Drug Saf (2014) 37:163–170
DOI 10.1007/s40264-014-0137-z

ORIGINAL RESEARCH ARTICLE

Signalling Paediatric Side Effects using an Ensemble of Simple Study Designs

Jenna M. Reps · Jonathan M. Garibaldi · Uwe Aickelin · Daniele Soria · Jack E. Gibson · Richard B. Hubbard

Published online: 19 February 2014

Abstract

Background Children are frequently prescribed medication 'off-label', meaning there has not been sufficient testing of the medication to determine its safety or effectiveness. The main reason this safety knowledge is lacking is due to ethical restrictions that prevent children from being included in the majority of clinical trials.

Objective The objective of this paper is to investigate whether an ensemble of simple study designs can be implemented to signal acutely occurring side effects effectively within the paediatric population by using historical longitudinal data. The majority of pharmacovigilance techniques are unsupervised, but this research presents a supervised framework.

Methods Multiple measures of association are calculated for each drug and medical event pair and these are used as features that are fed into a classifier to determine the likelihood of the drug and medical event pair corresponding to an adverse drug reaction. The classifier is trained using known adverse drug reactions or known non-adverse drug reaction relationships.

Results The novel ensemble framework obtained a false positive rate of 0.149, a sensitivity of 0.547 and a specificity of 0.851 when implemented on a reference set of drug and medical event pairs. The novel framework consistently outperformed each individual simple study design.

Conclusion This research shows that it is possible to exploit the mechanism of causality and presents a framework for signalling adverse drug reactions effectively.

Key Points

The ensemble of simple study designs outperformed each single simple study design when considering both the overall ability to rank adverse drug reactions and the signalling performance at a natural threshold.

The ensemble method is adaptable as it can incorporate any new measures of association that are proposed over time.

The results of the paper highlight the potential benefit of applying supervised learning for adverse drug reaction signalling.

1 Introduction

There is an abundance of evidence to support the impression that side effects in children currently present a significant public health problem [1, 2]. When there is a causal relationship between a drug and medical event it is termed an adverse drug reaction (ADR). Children of all ages can suffer from diseases that require them to take medication but the suitability of drugs in the paediatric population (0–17 years old) is generally unexplored. The majority of paediatric prescriptions are 'off-label' meaning the licensed medication is used in situations that have not had sufficient investigation to determine the drug efficiency or

Electronic supplementary material The online version of this article (doi:10.1007/s40264-014-0137-z) contains supplementary material, which is available to authorized users.

J. M. Reps (✉) · J. M. Garibaldi · U. Aickelin · D. Soria
IMA, The University of Nottingham,
Nottingham NG8 1BB, UK
e-mail: psxjr1@nottingham.ac.uk

J. E. Gibson · R. B. Hubbard
Community Health Sciences, Clinical Sciences Building,
Nottingham NG5 1PB, UK

safety. Examples of off-label prescriptions are prescribing a different dosage or frequency than recommended, prescribing a drug for a different indication than the drug was tested for or prescribing the drug to age groups such as children where the drug has not been extensively evaluated. The paediatric population are rarely involved in clinical trials, so there is little evidence available to determine if a medication is efficient and safe [3] and many drugs do not have licenses for use in children. A study that observed a paediatric hospital in Derby, UK found that 23 % of the prescribed drugs were off-label and this was lower than the off-label rate observed in four other European hospitals [2].

The problem with 'off-label' prescribing within the paediatric population is that there are clear physiological differences between adults and children, so the efficiency and safety knowledge discovered during clinical trials in the adult population cannot be accurately extrapolated for the paediatric population [4]. Consequently, there has been a recent demand for incorporating more children into randomised controlled trials [5] so drug efficiency and safety can be directly evaluated for children. In addition to being able to assess the efficiency of drugs within the paediatric population and determine suitable dosages, there are also advantages for the children enrolled in trials such as access to new medicine that may reduce morbidity. However, there are also many downsides including both physical and mental discomfort, separation from parents [5] and the standard risks associated with adult clinical trials [6]. These downsides may be magnified in children because of potential errors in initially estimating 'adult equivalent' doses, and because of additional impacts of drugs on still-developing tissues. In the worse case, a clinical trial could result in child mortalities. Therefore, if possible, it is more preferable to develop alternative means to identify ADRs that do not have these negative effects.

