ORIGINAL RESEARCH ARTICLE

Temporal Data Mining for Adverse Events Following Immunization in Nationwide Danish Healthcare Databases

Henrik Svanström,[1] *Torbjörn Callréus*[2] and *Anders Hviid*[1]

1 Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark
2 The Danish Medicines Agency, Copenhagen, Denmark

Abstract

Background: A rarely used opportunity in pharmacovigilance is data mining for adverse drug reactions (ADRs) in population-based healthcare databases.
Objective: To evaluate the potential of data mining for ADRs in the nationwide Danish healthcare databases. We specifically considered hospital contacts following measles, mumps and rubella (MMR) immunization.
Methods: We constructed a cohort consisting of all children born in Denmark from 1995 to 2007 (n = 918 831) with individual-level linked data on childhood vaccinations and hospital contacts from the nationwide Danish healthcare databases. We applied a cohort-based data mining methodology to compare the observed versus the expected incidence of adverse event in different time periods relative to immunization. With this approach we evaluated temporal associations between MMR immunization and 5915 different diagnoses occurring in the cohort. In order to evaluate the ability of our approach to detect signals, we singled out a set of four adverse events previously recognized as being associated with the MMR vaccine.
Results: We were able to link a total of 3 162 251 hospital contacts and 5915 different diagnoses to the children in the cohort. Previously recognized temporal associations between adverse events (febrile convulsions, idiopathic thrombocytopenic purpura, lymphadenopathy and rash) and MMR immunization were identified in the Danish databases by our method.
Conclusions: Data mining in the Danish population-based healthcare databases provides adequate ability to detect adverse events. Pharmacovigilance using electronic healthcare databases holds potential as an important supplement to traditional pharmacovigilance.

Background

Before the marketing of a new medical product, efficacy and safety have been thoroughly evaluated in randomized clinical trials. However, rare but important adverse reactions may not be readily detected even in large trials, and they may therefore remain unknown at the time of market

launch. Recent examples include intussusception among infants given oral rotavirus vaccine[1,2] and the risk of progressive multifocal leukoencephalopathy following exposure to natalizumab.[3] Hence, there is a need for effective postmarketing drug surveillance systems and, despite its limitations, spontaneous reporting of suspected adverse drug reactions (ADRs) has traditionally constituted the cornerstone of these systems.[4] Reports obtained from healthcare professionals and patients are registered in passive surveillance databases, and reporting patterns may identify safety signals worthy of further evaluation.

An interesting opportunity in postmarketing surveillance is data mining for ADR signals in linked healthcare databases.[4] These databases linking drug exposures, potential ADRs and relevant co-variates on the individual level are becoming more and more available due to increased use of electronic data collection and automation of medical records. Denmark has a number of linkable and comprehensive nationwide databases allowing for the construction of large prospective cohorts with detailed information on drug exposures, health outcomes and relevant co-variates.[5]

The objective of this study was to evaluate the potential of ADR signal detection in the Danish population-based healthcare databases. Specifically, we implemented a cohort-based data mining approach to evaluate temporal associations between measles, mumps and rubella (MMR) immunization and hospital contacts. A set of known ADRs was chosen to assess the validity of our approach

Materials and Methods

Overview

The general approach of our study was to analyse differences between the observed and the expected number of hospital contacts with a specific diagnosis in different time periods relative to MMR immunization. The analysis was performed with respect to each of 5915 different diagnoses that occurred in the cohort. Our methodology can be subdivided into three distinct steps (stratification of person-time and observed hospital contacts, calculation of the expected number of hospital

contacts and statistical analysis) that are described below.

Study Cohort

All people living in Denmark are assigned an individual ten-digit identification number in the Danish Civil Registration System. This registry was established in 1968 and contains information on all residents with reference to date of birth, death and emigration.[6] On the basis of the Civil Registration System we created a cohort comprising all children born in Denmark from 1 January 1995 through 31 December 2007. Using the unique personal identification numbers, we were able to individually link information on childhood vaccinations and hospital contacts to the children in the cohort.

Immunizations

The dates for MMR immunization were obtained from the National Board of Health. In Denmark, all immunizations are performed by general practitioners who are reimbursed only after reporting immunization dates and details to the National Board of Health. The MMR vaccine was introduced in Denmark in 1987 and since then has been routinely administered, with a first dose at 15 months of age. Before 2008, the second dose was administered at 12 years but is now given at 4 years of age. For this study, we included information on the first dose only. The live vaccine in use since 1987 is composed from the virus strains Enders-Edmonston (measles), Jeryl Lynn (mumps) and Wistar RA 27/3 (rubella).[7] Since its introduction, the MMR immunization coverage rate has fluctuated around 85–90% in Denmark.

