Implementing near real-time vaccine safety surveillance using the Clinical Practice Research Datalink (CPRD)

Andreia Leite [a,*], Sara L. Thomas [a], Nick J. Andrews [b]

[a] Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London WC1E 7HT, United Kingdom
[b] Statistics, Modelling and Economics Department, Public Health England, London NW9 5 EQ, United Kingdom

ABSTRACT

Introduction: Near real-time vaccine safety surveillance (NRTVSS) using electronic health records is increasingly used to rapidly detect vaccine safety signals. NRTVSS has not been fully implemented in the UK. We assessed the feasibility of implementing this surveillance using the UK Clinical Practice Research Datalink (CPRD).
Methods: We selected seasonal influenza vaccine/Guillain-Barré Syndrome (GBS) as an example of a rare outcome and measles-mumps-rubella (MMR) vaccine/febrile seizures as a positive control. For influenza/GBS we implemented a system for the 2013/2014 and 2014/2015 influenza seasons; for MMR/seizures the surveillance period was July 2014–June 2015. We used the continuous Poisson-based maximized sequential probability ratio test (PMaxSPRT), comparing observed-to-expected events, for both pairs. We calculated an age-sex-adjusted rate using 5 years of historic data and used this rate to calculate the expected number of events in pre-specified post-vaccination risk-window (GBS: 0–42 days, seizures: 6–21 days). For MMR/seizures we also implemented the system using the Binominal-based maximized sequential probability ratio test (BMaxSPRT). For this, we compared seizures in the risk-window (6–21 days) to a control window (0–5 and 22–32 days). Delays in recording outcomes influence the data available, so we adjusted the expected number of events using a historical distribution of delays in recording GBS/febrile seizures. Analyses were run using data up to each CPRD monthly release. We also performed power calculations for detecting increases in relative risk (RR) from 1.5 to 10.
Results: For influenza/GBS we implemented a system in both seasons with no signal. Power to detect a signal was >80% for RR ≥ 4. For MMR/seizures we were able to identify a signal with PMaxSPRT but not with BMaxSPRT. Power ≥ 80% for RR ≥ 2.5 for both tests.
Conclusion: CPRD is a potential data source to implement NRTVSS to exclude large increases in the risk of rare outcomes after seasonal influenza and lower increases in risk for more frequent outcomes.
© 2017 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Near real-time vaccine safety surveillance (NRTVSS) using electronic health records is amongst the tools available to perform post-licensure vaccine safety surveillance. NRTVSS is usually started shortly after a new vaccine is introduced and data is analysed at repeated points in time. Near real-time surveillance was introduced in the USA in 2005 first using the sequential probability ratio test and later its maximized version. It is now used routinely in this country [1]. It has allowed the identification of several safety signals [2].

In the UK, there are electronic health records available such as the Clinical Practice Research Datalink (CPRD). NRTVSS has been implemented in the UK using spontaneous reports to obtain the observed number of events and CPRD to calculate the expected number of events. This implementation inherits spontaneous reports limitations, including underreporting [3]. A NRTVSS fully relying on electronic health records has not been implemented to date.

When envisaging a new data source to implement NRTVSS timeliness is a key consideration. In CPRD, delays can happen due to: (i) delays in making a diagnosis after an initial consultation; (ii) delays in recording diagnosis made in other levels of care (e.g. hospital); (iii) delays in receiving data for analysis. To the best

of our knowledge, there has been no work to explore systematically the influence of (i) on recording patterns using CPRD data. For (ii), a previous analysis of CPRD data looking at conditions of interest for vaccine safety has shown that recording delays exist, but the majority of records accrue within a month after the date of the event [4]. Researchers receive CPRD data on a monthly basis (delay (iii)). Thus, from the evidence to date CPRD is a potential source of data to implement NRTVSS.

In addition to delays, several questions regarding the actual implementation of a system using CPRD data remain unaddressed, such as which statistical method to use, how to account for delays, and whether there is enough power to identify safety signals. To address these we sought to trial the implementation of NRTVSS using previously collected CPRD data.

2. Methods

2.1. Data source

We used data from CPRD, a primary health care database with anonymised health records from a broadly representative sample (~6.9%) of the UK population. CPRD includes information on demographics, coded diagnosis, therapies, vaccines, health-related behaviours, and referrals to secondary care [5]. Diagnoses recorded in CPRD include diagnoses made both in primary care and in hospital. Hospital diagnoses are fed back to GPs via letter, which are later coded in the system. Diagnoses are coded using Read-codes, a hierarchical thesaurus of clinical terms used in the UK since 1985 [6].

CPRD contains several relevant dates. For each patient there is information on the patient's current registration with the practice (crd) and the patient left the practice (tod). Each record contains the date when the record was entered into the system (system date) and the date deemed to represent when the event registered took place (event date). At the practice-level, CPRD includes the date when the practice met certain recording quality criteria (up-to-standard date, uts) and the date when data were last collected from the practice before each monthly release (last collection date, lcd) [5].

