the correlation between the demographics of the patients prescribed finasteride and those at highest background risk of the event, or if it is caused by biases such as increased recording due to increased engagement with healthcare services. If it was possible to better define cases then the data from CPRD may be of more value in placing any further spontaneous reports into context although, given the complexity of the event and its diagnosis and potential

progression, EHR data are likely to be of limited value in completely supporting any regulatory decision making, even with more in-depth analyses. However, the data should be considered alongside other sources of evidence when assessing the potential risk.

#### 4 Discussion

The methods used in this proof-of concept study have been explored previously although those studies identified that further research was needed to position the use of EHRs in routine signal detection and assessment [14, 22, 23]. The data from this study provide additional evidence on the value of routinely using CPRD data specifically to strengthen the early assessment of signals arising from spontaneous reporting and decision making within a UK regulatory environment and within current signal management processes. The review conducted in this study found a higher rate of the potential use of the CPRD (for 32 of 83 signals) than that suggested in a previous study,

which was able to evaluate 112 drug–event combinations using EHR data from UK THIN from 458 raised via disproportionality analyses in the WHO VigiBase database of individual case safety reports after excluding combination products [14]. This earlier study excluded DEC where < 100 first prescriptions could be identified whereas we excluded drugs that we expected would not be reliably captured in the CPRD based on pre-existing knowledge of the setting in which they are prescribed or accessed, and hence may have included DECs with limited EHR data. We took this approach in order to increase our confidence that exposure was fully and robustly captured within the EHR data used without excluding newer products, or those with limited indications for use.

The lower number of DECs reviewed in our study compared with the above-referenced study by Star et al. [14], despite a longer study period, is principally due to the stage at which they were identified. In our study, we only investigated DECs identified as signals that are a subset of those disproportionately reported. This was considered the most feasible stage within internal processes at which to include use of the platform as use earlier in the signal management process would have resulted in higher resource requirements and a likely increase in the number of false positive signals explored using the EHR data.

The value of the chronograph methodology for detecting and strengthening signals within EHR data has been demonstrated through a number of case studies previously and was reinforced in the three case studies presented here [14, 23]. We additionally found that having immediate access to data to place a signal into context was beneficial for timely signal assessment. Further, the opportunity to identify additional cases within the EHR enabled us to understand a wider range of patient experiences than those that are reported through spontaneous reporting systems. While the case studies did not suggest that, had the CPRD data been available in such a way at the time of the original assessment of the signal at the MHRA, the decisions would have substantially changed, it did suggest that it would have complemented the assessment of individual case reports and increased the robustness of the decision-making process.

Previous research into the use of EHR within signal detection and strengthening has raised the limitations of EHR data and the impact of these limitations [14–16, 22–24, 26]. However, there are several known limitations to the data in the CPRD, particularly when only considering the primary care database, that impacted on its value for routine signal strengthening within the MHRA in this study. Thirty percent (21/69) of new signals identified within this study related to drug substances prescribed almost exclusively in secondary care or widely available without a prescription. There is a paucity of individual

patient-level electronic data on the prescribing and use of medicines in the UK outside of primary care. This also impacts on those products where the decision to start treatment is made in secondary care with care then transferred to primary care at a later stage, as the start of treatment cannot be precisely estimated. This is particularly seen in the first case study on donepezil.

Data from databases such as the CPRD are not collected for research purposes and therefore coding may not be robust or consistent enough to accurately identify all adverse events, and the use of different coding systems (Read codes in primary care vs MedDRA<sup>®</sup> for adverse event reporting in pharmacovigilance databases) means that mapping between systems could be needed. Further, the complexity of the Read code terminology used in the CPRD means that for many events a single term is not adequate for identification. However, if there is flexibility in the approach taken to identify events within the analysis platform with custom groups of terms used for adverse events of special interest, which were already available for 16% of signals where both exposure and event could be identified in the CPRD, then the impact of inconsistencies in the recording of events is reduced.

Under-ascertainment of cases that are diagnosed in secondary care, which do not present to a healthcare professional, or that are not recorded in the EHR, and over-ascertainment where suspected but unconfirmed diagnoses are recorded in primary care are inherent limitations of EHR such as the CPRD and also need to be considered. The three case studies discussed here all relate to events likely to present to primary care, although under-recording may still pose a risk. Incorporating other healthcare databases such as the Hospital Episode Statistics data, which is routinely linked to the CPRD primary care data, could improve the identification of events for other signals.

These limitations of EHR highlight the need for a detailed understanding of the use of medicines in clinical practice, the diagnosis of outcomes, and the mechanisms of data capture in order to interpret the output of routine analyses coming from such an automated analysis platform.

