

Comparing data mining methods on the VAERS database[†]

David Banks PhD[‡], Emily Jane Woo MD, MPH*, Dale R. Burwen MD, MPH, Phil Perucci BS, M. Miles Braun MD, MPH and Robert Ball MD, MPH, ScM

The Office of Biostatistics and Epidemiology, Center for Biologics Evaluation and Research, Food and Drug Administration, Rockville, MD, USA

SUMMARY

Purpose Data mining may enhance traditional surveillance of vaccine adverse events by identifying events that are reported more commonly after administering one vaccine than other vaccines. Data mining methods find signals as the proportion of times a condition or group of conditions is reported soon after the administration of a vaccine; thus it is a relative proportion compared across vaccines, and not an absolute rate for the condition. The Vaccine Adverse Event Reporting System (VAERS) contains approximately 150 000 reports of adverse events that are possibly associated with vaccine administration.

Methods We studied four data mining techniques: empirical Bayes geometric mean (EBGM), lower-bound of the EBGM's 90% confidence interval (EB05), proportional reporting ratio (PRR), and screened PRR (SPRR). We applied these to the VAERS database and compared the agreement among methods and other performance properties, particularly focusing on the vaccine–event combinations with the highest numerical scores in the various methods.

Results The vaccine–event combinations with the highest numerical scores varied substantially among the methods. Not all combinations representing known associations appeared in the top 100 vaccine–event pairs for all methods.

Conclusions The four methods differ in their ranking of vaccine–COSTART pairs. A given method may be superior in certain situations but inferior in others. This paper examines the statistical relationships among the four estimators. Determining which method is best for public health will require additional analysis that focuses on the true alarm and false alarm rates using known vaccine–event associations. Evaluating the properties of these data mining methods will help determine the value of such methods in vaccine safety surveillance. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS — adverse event; empirical Bayes; proportional reporting ratio; vaccine

INTRODUCTION

The Vaccine Adverse Event Reporting System (VAERS) is a passive surveillance system to monitor vaccine safety and is co-managed by the Food and Drug Administration (FDA) and the Centers for Dis-

ease Control and Prevention (CDC).¹ VAERS receives more than 14 000 reports each year from vaccine manufacturers, healthcare professionals, and the general public. Each report describes one or more adverse events that, at least on temporal grounds, appear to be associated with the administration of a vaccine. Some of these associations are surely coincidental, in some cases the relationship is unclear, and in some cases (e.g., injection site reactions) the relationship is likely causal.

At the time of this analysis, the VAERS database included information on about 70 vaccines and 989 adverse event coding terms. The coding terms for the

* Correspondence to: Dr Emily Jane Woo, HFM-222, Food and Drug Administration, 1401 Rockville Pike, Rockville, MD 20852.
E-mail: wooj@cber.fda.gov

[†]No conflict of interest was declared.

[‡]Present address: The Institute of Statistics and Decision Sciences, Duke University, Durham, North Carolina, USA.

Received 1 March 2004

Revised 7 March 2005

Accepted 9 March 2005

adverse events are known as Coding Symbols for a Thesaurus of Adverse Reaction Terms (COSTARTs), and describe signs, symptoms, and diagnoses, such as headache, swelling at the vaccination site, arthritis, gastroenteritis, and so forth. A single report may list more than one vaccine and may generate several COSTARTs. Since COSTARTs may overlap (e.g., apathy and depression), the same condition may be coded in different ways, according to the judgment of the person assigning the COSTART codes.

Analysis of VAERS data focuses on describing clinical and demographic characteristics of reports and looking for patterns to detect 'signals' of adverse events plausibly linked to a vaccine. While pharmacoepidemiologists do not universally agree upon what constitutes a signal,² a signal can be generally defined as evidence that suggests an adverse event might be caused by vaccination and warrants further investigation or action. Evidence of a signal in case reports and case series of spontaneous reports may include the number of reports and any unexpected patterns in clinical conditions by such factors as age, gender, time to onset, and dose.

Limitations of spontaneous reporting systems such as VAERS include lack of verification of reported diagnoses, lack of consistent diagnostic criteria for all cases with a given diagnosis, wide range in data quality, underreporting, inadequate denominator data (doses administered or patients vaccinated), and absence of an unvaccinated control group. Signals detected through analysis of VAERS data almost always require confirmation through a controlled study. Data mining methods cannot address biases in reporting and should be used in conjunction with medical judgment. The lack of denominator data limits the use of VAERS to discover unforeseen safety problems that may be associated with particular vaccines. The analyst must rely on vaccine distribution data to estimate how many people receive a given vaccine, and does not know the demographic or clinical characteristics of the recipients. Thus, the ability to apply traditional methods of risk analysis, which depend upon estimation of the baseline incidence rates, is limited. Calculation of reporting rates (number of adverse events reported/number of doses of vaccine distributed) and reporting rate ratios that compare vaccines has been used to generate signals.³ Biases in reporting, inadequate denominator data, and lack of background rates for some conditions often limit the utility of the reporting rate approach. For spontaneous reporting systems such as VAERS, it is natural to worry about the effects of systematic underreporting or overreporting. Reporting rates vary from vaccine to vaccine, from adverse event