Pharmacovigilance is the study of prescription drug ADRs, including the collection of suitable data, and their detection and prevention. As the clinical trial data for the paediatric population is lacking, a key resource for paediatric pharmacovigilance is the spontaneous reporting system database [7, 8]. The spontaneous reporting systems amalgamate the suspected cases of ADRs that are voluntarily reported within a population. For example, in the UK if a patient is prescribed a drug and experiences an unexpected medical event then their doctor or the patient can report the suspected ADR via the yellow card scheme by filling out a form. All the yellow card scheme reports are then combined into a spontaneous reporting system database that is used to identify ADRs. As the database only contains reports detailing cases when a drug is taken and a suspected ADR occurs, and not the cases when a drug is taken and no ADRs is suspected, the frequency that drugs are prescribed is unknown. Furthermore, as the reporting is done voluntarily, it is common for data to be missing, incorrect or duplicated [9]. Consequently, there are no current algorithms that can be applied to spontaneous reporting system databases that are capable of detecting ADRs with a high accuracy, nor are the algorithms able to quantify the frequency that the ADRs occur. It is the combination of a lack of clinical trials coupled with the general spontaneous reporting system limitations that makes the paediatric population potentially more susceptible to ADRs. Research into developing novel algorithms that are able to efficiently and effectively discover qualitative and quantitative ADR information for the paediatric population is required to reduce child morbidities and mortalities.

A new type of database, called the longitudinal observational database, has started to emerge as a potentially new source of ADR information. The longitudinal observational database contains sequences of patients' medical data often spanning decades and offers a new perspective for ADR discovery. One example of a longitudinal observational database is The Health Improvement Network (THIN) database that contains medical records extracted directly from general practitioners' databases across the UK (http://www.thin-uk.com). The THIN database contains validated personal information about each patient in the database including their year of birth and gender. There is also a complete medical record and prescription history for each patient during the time they have been registered at the practice. However, patients may not inform their doctors of all the medical events they experience, especially minor ones, and use of over-the-counter medication for self-treatment is unlikely to be recorded. As the THIN database contains information about how many patients are prescribed a drug, it may be possible to extract quantitative information about ADRs. The effect of dosage and frequency of prescription could also be investigated to identify the optimal treatment for each child. This knowledge could then be used to help personalise medication for paediatric patients on the basis of their medical state, age and gender.

Several methods have been presented to identify ADR signals using longitudinal observational databases, although comparisons have concluded that the methods generally have a high false positive rate [10, 11]. These algorithms include cohort techniques [12], case-series approaches [13], case–control approaches [14], disproportionality analysis [15] or a mixture of the previously mentioned techniques [16]. In addition to being limited by a high false positive rate, many of these algorithms require the use of the patient's medical history years prior to the drug prescription of interest and this limits their ability for use on the paediatric population as a long medical history is often not available (if a child is very young). The case–control method compares the prevalence of the drug within the population of patients experiencing the medical event and the prevalence of the drug within a population of patients that have similar covariates but do not experience the medical event. However, the case–control technique applied to the paediatric population is likely to introduce confounding as children experiencing a medical event may have a serious disorder that makes them susceptible to other illnesses, whereas children not experiencing the medical event could be very healthy. Similarly, comparing children who take a drug with children not taking a drug may introduce confounding by indication [17] as children who do not take any medication are likely to be very healthy due to a lack of chronic or degenerative illnesses, whereas children who require certain medication are likely to be very unhealthy. It follows that a novel method that does not suffer from confounding by indication or require long periods of historic medical history should be more effective at signalling ADRs.

Recently, a supervised framework [18] that learns from known ADRs has been proposed to signal ADRs by using features based on the Bradford Hill causality criteria, criteria that are frequently used to determine causality. This framework was shown to perform well, and obtained a low false positive rate even when the frequency of the ADR was low. Unfortunately, the calculation of many of the Bradford Hill features requires knowledge of a long medical history and this is often not available for paediatric patients. Consequently, the framework using Bradford Hill criteria-based features faces difficulties when applied to detect ADRs within the paediatric population. One possible solution would be to implement an analogous framework that uses features based on the counterfactual method for causal inference, as these features can be chosen such that a large medical history is not required. This may enable rapid detection of paediatric ADRs and the framework may yield a low false positive rate. The main benefit of this method is that it does not have the risks associated with clinical trials.