2.2. Vaccine/outcome pairs

We selected two pairs: seasonal influenza vaccine/Guillain-Barré syndrome (GBS) and Measles-Mumps-Rubella (MMR) vaccine/febrile seizures. NRTVSS is of particular relevance to assess seasonal influenza vaccine due to the short time available for action, and GBS is a rare outcome of interest following influenza vaccine. Influenza vaccine/GBS was thus chosen to assess the potential of CPRD as a data source to implement NRTVSS for rare events. Febrile seizures are a known adverse reaction seen after MMR vaccine, so we selected this pair to represent a positive control and as an example of a somewhat less rare event with a childhood vaccine [7]. Appendix A presents code-lists used to identify GBS/seizures and Appendix B the algorithms used to identify vaccinated individuals.

3. Analysis

3.1. Statistical tests

Choice of the statistical test to use should be guided by the test characteristics (e.g. power and underlying assumptions), frequency of data updates and frequency of the outcome under study. One approach is to select first the general group of tests (continuous or group sequential) and then choose a specific version of the test [2]. For continuous tests, data are looked at as often as desired, and ideally when a new event is observed, while for group sequential tests data are interrogated at discrete points in time [2]. Previous work has shown that continuous sequential tests perform better than group sequential [8] and aggregate data (weekly or monthly) can be used in a continuous way [9]. As CPRD is updated monthly, we considered continuous sequential tests more appropriate.

Poisson-based Maximized Sequential Probability Ratio Test (PMaxSPRT), the Binomial-based Maximized Sequential Probability Ratio Test (BMaxSPRT), and the Conditional Maximized Sequential Probability Ratio Test (CMaxSPRT) are the continuous sequential tests available. PMaxSPRT involves a comparison observed-to-expected and its use has been proposed when less than 50 events are expected, as it is a more powerful test [2]. Disadvantages include limited ability to adjust for confounders and potential bias by secular or coding trends, as it relies on historical data. It also does not allow for uncertainty in the expected count (it is taken as a fixed expected number). BMaxSPRT compares the number of events in exposed-to-unexposed individuals or in periods within individuals. This allows further adjustment for potential confounders but lessens power. Unlike PMaxSPRT, CMaxSPRT was designed to account for uncertainty in the historical data. The comparison is made in terms of the cumulative person-time it took to observe a certain number of adverse events in the historical and surveillance data. It assumes event rates are constant in both versions of the data.

Given the rarity of GBS we selected PMaxSPRT for influenza vaccine/GBS. For seizures/MMR the number of expected events was still lower than 50 (see below), suggesting the use of PMaxSPRT. However, previous works have also considered the simultaneous use of PMaxSPRT and BMaxSPRT owing to their complementary strengths [9]. We preferred this approach as it allowed us to further identify challenges/potential solutions when using CPRD to perform NRTVSS. It has been previously suggested that PMaxSPRT gives biased results when a small sample is used to estimate the number of expected events [10]. To avoid this, we used a long historical period (5 years) to obtain more stable estimates and thus reduce uncertainty to negligible levels relative to uncertainty in the observed data. It has also been suggested as an *ad hoc* guideline that an alternative test (CMaxSPRT) should be used when the number of observed events in the historical data is less than five times the number of expected events in the surveillance data. We thus assessed whether this *ad hoc* rule held in our data.

Below we detail how we obtained the observed and expected numbers of events to implement PMaxSPRT for each pair. We start with an explanation for seasonal influenza/GBS followed by MMR/seizures. For the latter we emphasize aspects that differ from the first pair. For BMaxSPRT we used a case-only design and compared the number of cases during the exposed-to-unexposed periods, also detailed below. Analyses were performed using R package Sequential 2.3.1 [11].

3.2. Influenza/GBS

We studied the 2013/14 and 2014/15 seasons (1st September–31st March), using data released in July 2015 and 2016, respectively. Using these data releases allowed at least a year from the event date for them to be recorded. In all analyses we did not consider the small proportion of events that are recorded with a delay >1 year [4].

3.2.1. Historical rates, expected and observed number of events (PMaxSPRT)

For the historical comparison, we used the general background rate of GBS among individuals aged ≥65 years as this is the age in which seasonal influenza vaccine is routinely recommended and given in GP practices. For each study season, we calculated GBS

historical rates stratified by age (65–74, 75–84, ≥85 years old) and gender for the 5 previous seasons (2008/09–2012/13 and 2009/10–2013/14, respectively). Numerators were first-ever GBS cases for active patients. We have previously demonstrated that when GP systems are updated the system date (the date a record is added to a patient's file, assigned automatically by the general practice software) of some records can be altered to a later date [4]. For those records, it is not possible to estimate accurately the delay in recording the outcome. Hence, these records were identified using the approach proposed in [4] and were excluded. Active patients were defined as contributing follow-up time during each season. Start of follow-up was the latest of uts, crd (plus 1 year to exclude retrospective recording of previous diagnoses when registering with a new practice [12]), or 1st September 2008–13. End of follow-up was the earliest of date of tod, lcd, or 31st March 2009–14. We averaged seasonal GBS rates over the five historical seasons and applied this rate to the vaccine-exposed follow-up time in the study seasons, to obtain an expected number of events (adjusted by age and gender). For the study seasons end of follow-up was the earliest of tod, lcd or 42 days after vaccination (Appendix C) [13]. The observed number of events was the total number recorded in the vaccine risk-window at the time of each analysis.