Hospital Contacts

Information on hospital contacts from 1995 through 2007 was obtained from the Danish National Hospital Register.[8] This population-based registry has nationwide coverage and holds diagnosis-specific information, using the *International Classification of Diseases 10th revision* (ICD-10), on hospital admissions, emergency room visits and outpatient visits with practically

complete registration. We included ICD-10 codes with first letters A–N and R–T, thereby excluding diagnostic events in connection with pregnancy, the perinatal period, congenital malformations, codes following contact with health services and special purpose codes.

To avoid excessive influence from individuals with large numbers of hospital contacts, we focused on hospital contacts with an incident diagnosis only. Subsequent hospital contacts with a repeated diagnosis were censored accordingly.

Stratification of Person-Time and Observed Hospital Contacts

The children of the cohort were followed from birth until censoring (death, disappearance or emigration) or 31 December 2007, whichever occurred first. The resulting person-times of follow-up were summed according to strata defined by age (in 3-month intervals), calendar period (in 2-year intervals) and immunization status. The 84-day time period preceding immunization was stratified according to time to immunization in 14-day pre-immunization intervals. Exposed person-time lasted 364 days from immunization and was correspondingly stratified according to time since immunization in 14-day exposure intervals. All other person-time was classified as unexposed.

Records of hospital contacts for each specific diagnosis were stratified and counted using the same strata. Furthermore, with regard to each diagnosis, stratified person-time following first hospital contact was subtracted from the total amount of stratified person-time. Hence, we obtained one set of stratified person-time and cor-

responding number of hospital contacts for each different diagnosis.

In figure 1, the calculation and stratification of person-time is illustrated by examples of observed histories. Individual$_1$ and individual$_2$ are immunized and thus contribute age- and calendar-stratified unexposed, pre-immunization (lightly shaded) and exposed (darkly shaded) person time. The unimmunized individual$_3$ contributes age- and calendar-stratified unexposed person-time only. The follow-up of individual$_1$ is terminated by the occurrence of a hospital contact.

Calculation of the Expected Number of Hospital Contacts

The expected number of hospital contacts for each specific diagnosis in the respective pre-immunization and exposure strata was calculated by indirect standardization. First, unexposed age- and calendar-specific incidence rates were calculated by dividing the number of hospital contacts during unexposed person-time by the sum of unexposed person-time within each age and calendar stratum. Next, for each pre-immunization and exposure stratum, age- and calendar-specific person-time was multiplied by the corresponding unexposed age- and calendar-specific incidence rate. The sum of these products yielded the expected number of hospital contacts for the specific diagnosis and stratum.

Statistical Analysis

To obtain estimates of disproportionality, we adopted methodology and notation proposed by Norén et al.[9] With the object of identifying
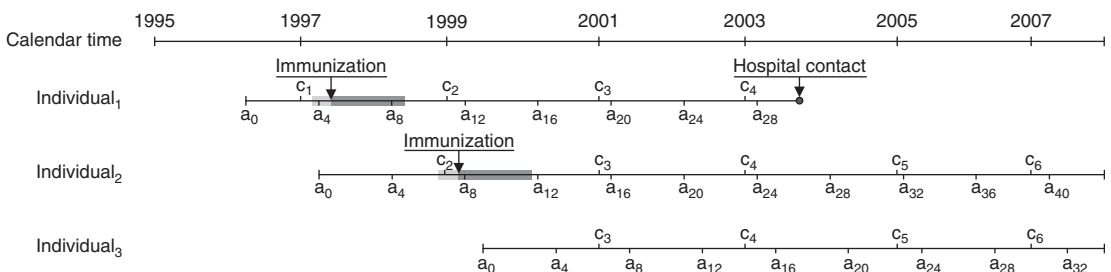
Fig. 1. Illustration of person-time calculation and stratification using examples of observed event histories. a_i indicates the start of age interval i and c_j indicates the start of calendar interval j. Pre-immunization intervals are lightly shaded and exposure intervals are darkly shaded.

temporal patterns, this approach produces an estimate that contrasts the observed to the expected incidence of an outcome event in different time periods relative to an index event (e.g. exposure to a medical product).

First, the logarithm of the observed-to-expected ratio can be defined as:

$$\log_2 \frac{O^t_{xy}}{E^t_{xy}} \qquad \text{(Eq. 1)}$$

In this expression, O and E denote the observed and the expected number of events, respectively. The subscript x refers to the index event of interest, subscript y refers to the outcome of interest and superscript t denotes a time period relative to the occurrence of index event x. Accordingly, the observed-to-expected ratio has a null-hypothesis value of zero, signifying no association between events x and y in time period t. A positive ratio indicates that event y has occurred more often than expected, whereas a negative value indicates that event y has occurred more rarely.

Due to sampling variation, the raw observed-to-expected ratio suffers from large variance when frequencies are small. This might in turn cause excessive measures of disproportionality, making it difficult to distinguish actual disproportionality from noise. In disproportionality analysis of cross-sectional data (e.g. spontaneous reports), sampling variation has been effectively adjusted by use of statistical shrinkage.[10,11] This technique accounts for sampling variation in a Bayesian manner by combining the observed real data with a moderating prior distribution. The resulting posterior distribution is thereby drawn towards the null-hypothesis value when frequencies are small. As frequencies increase, the impact of the prior distribution decreases.