In this paper we aim to investigate whether an ensemble of specifically chosen simple study designs can signal acute ADRs within the paediatric population with a low false positive rate. A comparison will be implemented to determine whether the ensemble offers an improvement over each individual simple study design. The ensemble requires generating multiple distinct measures of association between a drug and medical event on the basis of the counterfactual method for causal inference. However, each measure of association will be chosen such that a distinct type of main confounding effect will be introduced. The motivation of combining these measures of association via an ensemble is that it may be possible to exploit any causal mechanism structures. Using drug and medical event pairs definitively known to represent ADRs or non-ADRs, we calculate the various measures of association for each drug and medical event pair to generate the labelled data used to train a random forest classifier. The measures of association for any drug and medical event pair with an unknown ADR status can then be calculated and fed into the trained random forest to determine whether the pair corresponds to an ADR.

The objective of this paper is to investigate whether an ensemble of simple study designs can be implemented to signal acutely occurring side effects effectively within the paediatric population by using historical longitudinal data. The majority of pharmacovigilance techniques are unsupervised, but this research presents a supervised framework.

2 Material

The THIN database contains temporal medical and therapy records for over 11 million UK patients. We used a subset of the THIN database for this research which contained records for 4 million patients. Within the subset, there were a total of 30,191,726 medical events recorded for 1.7 million patients when they were 17 years old or less. For each patient, their year of birth, gender and other personal details are known. Each medical record specifies the patient that the record corresponds to, the record date and the medical event experienced by the patient. The medical events have a tree structure, with the medical event becoming more specific as its node depth (i.e. the length of the path from the root to the node) increases. Therefore, medical event nodes with a depth of 1 are the most general and medical event nodes with a depth of 5 are the most specific.

Each prescription within the THIN database contains details about the specific drug prescribed and contains a code corresponding to the drug known as the British National Formulary (BNF) code [19]. The BNF code has a hierarchal structure that can be used to identify similar drugs. For example, BNF codes starting with 05 (e.g. 05-xx-xx-xx) correspond to drugs used to treat infections, and BNF codes starting with 05-01-01 (e.g. 05-01-01-xx) correspond to penicillins. A drug family is the set of drugs with the same BNF code. For example, the drug family benzylpenicillin sodium and phenoxymethylpenicillin have a corresponding BNF code of 05-01-01-01, the drug family penicillinase-resistant penicillins have a BNF code of 05-01-01-02 and the drug family broad-spectrum penicillins have a BNF code of 05-01-01-03. These are the three drug families used to evaluate the novel framework presented in this paper. The number of records for each of the drug families in the THIN database is presented in Table 1. These drug families were chosen as they are frequently prescribed, so their ADRs are generally well known and the creation of a reference set of definitive ADRs and non-ADRs was possible.

Table 1 Details about the records within the subset of The Health Improvement Network (THIN) database for a selection of three penicillin drugs prescribed to patients aged 17 years or less

Drug family	BNF[a]	Number of prescriptions	
		Total	First in 3 months
Benzylpenicillin sodium and phenoxymethylpenicillin	05-01-01-01	1,520,866	456,926
Penicillinase-resistant penicillins	05-01-01-02	310,622	252,947
Broad-spectrum penicillins	05-01-01-03	6,490,455	1,448,563

[a] British National Formulary code

3 Methodology

3.1 Ensemble of Simple Studies Design Framework Overview

The proposed ensemble of simple studies design (ESSD) framework for signalling the acute ADRs that occur within the paediatric population is:

1. Generate simple studies labelled data
   - Choose n drug families of interest, where each drug family is denoted by $D_k, k \in [1, n]$.
   - Determine the risk medical events (RME_{D_k}) i.e. all the medical events that are potential acutely occurring ADRs to the drugs in D_k.
   - For each drug family D_k and medical event $\in \mathrm{RME}_{D_k}$ pair, determine whether the medical event is a known ADR or non-ADR of the drug family D_k and add labels. The label for the ith pair is denoted by y_i. For example, if the ith pair corresponds to an ADR then $y_i = 1$, but if the ith pair corresponds to a non-ADR then $y_i = 0$.
   - Generate the features for each pair ($D_k +$ event $\in \mathrm{RME}_{D_k}$) by applying the simple study designs to calculate multiple estimated causal effect values (the measure of association). The feature vector for the ith pair is denoted by $\mathbf{x_i}$.
2. Train a random forest model using the labelled data ($\{(\mathbf{x_i}, y_i)\}$)
   - Apply 20-fold cross-validation to tune the random forest classifier.
   - Select the optimal model parameter by considering the classifier's general ability to rank pairs corresponding to ADRs above pairs corresponding to non-ADRs.
3. Apply the trained random forest classifier to the simple study design features of any unlabelled drug family and medical event pair (not known to correspond to an ADR or non-ADR) and classify the pair as an ADR or non-ADR.