to adverse event, and from one segment of the population to another. Many factors may stimulate reporting, especially media reporting of suspected side effects, but also FDA and CDC communications. Moreover, the seriousness of an event is known to influence reporting: only a small minority of rashes after MMR vaccine are reported to VAERS, for example, but the majority of cases of paralytic polio after OPV are reported to VAERS.⁴

To address some of these limitations, various data mining techniques have been developed to help uncover potential signals in the data.⁵⁻⁷ The methods permit rapid analysis of large volumes of data that humans cannot possibly evaluate in detail. Although a spontaneous reporting system lacks a true control (i.e., people are not randomized to receive a placebo), data mining techniques permit analysis of a vaccine of interest, with all other vaccines as a quasi-control group for comparison. Data mining cannot eliminate reporting bias, but it does account for different reporting proportions for each vaccine. Specifically, the methods can identify conditions that comprise a larger proportion of reported events for a given vaccine, compared to other vaccines, but an absolute rate is not calculated. Moreover, data mining might identify rare conditions which may not appear during premarketing trials. We propose to apply a variety of methods to help shed light on the potential strengths and weaknesses of the methods with regard to vaccine adverse event data.

There are no 'gold standards' for the detection of vaccine-COSTART associations. The ability to confirm retrospectively a known connection between vaccination and a particular event (e.g., rotavirus vaccine and intussusception⁸) helps to validate data mining methods. Other known associations come from the Vaccine Injury Table, a list of vaccine-adverse event associations which the Institute of Medicine has determined are causal⁹. Agreement among data mining methods, i.e., two or more methods signal a given vaccine-COSTART association, may also be helpful. Our objective is to extend such empirical work through an examination of the statistical properties of the data mining methods that have been proposed.

METHODS

The VAERS database may be viewed as a contingency table with 70 rows (the vaccines) and 989 columns (the COSTARTs). Each cell in the table contains a value n_{ij} that gives the number of reports for the i th vaccine and the j th COSTART.

The usual multinomial model assumes that separate events are classified independently, and this is probably

approximately correct. A small deviation from this model is that sometimes a single report will generate a handful of different COSTARTs (e.g., both nausea and headache), but the effect of this is likely to be small. A second deviation is that sometimes a news story or popular television show can trigger a burst of reports, and these reports are not independent. But again, the overall magnitude of these effects is probably small, and the authenticity of signals generated in this way could be evaluated through examination of longitudinal trends.

One aspect of the VAERS data that has not yet played a substantial role in signal detection research is the association among COSTART terms. One could potentially 'borrow strength' by pooling signals from similar COSTART terms. For example, reports of dizziness and vertigo might be usefully combined to improve the power of the signal detection algorithm. However, we do not address this extension.

Instead, this paper focuses upon a statistical comparison of four signal detection methods that have been discussed in the literature. We call these methods proportional reporting ratio (PRR), screened PRR (SPRR), empirical Bayes geometric mean (EBGM), and lower-bound of the EBGM's 90% confidence interval (EB05). We do not address the relative risk,¹⁰ nor do we consider a conditional probability measure developed by Friedman *et al.*¹¹ and critiqued by DuMouchel *et al.*¹²

There are other methods that can be used for signal detection in large contingency tables without true measures of exposure. For example, the U.S. Census Bureau and the Consumer Product Safety Commission have explored the use of 'raking' to detect interactions in large tables (cf. Little and Wu¹³). Bate *et al.*¹⁴ propose using a Bayesian Confidence Propagation Neural Network for adverse event detection in the WHO database (but DuMouchel¹⁵ argues that this method is an approximation to EBGM based on beta-binomial Bayesian estimates). Hauben and Zhou⁷ review much of this literature.

Although these and other methods could be considered, this research has focused upon the four main techniques that have been piloted within the FDA to date; this paper is not intended to be a comprehensive overview of all currently available methods. A key concern is that methods used for official purposes ideally should be transparent and sufficiently interpretable that expert knowledge can guide the evaluation of new signals. Also, it is highly desirable that the signal detection system used in VAERS not be radically different from systems already in place.