3.2.2. Delays

For each patient we calculated the time between the risk-window midpoint and lcd (time = d) and then used the previously generated delay distribution [4] to calculate the probability (r_d) that an event that did occur within a year would be recorded by delay d. This was used to adjust the follow-up to obtain an adjusted follow-up. For example, if a patient had 30 days between the risk-window mid-point and lcd and r_d = 75%, then only 75% of this patient follow-up time was considered (Fig. 2 in Appendix D). We assumed no delays in vaccination data.

3.3. MMR/Seizures

The system was implemented for one year (July 2014–June 2015) using data released in July 2016.

3.3.1. Historical rates, expected and observed number of events (PMaxSPRT)

We calculated febrile seizures rates during the second year of life (12–23 months, timing of MMR 1st dose [13]), stratified by age (two weeks periods) and gender, for the five years previous to the study period (July 2009–June 2014). We first identified all febrile seizures events for eligible patients and excluded records likely to be duplicated (Appendix E).

We calculated follow-up time and the expected and observed number of events as described above (Appendix C) for the historical period July 2009–June 2014 and the study period July 2014–June 2015. A previous study looking at the risk of febrile seizures following MMR and using hospital data identified a risk-window of 6–11 days [7]. In this work, we used primary care data, which are likely to capture febrile seizures with some delay. This can happen if parents seek care outside primary care (e.g. emergency services) and GPs only receive and register the information regarding the seizure a few days after it has occurred. We thus allowed extra time, by using a risk window of 6–21 days to capture such events.

3.3.2. BMaxSPRT

To apply BMaxSPRT we used the same risk-window of 6–21 days post-vaccination and used a control period of 1–5 (c_1) and 22–32 (c_2) days post-vaccination, selected to be a period of the same length and close to the risk period.

3.3.3. Delays (BMaxSPRT)

For BMaxSPRT it was necessary to adjust for delays for each of the post-vaccination periods (the risk period and c_1, c_2). This was done by calculating an adjusted follow-up period for each of these intervals as shown in Fig. 3 in Appendix D. For each individual we then calculated a ratio of the corrected follow-up for control period compared to the risk period (see Appendix D for an example) and then obtained an average ratio across individuals. This average ratio was included in the calculations for the BMaxSPRT method as the matching ratio [9]. This final adjustment simultaneously accounted for delays in practices uploading data and partially accrued period.

3.4. Implementation

To mimic a NRTVSS using pre-existing data we first recreated how data accrued. To determine whether a record of interest would have been in each data release we used: release date; lcd (practice-level); event date of the record; and system date of the record. CPRD is released on a monthly basis, on the first Monday of each month. For a particular release we considered the outcome would be captured if the event date, system date, and last collection date all happened before the date of release. For example, an event taking place (event date) on 9/10/2014, with a system date of 10/11/2014, and lcd 28/10/2014 for the November release would not appear in the November. If lcd for the December release was 25/11/2014 then the event would appear in December.

As no signal was expected for influenza/GBS we further tested NRTVSS implementation by adding cases to generate an increase in risk of approximately 4 and 5-fold, which power calculations suggested should be detectable.

Implementation was done graphically by calculating the log-likelihood ratio test at the time of each data release. For PMaxSPRT the log-likelihood is based on the number of observed and expected events while for BMaxSPRT it considers the number of observed events occurring in the control and risk periods. The results from the log-likelihood ratio test were compared with the critical limit. For each vaccine/outcome pair and study period we calculated critical limits considering a minimum number of observed events to reject the null hypothesis of 1, 2, and 4.

3.5. Power and expected time to signal

Post-licensure vaccine safety surveillance aims at detecting signals that might have been missed before vaccine approval, due to the lack of power in the analyses conducted. When considering NRTVSS we thus need to assess power. The R package Sequential includes system performance tools, allowing calculating of power and expected time to signal [11].

Power is affected by several factors: incidence of the outcome (both background incidence and incidence following vaccination), vaccine uptake, vaccine risk-window length, length of the study period, delays in receiving the data, relative risk (RR) to be detected, events in the first look at the data, minimum number of events before rejecting the null, and level of significance. Calculations were performed for a plausible range of RR (1.5–10), considered no events in the first look at the data, and a level of significance of 5% (α = 0.05). For PMaxSPRT we also required 1, 2 or 4 events before rejecting the null [14] and the remaining factors were integrated through the expected number of events at the end of the surveillance period. For BMaxSPRT we considered the total number of events at the end of the surveillance period (both from risk-window and control periods).

Expected time to signal is conditional on having identified a signal and is obtained in the units of expected number of events. As CPRD data do not accrue at a constant rate, to know at which

release we would expect a signal we evaluated at which release the number of expected events would have been achieved.

4. Results

Table 1 presents the number of doses identified and the main characteristics of individuals receiving the vaccine of interest.