There are also limitations of the CVW Longitudinal Module version, as used in this study, that if addressed would increase the proportion of safety signals that could be explored using it. It has already been highlighted that, due to the complexity of the Read code terminology, adequately identifying events using a single code can be difficult; therefore, increasing the number of pre-defined custom groups of terms within the Module for identifying events will increase the proportion of signals that can be robustly explored. Further, allowing flexibility within the analysis platform in defining exposure by drug formulation or dose, although this data is not always fully captured in

CPRD, or in a subset of the population (12% of signals identified in this study), would also be beneficial as would allowing analyses looking at potential drug interactions. If cases cannot be robustly identified, then the CPRD may still provide valuable data on the treated and untreated populations or vice versa.

In order to be of full value within a regulatory environment, this approach would need to be embedded in routine signal management practices. Further work is required to establish how this should operate efficiently and consistently, and in line with GVP, while still maximising the scientific value of the data. In addition, while it is clear that even in its current form, the CVW Longitudinal Module facilitates analyses that would not be routinely feasible without such a tool, further improvements to it would also be highly beneficial in increasing the proportion of signals for which it could be used and also for increasing the strength of the analyses for individual signals. There may also be further scope for using CPRD through the CVW Longitudinal Module for active monitoring of selected adverse events of special interest, for assessing changes in risk over time, for supporting communications and for monitoring the impact of regulatory actions, although these have not been considered here.

## 5 Conclusion

Regulators are responsible for ensuring that accurate product information are available to support clinical decision making by healthcare professionals and patients, and this should be based on the most reliable evidence available. Routine access to relevant analyses of data from the CPRD would be valuable for supporting the management of a large proportion of safety signals and increasing the robustness and timeliness of their assessment within a regulatory environment. A further extended pilot of this approach is planned to further identify and implement improvements to the analysis platform and to optimally integrate it into signal management processes.

**Acknowledgements** The authors would like to thank Olutoyin Agbaje, Claire Davies, Temitope Ibitoye, and Jenny Wong for the comments they provided within early stages of the pilot and on the output from the three case studies, and Arlene Gallagher for providing additional expertise on the use of CPRD data.

**Author Contributions** The study was planned and led by KD, PT and RO. Preliminary screening of the DECs included was conducted by PT, RO and BB, with additional assessment support from KD, HB and AS for the second stage of the review. The three case studies were conducted by HB, BB and AS with input from KD. The paper was drafted by KD with contributions from all authors.

## Compliance with Ethical Standards

**Funding** This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors. Access to the CVW Longitudinal Module was provided by Commonwealth Informatics Inc. The CVW Longitudinal software is engineered by Commonwealth Informatics, Inc., and builds on research performed by the Uppsala Monitoring Centre.

**Conflict of interest** At the time of the study, all authors (Katherine Donegan, Rebecca Owen, Helena Bird, Brian Burch, Alex Smith, and Phil Tregunno) were employed by the MHRA, which is an Executive Agency of the Department of Health, but have no conflicts of interest that are directly relevant to the contents of this study. The MHRA has statutory responsibility to monitor the safety of medicinal products on the UK market, and the study was undertaken independently of the Department of Health.

## References

1. Foy M, Barrow P, Raine JM. Spontaneous reporting: United Kingdom. In: Andrews EB, Moore N, editors. *Mann's pharmacovigilance*. Oxford: Wiley; 2014.
2. Dal Pan GJ, Lindquist M, Gelperin K. Postmarketing spontaneous pharmacovigilance reporting systems. In: *Pharmacoepidemiology*. Oxford: Wiley; 2012.
3. Bate A, Evans SJW. Quantitative signal detection using spontaneous ADR reporting. *Pharmacoepidemiol Drug Saf*. 2009;18:427–36.
4. Candore G, Juhlin K, Manlik K, Thakrar B, Quarcoo N, Seabroke S, et al. Comparison of statistical signal detection methods within and across spontaneous reporting databases. *Drug Saf*. 2015;38:577–87.
5. Hazell L, Shakir SA. Under-reporting of adverse drug reactions: a systematic review. *Drug Saf*. 2006;29(5):385–96.
6. Pariente A, Gregoire F, Fourrier-Reglat A, Haramburu F, Moore N. Impact of safety alerts on measures of disproportionality in spontaneous reporting databases: the notoriety bias. *Drug Saf*. 2007;30(10):891–8.
7. Lopez-Gonzalez E, Herdeiro MT, Figueiras A. Determinants of under-reporting of adverse drug reactions: a systematic review. *Drug Saf*. 2009;32:19–31.
8. Ogdie A, Langan SM, Parkinson J, Dattani H, Kostev K, Gelfrand JM. Medical record databases. In: *Pharmacoepidemiology*. Wiley; 2012.
9. Smeeth L, Cook C, Fombonne E, Heavey L, Rodrigues LC, Smith PG, Hall AJ. MMR vaccination and persuasive developmental disorders: a case-control study. *Lancet*. 2004;365:963–9.
10. Oteri A, Mazzaglia G, Pecchioli S, Molokhia M, Ulrichson SP, Pedersen L, et al. Prescribing pattern of antipsychotic drugs during the years 1996–2010: a population-based database study in Europe with a focus on torsadogenic drugs. *Br J Clin Pharmacol*. 2016;82:487–97.
11. Douglas I, Evans S, Rawlins MD, Smeeth L, Tabrizi SJ, Wexler NS. Juvenile Huntington's disease: a population-based study using the General Practice Research Database. *BMJ Open*. 2013;3:e002085.
12. Donegan K, Beau-Lejdstrom R, King B, Seabroke S, Thomson A, Bryan P. Bivalent human papillomavirus vaccine and the risk of fatigue syndromes in girls in the UK. *Vaccine*. 2013;31:4961–7.
13. Mahaux O, Bauchau V, Van Holle L. Pharmacoepidemiological considerations in observed-to-expected analyses for vaccines. *Pharmacoepidemiol Drug Saf*. 2016;25:215–22.