Analysis

One objective of this comparison is to determine whether all four methods agree with each other, as shown by scatterplots and as measured by rank correlation. Comparison of methods' sensitivity and specificity is desirable, but the paucity of gold standards for vaccine–event causality limits the ability to estimate these properties. The theoretical properties of the procedures are also an important consideration. This paper addresses all three bases of comparison; we measure the agreement between methods, we discuss performance with respect to a handful of known adverse effects, and evaluate both kinds of information on the basis of the performance differences expected from theory.

The Vaccine Injury Table is a list of vaccine–adverse event associations that the Institute of Medicine has determined are causal.⁹ By operationalizing these associations as 32 vaccine–COSTART pairs (Table 1), we compare the ability of the methods to signal those pairs. Such operationalization is imperfect, since COSTARTs are applied without standardized definitions or diagnostic confirmation. For example, ARTHRITIS may refer to acute or chronic inflammation of joints. We then evaluate the efficiency of the methods by comparing the number of vaccine–COSTART pairs signaled by each method.

Injection site reactions are accepted as being caused by injectable vaccines. We also look at the methods' ability to signal injection site reactions, represented by COSTART codes ABCESS INJECT SITE, ATROPHY INJECT SITE, CYST INJECT SITE, EDEMA INJECT SITE, GRANULOMA INJECT SITE, HEM INJECT SITE, HYSN INJECT SITE, INFLAM INJECT SITE, INJECT SITE REACT, MASS INJECT SITE, NECRO INJECT SITE, and PAIN INJECT SITE. This comparison allows us to evaluate the methods' abilities to detect an adverse effect which is known to be caused by many vaccines.

A given method may be superior in some situations but inferior in others. There are six possible pairwise comparisons among the four data mining methods. Since our primary interest is to determine whether any method is the most effective for discovering adverse event risks, we focus on four comparisons that seem most informative in terms of identifying plausible vaccine–event pairs.

Data mining methods assessed

Proportional reporting ratio (PRR). The PRR approach was first described by Finney¹⁶ and further developed recently by Evans, Waller, and

Table 1. Vaccine–event associations from vaccine injury table, operationalized as vaccine–COSTART pairs

Association from vaccine injury table	COSTART	Vaccine code(s)
Anaphylaxis or anaphylactic shock after any of the following: tetanus toxoid-containing vaccines; pertussis antigen-containing vaccines; measles, mumps and rubella virus-containing vaccines in any combination; polio inactivated-virus containing vaccines; hepatitis B antigen-containing vaccines	Anaphylaxis	DT, DTAP, DTAPH*, DTP, HEP, IPV, MMR, TD
Chronic arthritis after rubella virus-containing vaccines	Arthritis	MMR, MR, MUR, R
Brachial neuritis after tetanus toxoid-containing vaccines	Brachial neuritis	DTAP*, DTAPH*, DT*, DTP*, TD*, TTOX
Encephalitis after any of the following: pertussis antigen-containing vaccines; measles, mumps and rubella virus-containing vaccines in any combination	Encephalitis	DTAP, DTAPH*, DTP, MMR
Encephalopathy after any of the following: pertussis antigen-containing vaccines; measles, mumps and rubella virus-containing vaccines in any combination	Encephalopathy	DTAP, DTAPH*, DTP, MMR
Intussusception after rotavirus vaccine	Intussusception	RV
Paralytic polio after polio live virus-containing vaccines	Poliomyelitis	OPV
Thrombocytopenia purpura after measles virus-containing vaccines	Thrombocytopenic purpura	M, MM*, MMR, MR

*VAERS did not contain any reports of these vaccine–COSTART pairs.

Davis.¹⁷ To describe the method, suppose we are interested in developing a measure for the strength of the association between vaccine i and COSTART j . Let

- denote n_{ij} , the number of reports for a given vaccine–COSTART combination;
- denote the number of times that any other COSTART is reported for vaccine i ;
- denote the number of times that COSTART j is reported for all other vaccines;
- denote the number of reports for any other vaccine–COSTART combination.

These values may be depicted in a contingency table:

	COSTART j	All Other COSTARTs
Vaccine i	a	b
All other vaccines	c	d

In this contingency table notation, the PRR signal for vaccine i and COSTART j is

$$\text{PRR}_{ij} = \frac{a/(a+b)}{c/(c+d)}$$

This fraction is the proportion of COSTART j reports for vaccine i divided by the proportion of COSTART j reports for all other vaccines. A large PRR for a specific vaccine–COSTART pair indicates that the COSTART has been disproportionately reported for that vaccine, compared with all the other vaccines in VAERS database.

There are several problems with PRR as a metric. First, it does not account for the number of cases n_{ij} . If this value is small, then the generated signal can have large variance. Second, an association might not be statistically significant, but the raw PRR does not reveal this fact since it lacks a well-defined null distribution. Third, the PRR is subject to major distortion due to artifacts in the reporting process. Nonetheless, PRR is an intuitive measure in the absence of exposure data.