3.1.1 Risk Medical Events

For each drug family D_k, the medical events investigated are determined using temporal information. As we are interested in acutely occurring ADRs, we restrict our attention to only investigate medical events that are observed within the month after a prescription of any drug within D_k is first prescribed. A month was chosen to be a suitable trade-off to enable a sufficient amount of time for the patient to report the medical event while not introducing a surplus quantity of noise. Therefore, given a drug family D_k, the risk medical events of D_k (RME_{D_k}) are defined as the set of all medical events that are observed for a minimum of three patients within the month after any prescription of a drug within D_k. We chose to add a limit of three or more patients experiencing the potential ADR as it is difficult to determine whether a medical event is an ADR if it is experienced by less than three patients.

3.1.2 Generating Features

For each drug family (D_k) and medical event $\in \mathrm{RME}_{D_k}$ pair, we extract six different estimates of the causal effect of D_k on the medical event. These are the six simple study designs. The target population is the patients prescribed D_k and the etiological time period (the period we investigate) is the month after the first prescription. The estimates of the causal effect are calculated by either using a different target population (target substitution) or using a different etiological time period in (etiological substitution).

x^1: Etiological substitution (SSD_1)—The causal effect is approximated by comparing the risk of the medical event during the month after the prescription for the target population with the risk during the month before the prescription for the target population. The main confounding effect is caused by a covariate of 'medical state' as the target population medical states are likely to change between the month before and the month after the prescription. This causal effect estimate is likely to be large for progressive medical events (e.g. progressions of the cause of taking the drug) even though they are not caused by the prescription.

x^2: Etiological substitution (SSD_2)—The causal effect is approximated by comparing the risk of the medical event during the month after the prescription for the target population with the risk during the year after for the target population. The main confounding effect is a covariate of 'medical sate', but unlike x^1 the causal effect estimate is likely to be small for progressive medical events that become more common as the population ages and large for medical events that occurred acutely after the drug.

x^3: Target substitution (SSD$_3$)—The causal effect is approximated by comparing the risk of the medical event during the month after the prescription for the target population with the risk during a randomly chosen month for a substitute population that matches the target population on age and gender. The main confounding effect is caused by a covariate of 'indication' as the target population all have certain illnesses causing them to require the drug but the substitute population does not. This causal effect estimate is likely to be large for medical events linked to the indication (i.e. the cause of taking the drug).

x^4: Target substitution (SSD$_4$)—The causal effect is approximated by comparing the risk of the medical event during the month after the prescription for the target population with the risk during the month after for a substitute population that are given a similar drug (i.e. a similar BNF code) and have the same indications. The main confounding effects are caused by a covariate of 'medical caution' as patients may be prescribed different drugs for the same indication owing to medical caution (i.e. one patient has kidney issues preventing them from having the standard medication) or covariates of 'age and gender' as age and gender may influence the choice of drug prescribed.

x^5: Etiological substitution (SSD$_5$)—First, a mapping is performed to 'generalise' the medical event descriptions. This is done by mapping medical event nodes that have a depth greater than 3 to their depth 3 'parent node'. The causal effect is then approximated using the mapped data by comparing the risk of the corresponding depth 3 'parent node' medical event during the month after the prescription for the target population with the risk during the month before for the target population. This causal effect measure is less vulnerable to covariates of 'medical event recording', where different medical events that correspond to the same/similar illness can be recorded because of redundancy or illness progression.

x^6: Etiological substitution (SSD$_6$)—First, a mapping is performed to 'generalise' the medical event descriptions. This is done by mapping medical event nodes that have a depth greater than 4 to their depth 4 'parent node'. The causal effect is then approximated by comparing the risk of the corresponding depth 4 'parent node' medical event during the month after the prescription for the target population with the risk during the month before for the target population. This causal effect measure is less vulnerable to a covariate of 'medical event recording', where different medical events that correspond to the same/similar illness can be recorded due to redundancy or illness progression.