To handle sampling variation in temporal pattern discovery, Norén et al.[9] propose a shrinkage approach in an adaptation of the Information Components (IC) disproportionality measure. We applied this adapted version of the IC, expressed as:

$$IC = \log_2 \frac{O^t_{xy} + \alpha}{E^t_{xy} + \beta} \qquad \text{(Eq. 2)}$$

Here, α and β are the parameters in a Gamma (α, β) prior distribution. With this specification, the posterior distribution for the IC is also Gamma, but with parameters ($O^t_{xy} + \alpha$) and ($E^t_{xy} + \beta$). The specified values of α and β can equally be regarded as a corresponding number of events being added to the observed and expected number of events, respectively. By altering the values of α and β, the strength and direction of shrinkage is adjusted. For this study, we applied parameter values $\alpha = 1$ and $\beta = 1$, resulting in shrinkage towards IC = 0. From the posterior distribution, a two-sided 95% credibility interval for the IC is defined by using the 2.5 and 97.5 percentiles as the lower and upper limits, respectively.

Adverse Events

In order to evaluate the resulting IC estimates, we employed a set of four adverse events (AEs) previously recognized as being associated with the MMR vaccine.

Febrile convulsions: MMR immunization has been shown to be followed by a transient increased risk of febrile seizures. Vestergaard et al.[12] found a 2- to 3-fold relative increased incidence rate of febrile seizures in the first 2 weeks following immunization. We used the ICD-10 code R56.0.

Idiopathic thrombocytopenic purpura (ITP): Thrombocytopenia is known to occur after natural infection of measles and rubella and is a documented AE following MMR immunization. In a study by Miller et al.,[13] a more than 3-fold change in the relative incidence of ITP in the first 6 weeks after immunization was described. We used the ICD-10 code D69.3.

Lymphadenopathy: Rubella vaccine is associated with lymphadenopathy, occurring in up to 9% of recipients.[14,15] We used the ICD-10 code R59.1.

Rash: A transient rash occurs in about 5% of all recipients of the measles vaccine.[16] The rash usually appears 7–10 days after immunization and persists for about 2 days. The ICD-10 code used was R21.

Results

The study cohort included 918 831 children, contributing a total of 5 958 774 person-years of

follow-up. Using personal identification numbers, we were able to link a total of 3 162 251 hospital contacts containing 5915 different diagnoses to the children of the cohort.

In the following section we present a screening of all IC estimates for the exposure interval of 0–13 days after immunization and IC estimates for exposure intervals up to 364 days after immunization for AEs listed in the Materials and Methods section.

Screening of Hospital Contacts 0–13 Days after Immunization

A total of 13 195 hospital contacts, involving 1095 different diagnoses, occurred in the exposure interval 0–13 days after immunization. The most common diagnoses were febrile convulsions (ICD-10 code R56.0), unspecified viral infection (B34.9) and unspecified fever (R50.9), accounting for 1344, 577 and 398 hospital contacts, respectively. For 43 different diagnoses, the IC estimate was statistically significantly higher than the null-hypothesis value of zero at the 0.05 significance level. In table I, we present diagnoses with the top ten highest lower 95% credibility bound values. As seen from the table, the three highest lower-bound values referred to diagnoses directly related to complications following vaccine administration.

In figure 2 we present a graphical representation, often referred to as a heat map, to summarize the lower 95% credibility bound for the IC for all diagnoses that occurred in the 0- to 13-day exposure interval. The width of each column corresponds to the relative occurrence of hospital contacts with a diagnosis code in the respective ICD chapters. Furthermore, each rectangle represents a specific diagnosis where the colour of the square corresponds to the size of the estimate and the area corresponds to its relative occurrence. As an example, the large red square in the column representing the ICD chapter on symptoms and signs (ICD block R00–R99) corresponds to febrile convulsions, recognized in table I as commonly occurring with a significant positive IC estimate.

As seen from the figure, significant positive estimates occurred across almost all ICD chapters and did not seem to display any clear pattern. The highest number of positive lower bound values occurred in the ICD chapters on injury, poisoning and other consequences of external cause (S00–T98), symptoms and signs (R00–R99) and diseases of the musculoskeletal system and connective tissue (M00–M99). Notably, most significant estimates corresponded to relatively common diagnoses.