Screened proportional reporting ratio (SPRR). Evans, Waller, and Davis¹⁷ proposed screening criteria to define SPRR: $n_{ij} \geq 3$, $\text{PRR} \geq 2$, and Yates-corrected chi-square ≥ 4 . (The Yates correction is a

continuity adjustment to improve the accuracy of the chi-squared approximation to the distribution of the Pearson's test for independence in a contingency table¹⁸). These requirements help to address the first two of the three concerns about use of the raw PRR score. The formula is

$$\text{Yates-corrected } X^2 = \sum (|O_{rs} - E_{rs}| - 0.5)^2 / E_{rs}$$

Here O_{rs} is the observed number in cell (r,s) for $r = 1,2$ and $s = 1,2$ and thus takes the values $a, b, c,$ and $d,$ as in the contingency table in 2.1. The E_{rs} are the numbers expected in those cells under the assumption that the adverse events are independent of the vaccine, and this is given by the row sum times the column sum divided by the total, so

$$E_{11} = (a + b)(a + c) / (a + b + c + d)$$

$$E_{12} = (a + b)(b + d) / (a + b + c + d)$$

$$E_{21} = (a + c)(c + d) / (a + b + c + d)$$

$$E_{22} = (b + d)(c + d) / (a + b + c + d)$$

Under the null hypothesis of no relationship between vaccine and COSTART, the Yates-corrected X^2 -statistic follows a chi-squared distribution with one degree of freedom. A value of 3.84 would be significant at the 0.05 level, which agrees closely with the screening criterion of Yates-corrected $X^2 \geq 4$.

Empirical Bayes geometric mean (EBGM). DuMouchel¹⁵ developed the empirical Bayes approach to analysis of spontaneous reporting systems such as VAERS. The empirical Bayes model assumes that the counts n_{ij} in each cell are random variables from Poisson distributions with unknown means μ_{ij} where the μ_{ij} are themselves random variables with a common distribution. Usually this common distribution is taken to be a mixture of two gamma distributions, one of which is centered at the null value corresponding to a coincidental adverse event, and the other of which is more dispersed and centered at a value corresponding to a true causal relationship between the vaccine and the adverse event. There are many alternative models that lead to similar results; this is a simple mixture model with two gamma components, one of which is highly dispersed and the other of which is concentrated near 1. Simple alternative models assume a mixture of different distributions and use the observed counts n_{ij} to estimate the parameters, but one could also consider nonparametric techni-

ques that allow the data to determine the shapes of the mixture components.

This kind of framework, called a hierarchical model, is widely used in Bayesian practice (see Carlin and Louis¹⁹ for details). It allows one to exploit a simple Bayesian computational structure for inference while avoiding the need to choose a subjective prior for the unknown distribution of μ_{ij} . Formally, the measure corresponding to vaccine i and COSTART j is given by

$$\log_2 \text{EBGM}_{ij} = E[\log_2(\mu_{ij}/E_{ij})|n_{ij}]$$

where the right-hand side of the equation denotes the expectation operator and E_{ij} is the value $(a + b)(a + c) / (a + b + c + d)$ in the notation in the SPRR section, and for the vaccine-COSTART pair of interest, i and j correspond to cell E_{11} . This expression calculates the expected value of the base 2 logarithm of the ratio between the estimated reporting ratio and that under the assumption of no causal relationship, given the observed count of the spontaneous reports for that vaccine and that COSTART. Large values suggest that vaccine i might provoke the adverse event described by COSTART j .

The practical effect of this hierarchical model framework is that it 'shrinks' the estimates of the reporting ratio parameters in the Poisson distributions towards each other, thereby reducing the effect of sampling variation in the data. The shrinkage is greatest when E_{ij} is small and/or n_{ij}/E_{ij} is small, which typically occurs when a or b is small. Another advantage is that the model preserves the interpretability of the parameters and their estimates. The main drawback of this approach is that it is computationally intensive, taking several minutes to run and requiring investment in well-tested, special-purpose code. The computational burden depends upon the number of rows and columns in the matrix, not the number of reports—so from the standpoint of scaling concerns, this performance is adequate for all foreseeable VAERS applications.