The vector $\mathbf{x}_{\ i} = (x_i^1, x_i^2, \ldots, x_i^6)$ contains the six estimates of the causal effect for the ith drug family and medical event pair. We also create three additional features from the original,

$$x_i^7 = \begin{cases} x_i^1/x_i^2 & \text{if} \quad |x_i^2| > 0 \\ x_i^1 & \text{else} \end{cases}$$

$$x_i^8 = \begin{cases} x_i^1/x_i^4 & \text{if} \quad |x_i^4| > 0 \\ x_i^1 & \text{else} \end{cases}$$

$$x_i^9 = \begin{cases} x_i^1/x_i^5 & \text{if} \quad |x_i^5| > 0 \\ x_i^1 & \text{else} \end{cases}$$

These additional features indicate how much the simple study design association measures deviate when considering time, similar patients or the specificity of the medical event. So the complete feature vector for each drug family and medical event pair is $\mathbf{x}_i = (x_i^1, x_i^2, \ldots, x_i^9) \in \mathbb{R}^9$.

3.1.3 Random Forest Classifier

A random forest is a supervised classifier. The task of supervised learning is to use the training data to learn a mapping between the feature vector and the class. This mapping can then be used to predict the class for unseen data. The random forest is known as a ensemble classifier as it trains and combines weak and diverse classifiers. Each weak classifier is a decision tree that uses a subset of the available features (the simple design study measures of association) to predict the class (ADR or non-ADR). The advantages of the random forest classifier are that it can have features that are both discrete and continuous and does not require the features to be pre-processed (e.g. centred and scaled). The parameter of the random forest that needs to be chosen is the number of features that each decision tree can use, which is referred to as *mtry*. In this research we used the R implementation of the random forest [20].

The random forest is trained using the labelled drug family and medical event pair data, $X^L = \{(\mathbf{x}_i, y_i) \mid \text{the label is known}\}$. We applied 20-fold cross-validation; this means that the data are partitioned into 20 sets and for each set the random forest is trained on the other 19 sets and then applies to predict the class of each data point within the set. The trained random forest is then evaluated by user-defined criteria, in our case the area under of receiver operating characteristic curve (AUC), and the average value corresponding to this measure over the 20 sets determines how well the random forest has performed. This performance measure is used to select the parameter *mtry*, as various random forests are trained with difference values of *mtry* and the *mtry* that results in the highest AUC is selected. When we refer to the trained random forest we mean the random forest that has been trained using the *mtry* value that was optimal.

3.2 Evaluation

To evaluate the ESSD framework, we created a reference set of drug families and medical events pairs that are known to be ADRs or non-ADRs. The drug families used to create the reference set are the penicillins with the BNF codes 05-01-01-01, 05-01-01-02 and 05-01-01-03. The reference is used to train the random forest and evaluate it. The reference set data corresponding to two of the drug families is used as labelled data to train the random forest and the reference set data corresponding to the remaining drug family is used for evaluation.

3.2.1 Creating the Reference Set

The reference set was created by investigating all the RME_{D_k} for each D_k. Medical events that are listed as side effects on the BNF website or the medical event states the occurrence of an adverse event or the medical event is candidiasis as antibiotics are known to cause this were labelled as ADRs. Any medical event with a cause that is known and is not related to the penicillins (e.g. impetigo, worms, diabetes) was labelled as non-ADRs and a selection of medical events that are likely to be related to the cause of taking the drug were also labelled as non-ADRs. The labels corresponding to medical events that are likely to be related are the most likely to be incorrect out of all the labels as it is difficult to show that a medical event is not an ADR, as even events that cause the drug to be taken could also be ADRs.

3.2.2 Evaluation Measures

The ESSD framework is evaluated by considering its ability to signal ADRs at its natural threshold and its ability to rank drug family and medical event pairs by how likely they correspond to ADRs. The ESSD is compared with each individual simple study design (SSD_i).