Furthermore, it can be seen that hospital contacts in the ICD chapters on injury, poisoning and other consequences of external cause, symptoms and signs and diseases of the respiratory system (J00–J99) were the most occurring whereas neoplasms (C00–D48), mental and behavioural disorders (F00–F99) and diseases of the circulatory system (I00–I99) occurred relatively rarely.
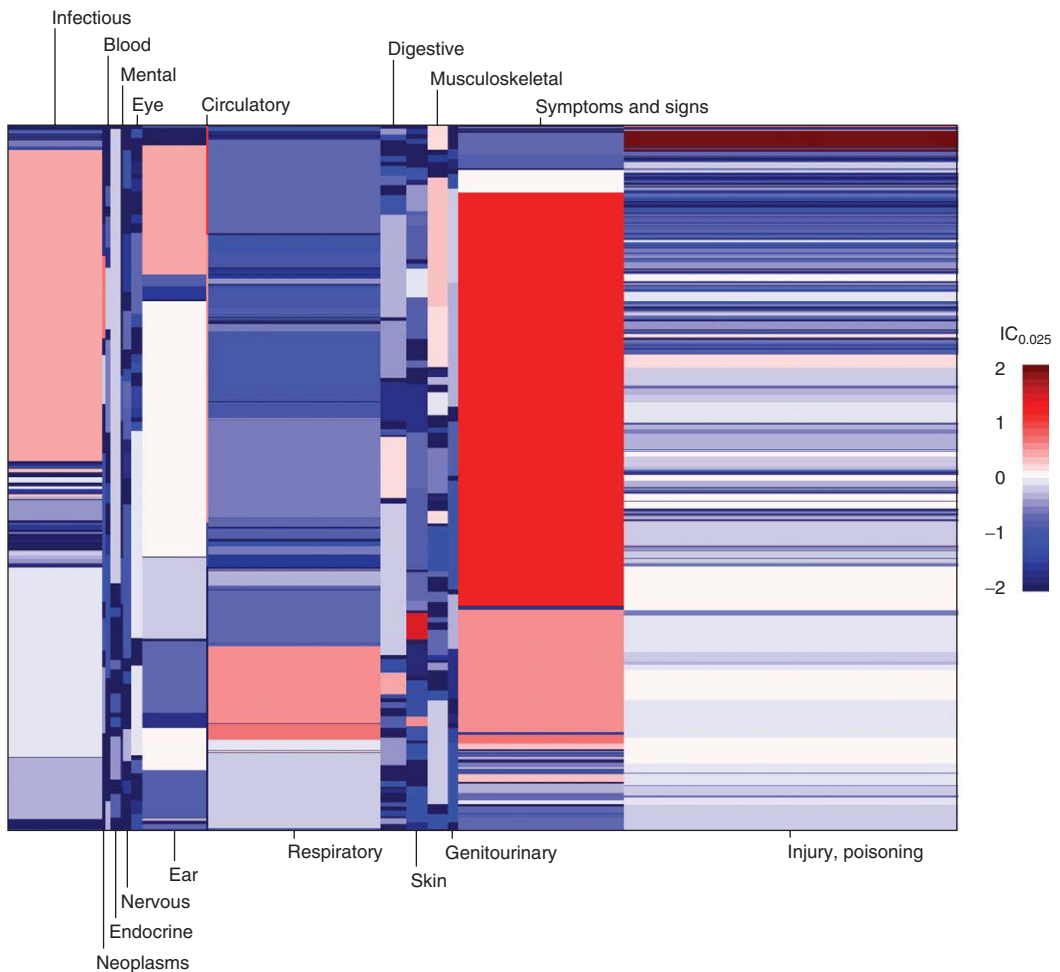
Table I. Adverse events with the top ten highest lower 95% credibility bound for the information components (IC) in the time period 0–13 days following measles, mumps and rubella (MMR) immunization

Adverse event	Observed events (n)	Expected events (n)	$IC_{0.025}$
Infection following immunization	21	0.7	3.02
Other complications following immunization	84	7.4	3.01
Other complications following infusion, transfusion and therapeutic injection	16	1.1	2.22
Acute lymphadenitis, unspecified	11	1.3	1.44
Febrile convulsions	1344	542.1	1.23
Lymphangitis	5	0.1	1.05
Sequelae of complications of surgical and medical care	14	3.8	0.81
Benign neoplasms of scalp and neck	5	0.3	0.71
Fever of other and unknown origin	29	11.8	0.67
Streptococcal tonsillitis	52	24.6	0.64

Adverse Events

In table II we present IC estimates and 95% credibility interval limits for AEs listed in the Materials and Methods section. The presented results are limited to exposure intervals up to 83 days after immunization. Febrile convulsions was by far the most common among the selected diagnoses, with a total of 30 979 incident events. Conversely, lymphadenopathy was the most uncommon diagnosis with 130 incident events in total.