Lower-bound of EBGM's 90% confidence interval (EB05). The EB05 is the lower-bound of the 90% confidence interval of EBGM. DuMouchel and Pregibon²⁰ recommend that one use the 5th percentile point of the posterior distribution of the ratio as the metric. If the 5th percentile is large, then the association is unlikely to be due to chance alone and warrants further exploration. The rationale for selecting the 5th percentile point is based upon a loose analogy with frequentist inference, in which one wants to indicate associations that are

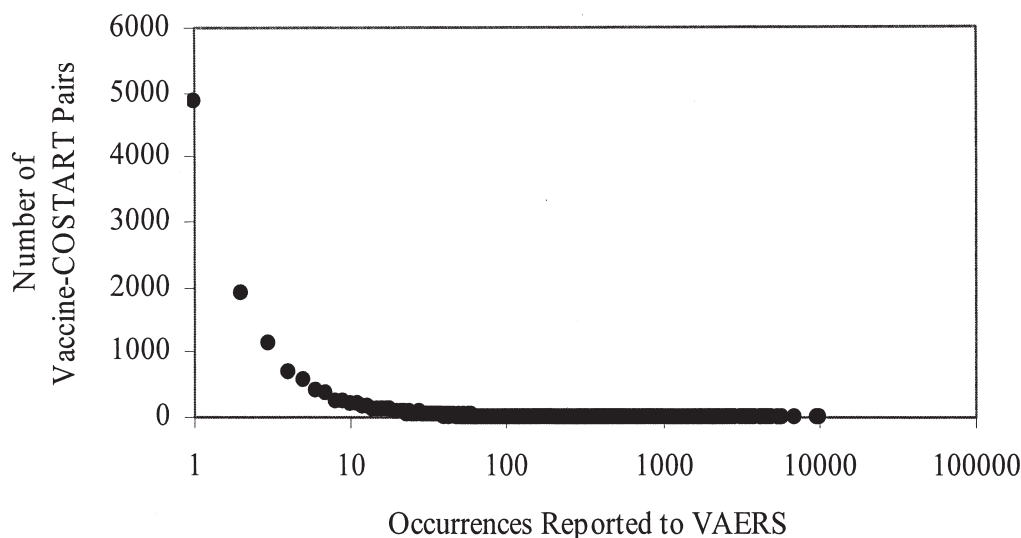


Figure 1. Frequency of occurrence of vaccine-COSTART pairs. This scatterplot illustrates the number of vaccine-COSTART pairs versus the total number of occurrences of a particular vaccine-COSTART pair, for the 14 800 pairs that occurred at least once in VAERS. For each of 4857 pairs, only one occurrence has been reported to VAERS (far left of graph). The pairs that occurred most frequently (far right of graph) correspond to pairs in which the COSTART is a common and expected event (such as fever) that occurs after many vaccines

significant at the 0.05 level. The EB05 signal is conservative and this quality should minimize false positives, but because it represents the lower bound of the confidence interval, it is theoretically less sensitive than EBGM.

A small modification of the EBGM method takes better account of the uncertainty in the posterior distribution of the ratio μ_{ij}/E_{ij} . As part of the EBGM computation, one finds the distribution of this ratio. This distribution can be asymmetric and highly dispersed, in which case use of the expected value could overemphasize the apparent relationship between the vaccine and the COSTART.

RESULTS

Of 69 230 theoretical vaccine-COSTART pairs, 14 800 actually occurred in VAERS at the time of this analysis. Figure 1 illustrates the number of vaccine-COSTART pairs versus the total number of occurrences in VAERS, for these 14 800 pairs. The point at (1, 4857) indicates that 4857 vaccine-COSTART pairs each occurred only once in VAERS. The pairs that occurred most frequently, at the far right of Figure 1, correspond to pairs in which the COSTART is a common and expected event (such as fever) that occurs after many vaccines. Many vaccine-COSTART pairs occurred rarely and some pairs occurred at high frequency, but overall the curve is very smooth.

Comparison of EBGM and PRR

Figure 2 displays the natural logarithm of the EBGM signal versus the natural logarithm of the PRR signal (175 points for which PRR is infinite are omitted from

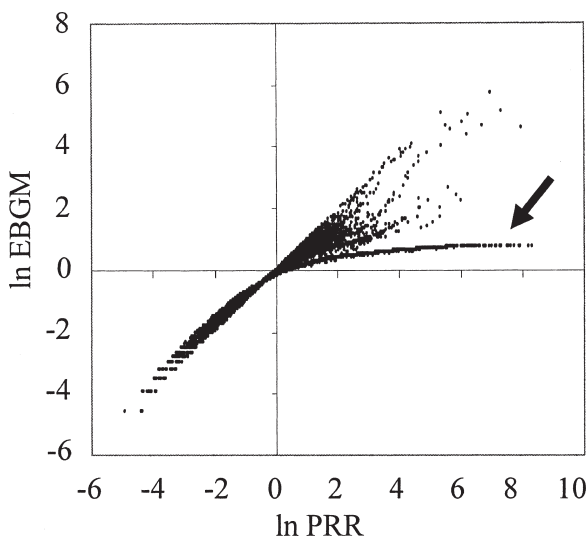


Figure 2. Scatterplot of ln EBGM vs ln PRR. This plot demonstrates a filament (arrow) that consists of vaccine-COSTART pairs for which only one report was received. For these singleton reports, the range of the PRR scores is large compared to that of the EBGM scores, suggesting that PRR gives undue weight to singleton reports relative to EBGM. EBGM: Empirical Bayesian Geometric Mean. PRR: Proportional Reporting Ratio