4.1. Seasonal influenza/GBS

We identified 1.89 and 1.66 expected events for season 2013–14 and 2014–15, respectively. The historical rates used were based on 33 observed events for each season. Hence, the use of

CMaxSPRT was not deemed necessary. Fig. 1 presents system implementation. No signal was identified for both seasons. When we added cases to generate an increase in risk of 4- and 5-fold we found, for an increase in risk of 4, the signal would be identified at the beginning of January and February for the season 2013/14 and 2014/15, respectively, if a minimum of 4 events was stipulated. For an increase of 5 times the risk the signals would be detected a month before (Fig. 2).

Table 2 presents power and expected time to signal for seasonal influenza/GBS and both seasons. In general, there was power $\geq 80\%$ to detect RR ≥ 4. If there was a signal this would be detected at the beginning of December for large increases in risk (6–8 times) and at the beginning of January for lower increases (4–5 times).

Table 1
Main characteristics of individuals receiving the vaccine of interest for the pairs included.

Characteristic	Vaccine/outcome pair		
	Influenza/GBS season 2013–14	Influenza/GBS season 2014–15	MMR/Febrile seizures
Number of doses (n)	533,110	477,454	28,249
Sex – n (%)			
Male	240,884 (45.2)	216,224 (45.3)	14,474 (51.2)
Female	292,226 (54.8)	261,230 (54.7)	13,775 (48.8)
Age (years) – n (%)			
65–74	270,690 (50.8)	242,168 (50.7)	–[a]
75–84	188,423 (35.3)	168,160 (35.2)	–[a]
$\geq$85	73,997 (13.9)	67,126 (14.1)	–[a]
Age (months) – n (%)			
12	–[a]	–[a]	11,460 (40.6)
13	–[a]	–[a]	10,049 (35.6)
14	–[a]	–[a]	3320 (11.8)
$\geq$15	–[a]	–[a]	3420 (12.1)

GBS – Guillain-Barre syndrome.
MMR – Measles-mumps-rubella.
[a] Age (at time of vaccination) is expressed in years for seasonal influenza/GBS and months for MMR/febrile seizures.
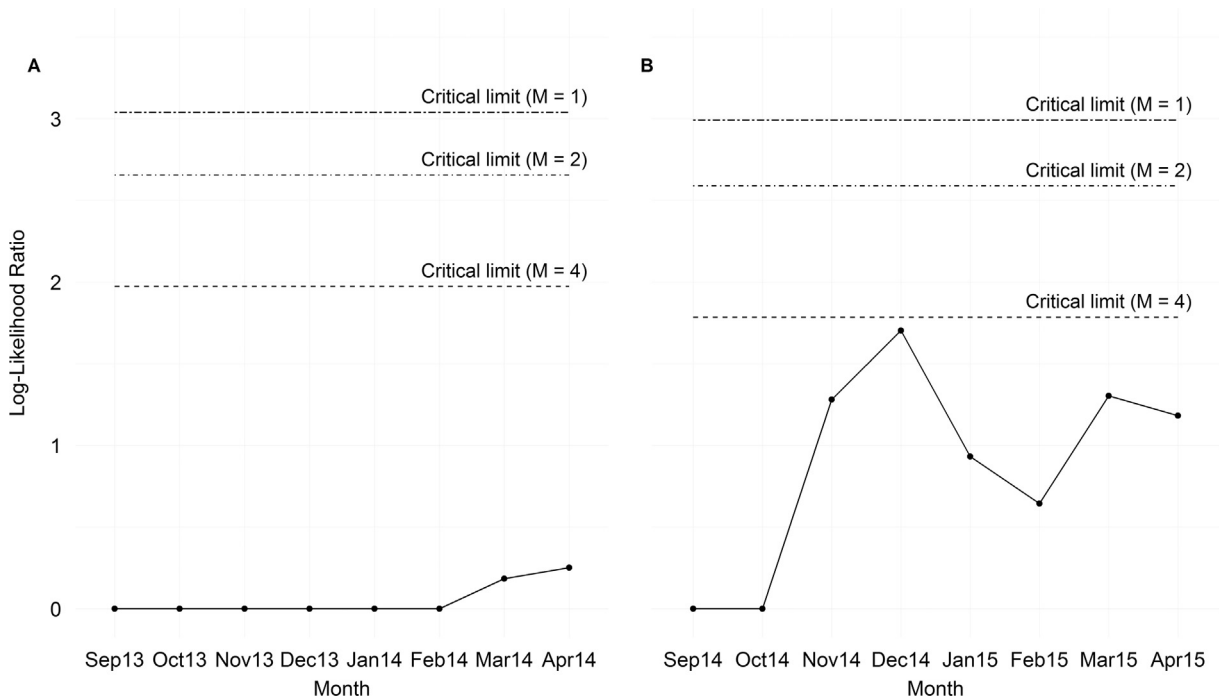