Regarding febrile convulsions, IC estimates for the exposure intervals 0–13 and 28–41 days after immunization proved statistically significantly higher than the null-hypothesis value of zero. The estimate for the 28- to 41-day interval was, however, relatively close to zero and had just marginal statistical significance. For rash, the positive IC for the exposure interval 0–13 days after immunization was the sole significant estimate. With respect to ITP, only the IC estimate 14–27 days after immunization proved significant. Similarly, for lymphadenopathy, only the positive IC estimate

Fig. 2. Heat map of the lower 95% credibility bound for the information components for diagnoses in the time period 0–13 days following measles, mumps and rubella virus (MMR) immunization. Column width corresponds to the relative occurrence of hospital contacts with a diagnosis code in the respective *International Classification of Diseases* chapters. Each rectangle represents a specific diagnosis where the colour corresponds to the size of the estimate and the area corresponds to its relative occurrence. Estimates are truncated at ±2.

Table II. Information component (IC) estimates and 95% credibility intervals (CI) for selected adverse events in the time period 0–83 days after measles, mumps and rubella (MMR) immunization

Adverse event	n	IC (95% CI)					
		0–13 days	14–27 days	28–41 days	42–55 days	56–69 days	70–83 days
Febrile convulsions	30 979	1.31 (1.23, 1.38)	0.05 (−0.07, 0.17)	0.13 (0.01, 0.25)	−0.01 (−0.14, 0.12)	−0.04 (−0.17, 0.09)	0.00 (−0.14, 0.14)
Rash	1 433	0.87 (0.27, 1.36)	0.15 (−0.66, 0.78)	−0.10 (−1.01, 0.59)	0.35 (−0.43, 0.96)	−0.84 (−2.16, 0.06)	0.40 (−0.43, 1.05)
ITP	529	0.45 (−0.61, 1.22)	1.40 (0.67, 1.98)	0.79 (−0.16, 1.50)	−1.13 (−3.41, 0.14)	−0.28 (−1.90, 0.76)	−1.41 (−4.45, 0.07)
Lymphadenopathy	130	0.66 (−1.61, 1.93)	1.88 (0.57, 2.78)	0.65 (−1.63, 1.92)	1.06 (−0.82, 2.19)	1.37 (−0.26, 2.40)	0.61 (−1.66, 1.88)

ITP = idiopathic thrombocytopenic purpura.

14–27 days after immunization deviated significantly from zero.

Temporal Patterns

Figure 3 displays IC estimates and 95% credibility interval limits according to time relative to MMR immunization. The graph regarding febrile convulsions demonstrates a clear peak in estimates in the immediate time period after immunization. Estimates then rapidly drop toward the null-hypothesis value of zero. However, from approximately 2–3 months after immunization, estimates again tend to increase, indicating a slight but significant increased risk of febrile convulsions in time periods up to more than 6 months after immunization. When it comes to ITP, the graph is characterized by significantly low estimates for time periods from 2 to 3 months prior up until immunization. The graph demonstrates a distinct peak in estimates in the time period up to 6 weeks after immunization. Around 7–8 months after immunization there is one estimate significantly higher than the null-hypothesis value. The graph regarding rash demonstrates significantly lowered estimates just prior to immunization and a transient peak in estimates shortly after immunization. All other estimates are fairly close to zero. Regarding lymphadenopathy, there is a similar peak in estimates in time periods just after immunization. Because of the limited number of observations, the graph is characterized by comparatively wide credibility intervals.

Discussion

In this study, we evaluated the potential of signal detection in population-based nationwide Danish healthcare databases by application of cohort-based disproportionality analysis. The assessment focused on potential ADRs following MMR immunization.

The screening of all hospital contacts in the 14 days immediately after immunization highlighted 43 different diagnoses. The most significant estimates corresponded to diagnoses specifically related to complications following immunization. The significantly elevated estimates of fever and febrile convulsions were anticipated and were consistent with the safety profile of the MMR vaccine. Similarly, lymphadenitis and lymphangitis could well be associated with this live vaccine. The increased risk of benign neoplasms of scalp and neck is most likely not causally related to MMR. More plausibly, this finding is due to detection bias, given that all immunized children had a recent medical visit.