the graph). The logarithmic plot demonstrates a strong filamentary structure. The lowest filament consists entirely of cases for which $n_{ij}=1$, which accounts for most of the largest (rightmost) values of the PRR scores. In fact, of the 175 vaccine–COSTART pairs with infinite values of PRR, 146 have $n_{ij}=1$, confirming that PRR gives undue weight to singleton reports, and thus is highly susceptible to sampling variation. Only two of these 175 vaccine–COSTART pairs are also in the top 100 EBGM. The known association of rotavirus vaccine and intussusception is not in the top 100 PRR scores, because the value—while very large—is finite.

Comparison of EBGM and SPRR

The screened version of PRR is intended to repair the deficiencies noted in the previous comparison. The SPRR drops cells for which $n_{ij} < 3$ and additionally requires both statistical significance and a raw $PRR \geq 2$; there are 1596 vaccine–COSTART pairs for which SPRR is defined. Figure 3 shows a plot of the natural logarithm of the EBGM score against the natural logarithm of the SPRR score, for the cells for which both SPRR and EBGM are defined (nine points for which SPRR is infinite are omitted from the graph). In comparison with Figure 2, the figure is left-truncated at 0.693, the natural logarithm of 2, because SPRR does not generate a score for cells with PRR less

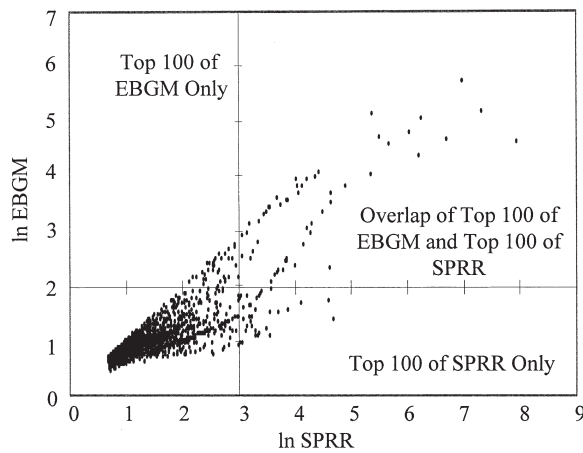


Figure 3. Scatterplot of \ln EBGM vs \ln SPRR. SPRR eliminates the overweighting of singleton reports, as evidenced by the absence of the lower filament seen in Figure 2. Lines indicate the cutoff for the top 100 pairs in each method ($EBGM \geq 7.16$, $\ln EBGM \geq 1.97$; $SPRR \geq 20.08$, $\ln SPRR \geq 3.0$). Fifty-four vaccine–COSTART pairs appear in the top 100 of both methods (upper right quadrant). EBGM: Empirical Bayesian Geometric Mean. SPRR: Screened Proportional Reporting Ratio

than 2. The lower filament, which consisted of cells for which $n_{ij}=1$, has disappeared. Several of the upper filaments are also gone, since they corresponded to cells in which Yates-corrected X^2 was not statistically significant. Note that because of the large number of points, there is considerable overplotting.

Among the top 100 vaccine–COSTART pairs from EBGM and SPRR, 54 appear in both, including nine pairs for which SPRR is infinite. Of those cells flagged by both methods, the Pearson correlation coefficient for the signal ranks is 0.543 ($p < 0.0001$). Among the top 100 EBGM scores ($EBGM \geq 7.16$, $\ln EBGM \geq 1.97$), there are nine cases in which the SPRR method does not signal because $n_{ij} < 3$. There are 37 cases in which all criteria are met, but the rank is simply greater than 100. Among the top 100 SPRR scores ($SPRR \geq 20.08$, $\ln SPRR \geq 3.0$), there are 46 cases in which the EBGM score is not in the top 100 of scores by that method. The top 100 EBGM scores include the rotavirus–intussusception and rubella–arthritis associations, whereas the top 100 SPRR scores include the rotavirus–intussusception and oral polio vaccine–poliomyelitis associations. The top 100 SPRR scores include three injection site reaction COSTARTs, two of which are also among the top 100 EBGM scores.