Fig. 1. Implementation of a system for influenza vaccine/GBS for season 2013–14 (A) and season 2014–15 (B). No signal is detected in any of the seasons.
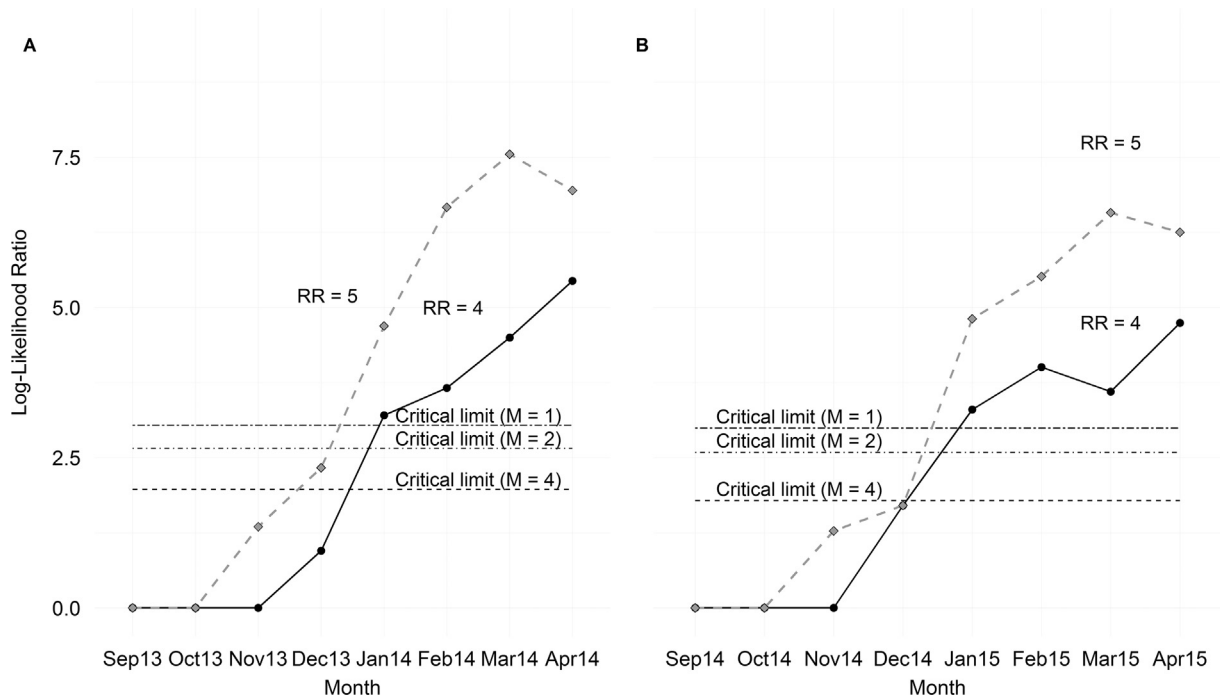
Fig. 2. Implementation of a system for influenza vaccine/GBS for season 2013–14 (A) and season 2014–15 (B), assuming an increase in risk of 4 and 5 times. A signal is detected at different points in time depending on critical limits considered.

Table 2
Power and expected time to signal for seasonal influenza/GBS (seasons 2013–14 and 2014–15) using Poisson-based Maximized Sequential Probability Ratio.

Minimum events	Season	Data available at	Power (time to signal in months from beginning of surveillance)[a] Relative risk								
			1.5	2	2.5	3	4	5	6	8	10
1	2013–14	07–04–2014	13	25	40	55 (4)	78 (4)	**91 (3)**	**97 (3)**	**100 (3)**	**100 (3)**
	2014–15	06–04–2015	12	23	37	51 (4)	74 (4)	**88 (4)**	**95 (4)**	**99 (3)**	**100 (3)**
2	2013–14	07–04–2014	14	28	44	60 (4)	**82 (4)**	**93 (3)**	**98 (3)**	**100 (3)**	**100 (3)**
	2014–15	06–04–2015	14	26	41	55 (4)	77 (4)	**90 (4)**	**96 (4)**	**100 (3)**	**100 (3)**
4	2013–14	07–04–2014	16	33	50	65 (4)	**86 (4)**	**95 (4)**	**98 (4)**	**100 (3)**	**100 (3)**
	2014–15	06–04–2015	16	31	47	62 (4)	**83 (4)**	**93 (4)**	**98 (4)**	**100 (4)**	**100 (4)**

Cells in bold refer to power ≥80%.
PMaxSPRT - Poisson-based Maximized Sequential Probability Ratio.
[a] Time to signal is only displayed for cells where equivalent power≥50%.

4.2. MMR/seizures

After investigation of duplicated records of febrile seizures we decided to exclude those occurring with three days of one another (Appendix E). We identified 11.3 expected episodes in the study period and the historical rates were based on 2693 observed events. Fig. 3 presents NRTVSS implementation. We identified a signal using PMaxSPRT. For BMaxSPRT the signal was just missed.

Table 3 presents power and expected time to signal for febrile seizures/MMR based on a one-year surveillance period. We observed power ≥ 80% to detect RR ≥ 2.5. If there was a signal this would be detected at the beginning of September (2 months after beginning of surveillance) using PMaxSPRT for RR of ≥5, and in subsequent months for lower increases in risk. Power for BMaxSPRT was lower but would still allow detection of an RR of ≥2.5.

5. Discussion

We systematically assessed the feasibility of implementing a NRTVSS using data from CPRD. Our study shows that it is feasible

to use CPRD and it would enable detection of medium/large increases in risk of GBS following seasonal influenza vaccine among individuals aged ≥65 years, and smaller increases in the risk of febrile seizures following first dose of MMR.