With regard to the selected AEs, the highlighted temporal associations were mostly consistent with prior research. A transient increased risk of febrile convulsions, rash and lymphadenopathy shortly after immunization is well recognized. Regarding ITP, the elevated estimates for the first 42 days and the significantly high estimate 14–27 days after immunization similarly agree with earlier research. In the previously cited study by Miller et al.,[13] the highest relative incidence of ITP was found between 15 and 28 days after
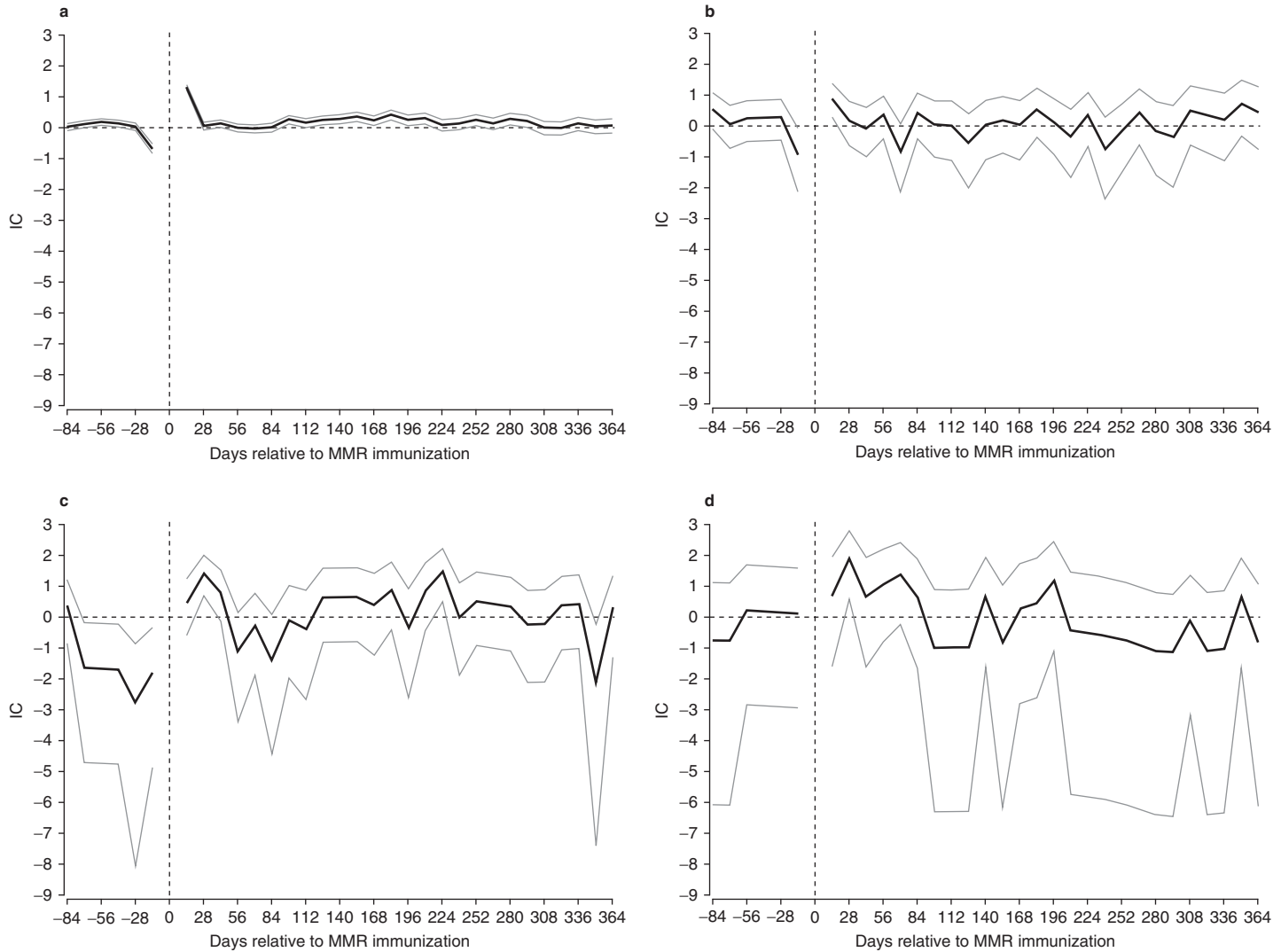
Fig. 3. Information component (IC) estimates (black line) and 95% credibility limits (grey lines) for (**a**) febrile convulsions, (**b**) rash, (**c**) idiopathic thrombocytopenic purpura and (**d**) lymphadenopathy according to time relative to measles, mumps and rubella (MMR) immunization.

MMR immunization. However, the persistent pattern of significantly high estimates for febrile convulsions from approximately 3 months after immunization is unexpected and is unlikely to represent an actual causal association. The noticeably low estimates in time periods just prior to immunization regarding febrile convulsions, rash and ITP are likely explained by 'the healthy vaccinee effect' – healthy children are more likely to be vaccinated, thus inducing a downward bias on vaccine-effect estimates.

In summary, the highlighted temporal associations were generally consistent with earlier research. The results suggest that the methodology and implementation provide adequate ability to detect AEs.

Strengths and Limitations

Although an important tool for drug safety surveillance, spontaneous reporting systems have a number of widely recognized limitations.[17-19] The utility of spontaneous reports is influenced by the temporal relationship between the AE and the exposure and the incidence of the AE in the underlying population.[20] AEs that occur late after use, from long-term use or emerge as an increased risk of a relatively common disease are less likely to be detected from spontaneous reports. By utilizing healthcare databases in pharmacovigilance, a number of these limitations can be addressed. The data employed in this study yielded practically complete coverage of serious AEs (as defined by hospital contacts) in the population.

In addition, the population-based Danish healthcare databases provided the opportunity for signal detection in a large, stable, well defined population. The cohort nature of the data, furthermore, allowed assessment of the absolute number of specific AEs in different time periods relative to immunization.