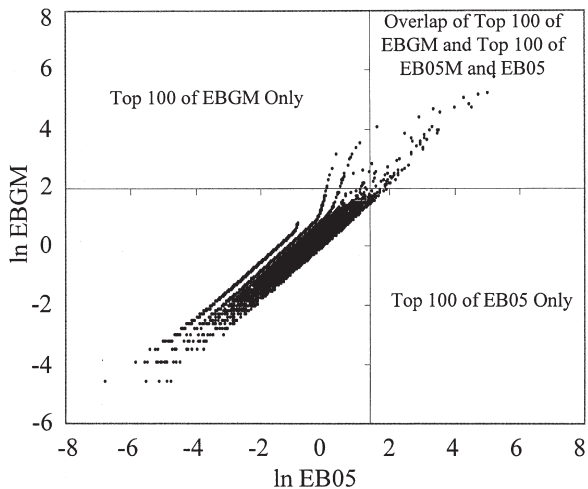


Figure 4. Scatterplot of \ln EBGM vs \ln EB05. Lines indicate the cutoff for the top 100 pairs in each method ($EBGM \geq 7.16$, $\ln EBGM \geq 1.97$; $EB05 \geq 3.98$, $\ln EB05 \geq 1.38$). There is greater agreement between EBGM and EB05 (67 vaccine costart pairs appear in top 100 of both) than between any of the other methods compared. The natural logs of the two scores generally have a linear relationship, but there is divergence from this linear relationship for some vaccine–costart pairs (upper left quadrant) indicating that these methods are not interchangeable EBGM: Empirical Bayesian Geometric Mean. EB05: Lower-Bound of the 90% Confidence Interval of the Empirical Bayesian Geometric Mean.

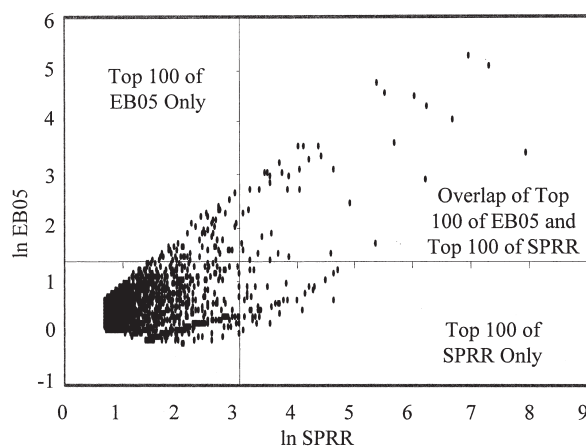


Figure 5. Scatterplot of \ln EB05 vs \ln SPRR. Lines indicate the cutoff for the top 100 pairs in each method ($\text{EB05} \geq 3.98$, $\ln \text{EB05} \geq 1.38$; $\text{SPRR} \geq 20.08$, $\ln \text{SPRR} \geq 3.0$). Forty-two vaccine–COSTART pairs appear in the top 100 of both methods (upper right quadrant). EB05: lower-bound of the 90% confidence interval of the empirical Bayesian geometric mean. SPRR: screened proportional reporting ratio

Comparison of EBGM and EB05

As shown in Figure 4, the natural logs of EBGM and EB05 generally have a linear relationship, as is expected since the posterior distributions are reasonably symmetrical. Sixty-seven vaccine–COSTART pairs appear in the top 100 scores of both EBGM and EB05. The top 100 EBGM scores ($\text{EBGM} \geq 7.16$, 7.16 , $\ln \text{EBGM} \geq 1.97$) include the rotavirus–intussusception and rubella–arthritis associations; the top 100 EB05 scores ($\text{EB05} \geq 3.98$, $\ln \text{EB05} \geq 1.38$) include these two, as well as the oral live polio vaccine–poliomyelitis association. The top 100 EBGM scores include two injection site reaction COSTARTs, one of which appears among the top 100 EB05 scores.

Comparison of EB05 and SPRR

Figure 5 plots the natural log of EB05 against the natural log of SPRR for the cells for which both SPRR and EB05 are defined (nine points for which SPRR is infinite are omitted from the graph). From the plot of the natural logs of the scores, it is clear that many of the top-ranked signals from one method are not the same as the top-ranked signals from the other method.

We have examined the top 100 vaccine–COSTART pairs flagged by the SPRR method and the EB05 method. Among these, 42 are in common, including one infinite value of SPRR. The discrepancies are due

purely to differences in score ranks (top 100 or not) and not due to restrictions on count, Yates-corrected chi-squared, or SPRR value. Of those cells flagged by both methods, the Pearson correlation coefficient for the signal ranks is 0.416 ($p < 0.0001$). The top 100 SPRR scores ($\text{SPRR} \geq 20.08$, $\ln \text{SPRR} \geq 3.0$) include the rotavirus–intussusception and oral live polio vaccine–poliomyelitis associations; the top 100 EB05 scores ($\text{EB05} \geq 3.98$, $\ln \text{EB05} \geq 1.38$) include these two, as well as the rubella–arthritis association. The top 100 SPRR scores include three injection site reaction COSTARTs, one of which appears among the top 100 EB05 scores.