For influenza/GBS, CPRD would only enable detection of large increases in risk. In addition, the signal would only be detected around mid-season (beginning of January) which might be late, as the vaccine is recommended early in the season [15]. Despite limited power to detect an increased risk, our finding of no increased risk of GBS following seasonal influenza vaccine seems consistent with the existing literature. For example, a recent work assessing GBS following influenza vaccine in the USA between 2010/11 and 2013/14 found no signal for the season 2013/14, the season we also assessed as part of our work [16]. Overall, we believe the system now proposed addresses some of the limitations of the existing system, which is based on spontaneous reports and thus is limited by underreporting [3].

We were able to replicate a known signal for febrile seizures following MMR based a one-year surveillance period. This signal was identified only with PMaxSPRT (after 3 months of surveillance, if a minimum events of 2 events was stipulated). Although BMaxSPRT did not quite signal as it is a less powerful test, it has the advantage
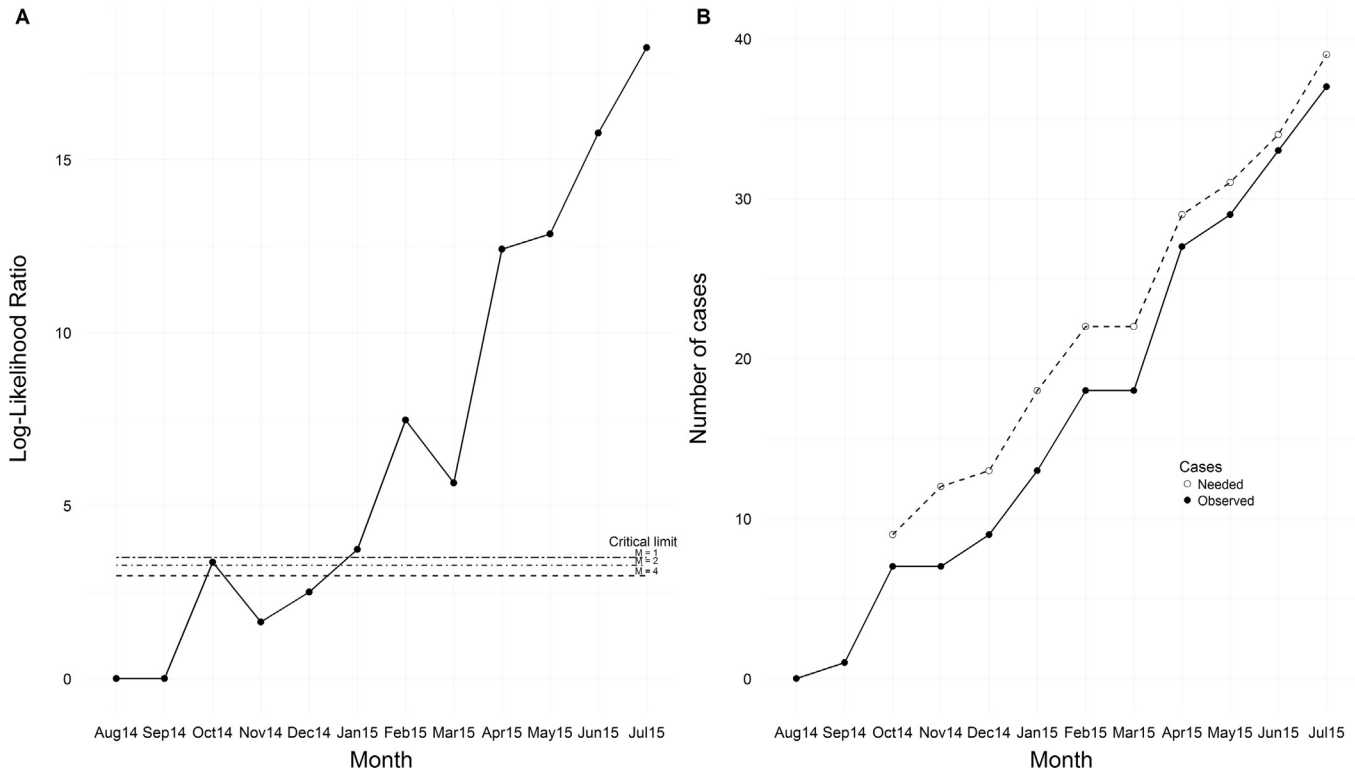
Fig. 3. Implementation of a system for MMR/febrile seizures using PMaxSPRT (right) and BMaxSPRT (left). Only for PMaxSPRT a signal is detected.

Table 3
Power and expected time to signal for MMR/febrile seizures after, using Poisson and Binomial-based Maximized Sequential Probability Ratio.