The specific implementation provided additional benefits. Using this approach, we were able to assess risk over time relative to immunization. This feature provided valuable information regarding the relative timing of events as well as a better understanding of temporal associations over time. In particular, this approach enabled us to detect delayed-onset ADRs up to 1 year after immunization. As an example, we were able to clearly identify the relatively delayed increased risk of ITP.

The applied approach has a broad applicability and may well be modified to handle a variety of different exposures and outcome events. In Denmark, the present approach could be applied to exposure data from the nationwide Danish National Prescription Registry, which holds individual-level information on all prescriptions filled at pharmacies.

We recognize several limitations of our study. The study was limited to data on hospital contacts only. The sensitivity for mild and moderate AEs and diagnostic events rarely treated in a hospital setting was accordingly low. Similarly, the sensitivity for sudden fatal events not leading to hospital contact was expected to be low.

We also acknowledge the possibility of residual confounding. Estimates were adjusted for age and calendar effects only and, accordingly, other potential significant confounders that may exist were not controlled for. In particular, the unexpected results with respect to febrile convulsions may possibly result from confounding factors in the underlying population. It should also be noted that due to the large number of intervals considered simultaneously, a number of false significant estimates is expected.

Lastly, it should be noted that associations detected through disproportionality analysis should merely be considered signals for further evaluation and do not imply causality.[21] Conversely, the absence of a signal does not automatically mean that there is no causal relationship. The results should aid in hypotheses generation concerning a possible linkage between MMR vaccine and the respective diagnoses.[22] In order to confirm these hypotheses additional investigations using analytical study designs are needed.

Similar Studies

Recent research on data mining in electronic healthcare databases has primarily involved sequential analytic methods, focusing on real-time surveillance of the probabilities of selected ADRs

of specific interest. Population-based real-time surveillance methodology has been applied to evaluate the association between selected AEs and influenza vaccination and selected drugs, respectively,[23,24] risk of selected AEs following the shift from whole cell diphtheria-tetanus-pertussis (DTP) to acellular DTP vaccine,[25] risk of intussusception following rotavirus vaccination[26] and the risk of Guillain-Barre-syndrome following exposure to meningococcal conjugate vaccination.[27] Furthermore, in a recent study, Hocine et al.[28] developed a sequential data mining method based on the self-controlled case-series method to assess the risk of Bell's palsy and bleeding disorders following influenza and MMR immunization, respectively. The application of sequential analytic methods allows for early detection of AEs from newly introduced drugs or vaccines, while at the same adjusting for multiple testing. However, in contrast to the approach of our study and the methodological framework by Norén et al.,[9] these methodologies are best suited for evaluation of short-term AEs and do not provide explicit information of temporal associations over time.

Conclusions

Electronic healthcare databases have become increasingly accessible for research. Although collected for other purposes, the information in these data sources could play an important role as a supplement to traditional pharmacovigilance for ADRs. The results from this study demonstrate that pharmacovigilance by use of electronic healthcare databases has the potential to contribute valuable information with respect to public health and, furthermore, allow signal generation with adequate ability to detect AEs. Signal generation in electronic healthcare databases holds particular potential as a tool for understanding temporal patterns and for detection of delayed-onset ADRs.

The abundance of information in healthcare databases requires techniques for research to move beyond traditional epidemiological study designs, inherited from an era when data collection was expensive and yielded only a few facts

for each study subject.[29] Our study should help generate further research in this area. There is an apparent need for more studies in order to compare various methods and fully understand the benefits and risks associated with each approach.[30] Adequate methodologies must be properly validated with proven ability in terms of detecting ADRs and minimizing the risk of highlighting spurious associations.

Acknowledgements

References

1. Centers for Disease Control and Prevention (CDC). Intussusception among recipients of rotavirus vaccine: United States, 1998-1999. MMWR Morb Mortal Wkly Rep 1999 Jul 16; 48 (27): 577-81

2. Murphy TV, Gargiullo PM, Massoudi MS, et al. Intussusception among infants given an oral rotavirus vaccine. N Engl J Med 2001 Feb 22; 344 (8): 564-72

3. Major EO. Reemergence of PML in natalizumab-treated patients: new cases, same concerns. N Engl J Med 2009 Sep 10; 361 (11): 1041-3

4. Berlin JA, Glasser SC, Ellenberg SS. Adverse event detection in drug development: recommendations and obligations beyond phase 3. Am J Public Health 2008 Aug; 98 (8): 1366-71

5. Hallas J. Conducting pharmacoepidemiologic research in Denmark. Pharmacoepidemiol Drug Saf 2001 Dec; 10 (7): 619-23

6. Malig C. The civil registration system in Denmark. IIVRS technical paper. Bethesda (MD): International Institute for Vital Registration and Statistics, 1996. Report no. 66