The natural threshold evaluation measures are the number of true positives (TP), which is the number of pairs corresponding to ADRs that the method classes as ADRs, and the number of false positives (FP), which is the number of pairs corresponding to non-ADRs that the method classes as ADRs. Similarly, the number of false negatives (FN) is the number of pairs that correspond to ADRs but the method classes as non-ADRs and the number of true negatives (TN) is the number of pairs that correspond to non-ADRs and the method classes as non-ADRs. The natural threshold measures can be calculated as

$$\begin{aligned} \text{Sensitivity} &= \mathrm{TP}/(\mathrm{TP}+\mathrm{FN}) \\ \text{Specificity} &= \mathrm{TN}/(\mathrm{FP}+\mathrm{TN})\,. \\ \text{False positive rate (FPR)} &= \mathrm{FP}/(\mathrm{FP}+\mathrm{TN}) \end{aligned} \qquad (1)$$

The general ranking ability of the methods are measured by the average precision (AP) and the AUC. The AUC is the area under the curve of the sensitivity plotted against 1 minus the specificity for various thresholds. The AUC can be interpreted as the probability of a uniformly chosen ADR pair being ranked above a uniformly chosen non-ADR pair. If a method performs poorly at its natural threshold but has a high AUC, then this may mean the natural threshold needs to be modified.

4 Results

The ESSD framework was evaluated three times. Each time the reference set data for two of the drug families were used to train the random forest and the reference set data for the third drug family was used for evaluation. A table containing the full reference set data and the confidences returned by the ESSD framework can be found in the electronic supplementary material. A summary of each of the three evaluations is presented in Table 2. The optimal *mtry* obtained when training the model is presented and also the result of the AUC for the cross-validation.

At their natural thresholds, the ESSD had a lower overall FPR of 0.149 compared to the other methods that had FPRs between 0.184 and 0.716. The ESSD had a sensitivity of 0.547. Although other methods had a higher sensitivity, they also had a very high false positive rate ($\geq$0.532) which is not desirable. The results of methods at their natural thresholds for each individual evaluation are presented in Table 3 and the overall results with the sensitivity, specificity and false positive rate are displayed in Table 4.

The general ranking ability of the ESSD was consistently higher than the other methods with AUC values of 0.814, 0.806 and 0.813 for the evaluations 1–3 respectively. The highest AUC value out of all the other methods was 0.813,

Table 2 Details of the evaluation experiments

Evaluation	Training/ testing set	Optimal *mtry*	Training AUC[a]	Evaluation set
1	05-01-01-{02, 03}	4	0.885	05-01-01-01
2	05-01- 01-{01, 03}	6	0.827	05-01-01-02
3	05-01-01-{01, 02}	9	0.875	05-01-01-03

[a] Area under the ROC curve obtained by the 20-fold cross-validation

Table 3 Number of TP, FP, FN and TN returned for the ensemble of simple study designs (ESSD) and each individual simple study design (SSD_i)

Method	1				2				3			
	TP	FP	FN	TN	TP	FP	FN	TN	TP	FP	FN	TN
ESSD	9	9	8	30	9	6	9	40	17	6	12	50
SSD_1	17	21	0	18	15	31	3	15	26	33	3	23
SSD_2	17	22	0	17	18	42	0	4	29	37	0	19
SSD_3	5	4	12	35	5	13	13	33	17	17	12	39
SSD_4	2	4	15	35	2	17	16	29	4	5	25	51
SSD_5	17	19	0	20	17	26	1	20	25	30	4	26
SSD_6	17	19	0	20	15	29	3	17	24	31	5	25

TP true positive, *FP* false positive, *FN* false negative, *TN* true negative

Table 5 Comparison of the general ADR ranking ability of the ensemble of simple study designs (ESSD) and each individual simple study design (SSD_i)

Method	1		2		3		Average	
	AUC	AP	AUC	AP	AUC	AP	AUC	AP
ESSD	0.814	0.615	0.806	0.659	0.813	0.712	0.811	0.662
SSD_1	0.797	0.559	0.606	0.362	0.678	0.457	0.694	0.459
SSD_2	0.785	0.548	0.741	0.430	0.729	0.493	0.752	0.490
SSD_3	0.602	0.499	0.484	0.292	0.629	0.433	0.572	0.408
SSD_4	0.477	0.316	0.322	0.217	0.459	0.333	0.419	0.289
SSD_5	0.797	0.558	0.794	0.508	0.698	0.448	0.763	0.505
SSD_6	0.813	0.587	0.621	0.356	0.655	0.435	0.696	0.459

AUC area under the ROC curve, AP average precision

Table 4 Average number of TP, FP, FN and TN returned for the ensemble of simple study designs (ESSD) and each individual simple study design (SSD_i) and the overall specificity, sensitivity and false positive rate (FPR)