Minimum events	Test	Data available at	Power (time to signal in months from beginning of surveillance)[a] Relative risk								
			1.5	2	2.5	3	4	5	6	8	10
1	PMaxSPRT	06–07–2015	30	73 (5)	**95 (4)**	**99 (3)**	**100 (2)**	**100 (1)**	**100 (1)**	**100 (1)**	**100 (1)**
	BMaxSPRT	06–07–2015	28	63 (6)	**85 (6)**	**95 (5)**	**99 (5)**	**100 (4)**	**100 (3)**	**100 (3)**	**100 (3)**
2	PMaxSPRT	06–07–2015	33	76 (5)	**96 (4)**	**100 (3)**	**100 (2)**	**100 (1)**	**100 (1)**	**100 (1)**	**100 (1)**
4	PMaxSPRT	06–07–2015	36	79 (5)	**96 (4)**	**100 (3)**	**100 (2)**	**100 (1)**	**100 (1)**	**100 (1)**	**100 (1)**

Cells in bold refer to power≥80%.
BMaxSPRT - Binomial-based Maximized Sequential Probability Ratio.
PMaxSPRT - Poisson-based Maximized Sequential Probability Ratio Test.
  [a] Time to signal is only displayed for cells where equivalent power≥50%.

of having a much more relevant comparator period that should be less prone to bias and would likely have signalled with an extended surveillance period. We would therefore suggest that despite the low number of expected events (11) it is still worthwhile using this method in addition to PMaxSPRT to make the signal more robust. Others have suggested a minimum number of expected events of 50 [2].

A further aspect is the minimum number of events required to reject the null hypothesis. As previous work has suggested, rejecting the null hypothesis only after a certain number of events increases power [14]. Given we have limited power for seasonal influenza/GBS we would recommend implementing a system with a requirement of 4 events before rejecting the null.

Vaccine safety studies require careful specification of risk-windows and, if applicable, comparator windows. This includes not only knowledge of the characteristics of the vaccine/outcome pair under study but also the data available for analysis. In the case of MMR/seizures we decided to use a longer risk-window than previously suggested (6–21 days instead of 6–11 days) to account for delayed recording of seizures in the primary care data. If our choice

resulted in an unduly long risk-window the result would be an underestimation of the risk and thus a reduction in the power to detect a signal. In practice, a way to address uncertainty in the specification of risk-windows is to conduct a sensitivity analysis using an alternative risk/comparator window. Alternatively, this uncertainty can be addressed at the confirmatory stage by looking at the distribution of cases within the risk-window.

Data quality should also be considered. Our previous assessment of completeness of records first diagnosed in hospital showed that CPRD had low sensitivity to capture GBS. However, if this sub-optimal sensitivity is constant over time, for the purposes of the current system the effect would be a decrease in power to detect an event [2]. We know of no studies assessing the positive predictive value of the outcomes included. As for the vaccination data, the vaccines we selected are administered in general practices and GPs are financially incentivised to achieve certain thresholds of vaccine uptake. It is thus expected that individuals classified as vaccinated are indeed so.

Our study presents several limitations. The use of PMaxSPRT is susceptible to uncertainty in historical rates and a conditional test

was proposed to address this issue. We tried to minimize this by using data from the 5 previous seasons/years to estimate historical rates. Given the amount of observed events in the historical data is substantial larger (more than five times) than the number of expected events in the study period we considered that the use of a conditional test was not necessary. Secondly, for our study period we considered only vaccinated individuals while for historical rates we included both vaccinated and non-vaccinated. Including vaccinated individuals in historical periods could have led to a slight overestimate of the background rate and underestimate of the RR and thus miss a signal. However, even if there were increases in risk in the historical periods due to the vaccine, the increase in the attributable risk would be small, thus minimizing this issue. Nevertheless, we were able to detect a signal for seizures/MMR, which is reassuring. Finally, our study is limited by assumptions of the method used, including homogenous distribution of a potential risk during the risk-window and throughout the study period and that if there is an increase in risk these additional cases would be also recorded in CPRD.

We proposed a new adjustment for delays but it might still not fully capture existing delays [17]. We only considered a mid-point for adjustment, which simplifies the data accrual process. Furthermore, we considered a delay distribution based on historical data and recording patterns might have changed, although previous work looking at ten years' worth of data shows consistent recording patterns [4]. Overall, we believe our adjustment reduces bias due to data availability and enables an earlier start of surveillance.

As previously pointed out there are few strategies available to deal with potential confounding factors [2]. For influenza/GBS we were able to account for gender and partially for age. If there is a signal, further adjustment for confounders is one of the initial steps [1], potentially including more detailed adjustment for age (we only considered 10-year age groups) and for other potential confounders. Influenza incidence may be one of these potential confounders, as GBS is known to be associated with influenza-like illness [18]. Rapid yearly estimates are provided for influenza incidence and could potentially be used in this context. For seizures/MMR, we were able to account for age and gender in the PMaxSPRT analysis but we did not explicitly account for age in the BMaxSPRT. Febrile seizure rates are known to change rapidly with age [19] but the use of a control period before and after the risk period should have helped to limit potential confounding due to age.

Our study made use of previously collected data to mimic a new system. However, CPRD is expanding, to include practices using different softwares [20]. While this can be seen as an opportunity to increase power to detect lower increases in risk for rare outcomes, there might be differences in coding systems and behaviour that could limit the applicability of the results of our previous studies. Alternatively, these new practices could be used for signal confirmation should a signal be identified. This strategy would be a way to avoid using the same data for signal identification and confirmation.