7. Hviid A. Postlicensure epidemiology of childhood vaccination: the Danish experience. Expert Rev Vaccines 2006 Oct; 5 (5): 641-9

8. Andersen TF, Madsen M, Jorgensen J, et al. The Danish National Hospital Register: a valuable source of data for modern health sciences. Dan Med Bull 1999 Jun; 46 (3): 263-8

9. Norén GN, Hopstadius J, Bate A, et al. Temporal pattern discovery in electronic patient records. Data Min Knowl Discov 2010; 20 (3): 361-87

10. DuMouchel W. Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. Am Stat 1999; 3 (53): 177-202

11. Bate A, Lindquist M, Edwards IR, et al. A Bayesian neural network method for adverse drug reaction signal generation. Eur J Clin Pharmacol 1998 Jun; 54 (4): 315-21

12. Vestergaard M, Hviid A, Madsen KM, et al. MMR vacci-
nation and febrile seizures: evaluation of susceptible sub-
groups and long-term prognosis. JAMA 2004 Jul 21;
292 (3): 351-7

13. Miller E, Waight P, Farrington CP, et al. Idiopathic
thrombocytopenic purpura and MMR vaccine. Arch Dis
Child 2001 Mar; 84 (3): 227-9

14. Jefferson T, Price D, Demicheli V, et al. Unintended events
following immunization with MMR: a systematic review.
Vaccine 2003 Sep 8; 21 (25-26): 3954-60

15. Plotkin SA, Reef ER. Rubella vaccine. In: Plotkin SA,
Orenstein WA, Offit PA, editors. Vaccines. 5th ed. Phila-
delphia (PA): Saunders, 2008: 735-71

16. Strebel PM, Papania MJ. Measles vaccine. In: Plotkin SA,
Orenstein WA, Offit PA, editors. Vaccines. 5th ed. Phila-
delphia (PA): Saunders, 2008: 353-98

17. Edwards R, Faich G, Tilson H. Points to consider: the roles
of surveillance and epidemiology in advancing drug safety.
Pharmacoepidemiol Drug Saf 2005 Sep; 14 (9): 665-7

18. Hazell L, Shakir SA. Under-reporting of adverse drug re-
actions: a systematic review. Drug Saf 2006; 29 (5): 385-96

19. Stephenson WP, Hauben M. Data mining for signals in
spontaneous reporting databases: proceed with caution.
Pharmacoepidemiol Drug Saf 2007 Apr; 16 (4): 359-65

20. Brewer T, Colditz GA. Postmarketing surveillance and ad-
verse drug reactions: current perspectives and future needs.
JAMA 1999 Mar 3; 281 (9): 824-9

21. Hauben M, Patadia V, Gerrits C, et al. Data mining in
pharmacovigilance: the need for a balanced perspective.
Drug Saf 2005; 28 (10): 835-42

22. Almenoff JS, Pattishall EN, Gibbs TG, et al. Novel statis-
tical tools for monitoring the safety of marketed drugs.
Clin Pharmacol Ther 2007 Aug; 82 (2): 157-66

23. Brown JS, Kulldorff M, Chan KA, et al. Early detection of
adverse drug events within population-based health net-
works: application of sequential testing methods. Pharmaco-
epidemiol Drug Saf 2007 Dec; 16 (12): 1275-84

24. Greene SK, Kulldorff M, Lewis EM, et al. Near real-time
surveillance for influenza vaccine safety: proof-of-concept
in the Vaccine Safety Datalink Project. Am J Epidemiol
2010 Jan 15; 171 (2): 177-88

25. Davis RL, Kolczak M, Lewis E, et al. Active surveillance of
vaccine safety: a system to detect early signs of adverse
events. Epidemiology 2005 May; 16 (3): 336-41

26. Belongia EA, Irving SA, Shui IM, et al. Real-time surveil-
lance to assess risk of intussusception and other adverse
events after pentavalent, bovine-derived rotavirus vaccine.
Pediatr Infect Dis J 2010 Jan; 29 (1): 1-5

27. Lieu TA, Kulldorff M, Davis RL, et al. Real-time vaccine
safety surveillance for the early detection of adverse events.
Med Care 2007 Oct; 45 (10 Suppl. 2): S89-95

28. Hocine MN, Musonda P, Andrews N, et al. Sequential case
series analysis for pharmacovigilance. J R Stat Soc Ser A
2009; 172 (1): 213-36

29. Walker AM. Pattern recognition in health insurance claims
databases. Pharmacoepidemiol Drug Saf 2001 Aug; 10 (5):
393-7

30. Platt R, Madre L, Reynolds R, et al. Active drug safety
surveillance: a tool to improve public health. Pharmaco-
epidemiol Drug Saf 2008 Dec; 17 (12): 1175-82

Correspondence: *Henrik Svanström*, Department of Epide-
miology Research, Statens Serum Institut, Artillerivej 5,
2300 Copenhagen, Denmark.
E-mail: htr@ssi.dk