14. Star K, Watson S, Sandberg L, Johansson J, Edwards IR. Longitudinal medical records as a complement to routine drug safety signal analysis. *Pharmacoepidemiol Drug Saf.* 2015;24:486–94.
15. Ball R, Robb M, Anderson SA, Dal Pan G. The FDA's Sentinel initiative—a comprehensive approach to medical product surveillance. *Clin Pharmacol Ther.* 2016;99:1532–5.
16. Pacurariu AC, Straus SM, Trifirò G, Schuemie MJ, Gini R, Herings R, et al. Useful interplay between spontaneous ADR reports and electronic healthcare records in signal detection. *Drug Saf.* 2015;38:1201–10.
17. Guideline on good pharmacovigilance practices (GVP) Annex 1—Definitions (Rev 4). European Medicines Agency, 2017. Available at: [http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Scientific\\_guideline/2013/05/WC500143294.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2013/05/WC500143294.pdf). Accessed 8 Jan 2018.
18. Guideline on good pharmacovigilance practices (GVP) Module IX—Signal Management (Rev 1). European Medicines Agency, 2017. Available at: [http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Scientific\\_guideline/2017/10/WC500236408.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2017/10/WC500236408.pdf). Accessed 8 Jan 2018.
19. A guideline on summary of product characteristics (SmPC). European Commission, September 2009. Available at: [https://ec.europa.eu/health/sites/health/files/files/eudralex/vol-2/c/smpc\\_guideline\\_rev2\\_en.pdf](https://ec.europa.eu/health/sites/health/files/files/eudralex/vol-2/c/smpc_guideline_rev2_en.pdf). Accessed 8 Jan 2018.
20. Herret E, Gallagher AM, Bhaskeran K, Forbes H, Mathur R, van Staa T, Smeeth L. Data resource profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol.* 2015;44:827–36.
21. ISAC protocol 15\_217: Descriptive characteristics of the Clinical Practice Research Datalink (CPRD) EMIS database. Available at: [https://www.cprd.com/isac/Protocol\\_15\\_217.asp](https://www.cprd.com/isac/Protocol_15_217.asp). Accessed 24 July 2017.
22. Norén N, Hopstadius J, Bate A, Star K, Edwards IR. Temporal pattern discovery in longitudinal electronic patient records. *Data Min Knowl Disc.* 2010;20:361–87.
23. Cederholm S, Hill G, Asiimwe A, Bate A, Bhayat F, Persson Brobert G, et al. Structured assessment for prospective identification of safety signals in electronic medical records: evaluation in The Health Improvement Network. *Drug Saf.* 2015;38:87–100.
24. Trifirò G, Pariente A, Coloma PM, Kors JA, Polimeni G, Miremont-Salame G, et al. Data mining on electronic health record databases for signal detection in pharmacovigilance: which events to monitor? *Pharmacoepidemiol Drug Saf.* 2009;18:1176–84.
25. National Institute for Clinical Excellence, Social Care Institute for Excellence. *Dementia: Supporting People with Dementia and Their Carers in Health and Social Care.* NICE/SCIE, 2006.
26. Bate A, Brown EG, Goldman SA, Hauben M. Terminological challenges in safety surveillance. *Drug Saf.* 2012;35:79–84.