As we have further knowledge on NRTVSS and its application using CPRD next steps include application to new vaccines. In addition, there is the need to define which steps to undertake if a signal is detected. Yih et al. [1] proposed a series of steps in case a signal is found, broadly including: to check data and code, to examine descriptive statistics for patterns in time between the exposure and outcome, to adjust for additional confounders, to conduct a non-sequential analysis with a different comparator, to conduct a review of records, to compare the results with similar outcomes or other existing data, to analyse new data or to design a new study. All steps can be conducted using CPRD data. However, there is limited ability to perform a timely confirmation of the patient's recorded diagnoses. Currently, when GPs are asked to validate diagnoses identified from coded information this process may take

several months. Future discussions with data providers and medicines regulatory authorities may help to facilitate the process of data validation. An in-depth presentation of the steps required following a signal is beyond the scope of this work.

In conclusion, our results suggest the implementation of NRTVSS using CPRD as a way to complement existing methods, by allowing timely identification of signals for more frequent outcomes and by excluding large increases in risk for less frequent outcomes.

Ethics statement

All data were anonymised prior to receipt by the authors. Approval for the study was obtained from the Independent Scientific Advisory Committee of the Medicines and Healthcare Products Regulatory Agency (ISAC number: 15_230) and from the Ethics Committee of the London School of Hygiene & Tropical Medicine (LSHTM reference: 10421). The procotol for the overall programme of work was made available for reviewers.

Author contributions

NA conceived the idea for the study. All authors contributed to the design of the study. AL analysed the data and wrote the first and final drafts of the manuscript, with input from NA and ST. All authors contributed to and approved the final manuscript.

Conflict of interest

The authors state no conflict of interest.

Acknowledgments

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.vaccine.2017.09.022.

References

[1] Yih WK et al. Active surveillance for adverse events: the experience of the vaccine safety datalink project. Pediatrics 2011;127:S54–64.
[2] Leite A, Andrews NJ, Thomas SL. Near real-time vaccine safety surveillance using electronic health records-a systematic review of the application of statistical methods. Pharmacoepidemiol Drug Saf 2016;25:225–37.
[3] Donegan K et al. Bivalent human papillomavirus vaccine and the risk of fatigue syndromes in girls in the UK. Vaccine 2013;31:4961–7.
[4] Leite A, Andrews NJ, Thomas SL. Assessing recording delays in general practice records to inform near real-time vaccine safety surveillance using the Clinical Practice Research Datalink (CPRD). Pharmacoepidemiol Drug Saf 2017;26:437–45.
[5] Herrett E et al. Data resource profile: Clinical Practice Research Datalink (CPRD). Int J Epidemiol 2015;44:827–36.
[6] DATA.GOV.UK. UK Read Codes. <https://data.gov.uk/dataset/uk-read-code> [accessed 22 August 2017].

[7] Miller E et al. Risks of convulsion and aseptic meningitis following measles-mumps-rubella vaccination in the United Kingdom. Am J Epidemiol 2007;165:704–9.

[8] Silva IR, Kulldorff M. Continuous versus group sequential analysis for post-market drug and vaccine safety surveillance. Biometrics 2015;71:851–8.

[9] Kulldorff M et al. A maximized sequential probability ratio test for drug and vaccine safety surveillance. Seq Anal 2011;30:58–78.

[10] Li L, Kulldorff M. A conditional maximized sequential probability ratio test for pharmacovigilance. Stat Med 2010;29:284–95.

[11] Silva IR, Kulldorff M. Sequential: exact sequential analysis for poisson and binomial data. R Package Version 2.3.1. 2017. <http://CRAN.R-project.org/package=Sequential> [accessed 19.05.17]

[12] Lewis JD et al. The relationship between time since registration and measured incidence rates in the General Practice Research Database. Pharmacoepidemiol Drug Saf 2005;14:443–51.

[13] Public Health England. Measles: the green book. In: The green book. PHE; 2014.

[14] Kulldorff M, Silva IR. Continuous post-market sequential safety surveillance with minimum events to signal; 2015. arxiv:1503.01978 [stat.ap].

[15] Public Health England. Influenza: the green book. In: The green book. PHE; 2014.

[16] Sandhu SK et al. Near real-time surveillance for Guillain-Barre syndrome after influenza vaccination among the Medicare population, 2010/11 to 2013/14. Vaccine 2017;35:2986–92.

[17] Greene SK et al. Near real-time vaccine safety surveillance with partially accrued data. Pharmacoepidemiol Drug Saf 2011;20:583–90.

[18] Alcalde-Cabero E et al. Guillain-Barré syndrome following the 2009 pandemic monovalent and seasonal trivalent influenza vaccination campaigns in Spain from 2009 to 2011: outcomes from active surveillance by a neurologist network, and records from a country-wide hospital discharge database. BMC Neurol 2016;16:75.

[19] Sammon CJ et al. The incidence of childhood and adolescent seizures in the UK from 1999 to 2011: A retrospective cohort study using the Clinical Practice Research Datalink. Vaccine 2015;33:7364–9.

[20] The Clinical Practice Research Datalink (CPRD). Descriptive characteristics of the Clinical Practice Research Datalink (CPRD) EMIS database. 2017. <https://www.cprd.com/isac/Protocol_15_217.asp> [accessed 19.05.17].