Method	TP	FP	FN	TN	Sensitivity	Specificity	FPR
ESSD	35	21	29	120	0.547	0.851	0.149
SSD_1	58	85	6	56	0.906	0.397	0.603
SSD_2	64	101	0	40	1	0.284	0.716
SSD_3	27	34	37	107	0.422	0.759	0.241
SSD_4	8	26	56	115	0.125	0.816	0.184
SSD_5	59	75	5	66	0.922	0.468	0.532
SSD_6	56	79	8	62	0.875	0.440	0.560

TP true positive, *FP* false positive, *FN* false negative, *TN* true negative

0.794 and 0.729 for the evaluations 1–3 respectively. The AP value obtained by the ESSD was also greater for each evaluation, with the ESSD being 0.615, 0.659 and 0.717 for evaluation 1–3 respectively, whereas the highest AP obtained by any other method over evaluation 1–3 was 0.587, 0.508 and 0.493 respectively. The results of the general ranking ability are presented in Table 5.

5 Discussion

The results show that the ESSD framework consistently outperformed the individual simple study design measures. It consistently had a higher AP and AUC for all three BNF families investigated. At its natural threshold the ESSD framework was able to signal just over half the true ADRs while only signalling approximately 15 % of the non-ADRs. However, as it is difficult to prove a medical event is not an ADR, some of the non-ADR labels may be incorrect, so the probability of signalling an non-ADRs may actually be lower (i.e. the FPR is probably less than the value obtained within this research). The ESSD framework is also efficient, as the simple measures can be calculated readily and the classifier can be trained quickly. Once trained, the prediction is fast and could be implemented regularly when new data is added to the longitudinal database. As the ESSD has a low false positive rate and is efficient, it may be a useful framework to implement for signal generation. However, the signals that are generated will still need to be refined as it does not have a zero false positive rate.

There have been few attempts to apply supervised learning to the field of pharmacovigilance but this is an interesting area, as the results of a supervised classifier can be improved when more labelled data is available. Therefore, supervised methods should improve over time as more ADR knowledge is gained. It may be possible to feedback the results of the supervised methodologies to further improve them via methods such as semi-supervised techniques. The ESSD also has the advantage of being adaptable, as it could incorporate new measures of association that get proposed over time as features.

In this paper we have trained the ESSD using similar BNFs to the BNF used for the evaluation. The justification for this is that similar BNF drugs are likely to have similar underlying patterns. It would be interesting to see whether the random forest within the ESSD framework could be trained on a variety of different BNFs and still perform well. This would indicate whether the underlying structures are independent of the drug.

One difficulty that was noticed with the ESSD framework is the choice of BNF to use to calculate the similar BNF measure of association (SSD_4). For the penicillins that was easy as the BNF codes are numerous; however, for BNF families such as the proton pump inhibitors, a closely resembling BNF family is difficult to determine and the wrong choice could lead to different models with various performances. To overcome this the methodology may need to be applied to individual drugs rather than BNF families, but this could be problematic when the drug is rarely prescribed and therefore rarely recorded within the database. However, with the combination of longitudinal healthcare databases now becoming common [21], even rarely prescribed drugs should have a sufficient number of occurrences in the combined database.

6 Conclusion

In this paper we have proposed a novel framework, called the ensemble of simple study designs (ESSD), specifically for signalling acute ADRs within the paediatric population. The framework does not require knowledge of a patient's medical history of more than a month prior to the prescription. The results show that the ESSD can outperform each individual simple study design measure and appears to be more consistent. The ensemble still misclassifies some non-ADRs but it had a false positive rate of 0.149, making it competitive with existing methods. The advantage of this methodology is that it is supervised, so as new ADRs are discovered its performance should increase as more labelled data will be available for training it.

Future work could involve researching new methods for refining the ADR signals that are generated and reducing the false positive rate further or investigating different features based on alternative substitutions for the hypothetical counterfactual situation. For example, features that deal with alternative forms of confounding such as the time of the year could be incorporated. It is also of interest to see whether adding more complex study design methods such as Temporal Pattern Discovery [16] or HUNT [12] can improve the ensemble and a comparison between these methods and the ESSD would be useful.

Acknowledgments This study was funded by the Department of Computer Sciences, University of Nottingham, UK. J.M. Reps, J.M. Garibaldi, U. Aickelin, D. Soria, J.E. Gibson and R.B. Hubbard have no conflicts of interest that are directly relevant to the content of this study.

References

1. Aagaard L, Christensen A, Hansen EH. Information about adverse drug reactions reported in children: a qualitative review of empirical studies. Br J Clin Pharmacol. 2010;70(4):481–91.
2. Conroy S, Choonara I, Impicciatore P, Mohn A, Arnell H, Rane A, Knoeppel C, Seyberth H, Pandolfini C, Raffaelli MP, et al. Survey of unlicensed and off label drug use in paediatric wards in European countries. Br Med J. 2000;320(7227):79–82.
3. Rose K, van den Anker JN. Guide to paediatric drug development and clinical research. Basel: Karger; 2010.
4. Bwakura-Dangarembizi M, Musesengwa R, Nathoo KJ, Takaidza P, Mhute T, Vhembo T. Ethical and legal constraints to childrens participation in research in Zimbabwe: experiences from the multicenter pediatric HIV ARROW trial. BioMed Central Med Ethics. 2012;13(1):17.
5. Caldwell PH, Murphy SB, Butow PN, Craig JC. Clinical trials in children. Lancet. 2004;364(9436):803–11.
6. Lidz CW, Appelbaum PS, Grisso T, Renaud M. Therapeutic misconception and the appreciation of risks in clinical trials. Soc Sci Med. 2004;58(9):1689–97.
7. Avery A, Anderson C, Bond C, Fortnum H, Gifford A, Hannaford P, Hazell L, Krska J, Lee A, McLernon D, et al. Evaluation of patient reporting of adverse drug reactions to the UK 'Yellow card scheme': Literature review, descriptive and qualitative analyses, and questionnaire surveys. Health Technol Assess. 2011;15(20):1–234.
8. Aagaard L. Knowledge creation about adverse drug reactions in the paediatric population. Ugeskr Laeger. 2013;175(6):342–45.
9. Bate A, Evans S. Quantitative signal detection using spontaneous ADR reporting. Pharmacoepidemiol Drug Saf. 2009;18(6): 427–36.
10. Ryan PB, Madigan D, Stang PE, Marc-Overhage J, Racoosin JA, Hartzema AG. Empirical assessment of methods for risk identification in healthcare data: results from the experiments of the Observational Medical Outcomes Partnership. Stat Med. 2012;31(30):4401–15.
11. Reps JM, Garibaldi JM, Aickelin U, Soria D, Gibson J, Hubbard R. Comparison of algorithms that detect drug side effects using electronic healthcare databases. Soft Comput. 2013;17(12): 2381–97. doi:10.1007/s00500-013-1097-4.
12. Jin H, Chen J, He H, Kelman C, McAullay D, O'Keefe CM. Signaling potential adverse drug reactions from administrative health databases. IEEE Trans Knowl Data Eng. 2010;22(6): 839–53.
13. Simpson SE. Self-controlled methods for postmarketing drug safety surveillance in large-scale longitudinal data. Ph.D. thesis, Columbia University; 2011.
14. Harpaz R, DuMouchel W, Shah NH, Madigan D, Ryan P, Friedman C. Novel data-mining methodologies for adverse drug event discovery and analysis. Clin Pharmacol Ther. 2012;91(6): 1010–21.
15. Zorych I, Madigan D, Ryan P, Bate A. Disproportionality methods for pharmacovigilance in longitudinal observational databases. Stat Methods Med Res. 2013;22(1):39–56.
16. Norén GN, Hopstadius J, Bate A, Star K, Edwards IR. Temporal pattern discovery in longitudinal electronic patient records. Data Min Knowl Discov. 2010;20(3):361–87.
17. Greenland S, Neutra R. Control of confounding in the assessment of medical technology. Int J Epidemiol. 1980;9(4):361–67.
18. Reps JM, Garibaldi JM, Aickelin U, Soria D, Gibson JE, Hubbard RB. A novel prescription drug side effect classifier. IEEE Trans Knowl Data Eng. 2014 (submitted)
19. Joint Formulary Committee. British National Formulary, 66 ed. London: BMJ Group and Pharmaceutical Press; 2013.
20. Liaw A, Wiener M. Classification and regression by random forest. R News. 2002;2(3):18–22. http://CRAN.R-project.org/doc/Rnews/
21. Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. J Am Med Inf Assoc. 2012;19(1):54–60.