DISCUSSION

Data mining methods have been proposed as screening tools for improving the efficiency of adverse event reports. This is the first analysis comparing several proposed methods using the VAERS database. Several data mining methods exist, and our purpose is to compare four approaches that have been piloted within the FDA. The qualitative features of the comparisons are as follows. The PRR signal appears less useful for postmarketing safety surveillance than SPRR, EBGM, and EB05. The large number of PRR signals for singleton reports could result in many false alarms and divert resources from more consequential relationships. Because of these limitations, PRR was removed from further consideration in the analysis.

Even the best method for detecting clinically important signals among spontaneous report data is subject to limitations. First, if nearly all vaccines are associated with the same adverse event, such as injection site reactions, then automatic signal detection systems are unlikely to discover this association from VAERS data. No single vaccine would likely emerge as markedly different from others, with regard to this event, even if the event were extremely common. Some vaccines are commonly administered simultaneously, e.g., *Hemophilus influenzae* type B vaccine, inactivated polio vaccine, pneumococcal conjugate vaccine, and diphtheria and tetanus toxoids with acellular pertussis vaccine in children. Determining whether a given adverse event results from one of several simultaneously administered vaccines (thereby exonerating the ‘innocent bystanders’), from the simple additive effects of multiple vaccines, or from the synergistic effect of multiple vaccines, is a topic for further research.

We found that the SPRR method was generally competitive with the EBGM method. In comparing EBGM versus SPRR, one should consider the

bias-variance tradeoff.²¹ SPRR estimates have large variance; EBGM estimates are shrunk towards a common mean, which reduces variance at the expense of a small bias. From a public health standpoint, good methods will agree on the strongest signals; close correlation among the other signals is not as helpful. EB05 is designed with statistical principles in mind and takes explicit account of the asymmetry in the distribution of signals. However, these properties may not ensure superior performance. We have evaluated the ability of the different methods to detect some well-known adverse effects. The causal relationship of the vast majority of vaccine–event pairs is unknown, making estimates of sensitivity and specificity unreliable. This paper brings together the comparative information that is currently available, relying on both theory and some empirical work. The number of vaccine–COSTART pairs that ranked in the top 100 by each of two methods (EBGM, EB05, or SPRR) ranged from 42 to 67. Few known associations were in the top 100 scores of any of the methods that we studied, but the known associations that were signaled overlapped and were more similar than different. Under the limitations described above, our research finds that each method has strengths and limitations, and knowledge of these differences has practical value.

ACKNOWLEDGEMENTS

The study was part of routine activities by the Office of Biostatistics and Epidemiology/Food and Drug Administration and therefore did not require supplemental funding. The authors thank Dr. Susan Ellenberg for helpful critique and Dr. Vitali Pool for assistance with data mining method definitions. We also greatly appreciate the efforts of the VAERS Working Group for their dedication to the maintenance of VAERS. The members of the VAERS Working Group include: Marthe Bryant-Genevier, Soju Chang, Hector Izurieta, Ann W. McMahon, Lise Stevens, Frederick Varricchio, and Robert Wise (Food and Drug Administration); Scott Campbell, Robert Chen, Penina Haber, John Iskander, Alena Khromova, Elaine Miller, Gina T. Mootrey, Vitali Pool, and Sean Shadomy (Centers for Disease and Prevention).

REFERENCES

- Chen RT, Rastogi SC, Mullen JR, *et al.* The vaccine adverse event reporting system (VAERS). *Vaccine* 1994; **12**: 542–550.
- Meyboom RH, Egberts AC, Edwards IR, Hekster YA, de Koning FH, Gribnau FW. Principles of signal detection in pharmacovigilance. *Drug Saf* 1997; **16**: 355–365.
- Martin M, Weld LH, Tsai TF, *et al.* Advanced age a risk factor for illness temporally associated with yellow fever vaccination. *Emerg Infect Dis* 2001; **7**: 945–951.
- Rosenthal S, Chen R. The reporting sensitivities of two passive surveillance systems for vaccine adverse events. *Am J Public Health* 1995; **85**: 1706–1709.
- Szarfman A, Machado SG, O'Neill RT. Use of screening algorithms and computer systems to efficiently signal higher-than-expected combinations of drugs and events in the US FDA's spontaneous reports database. *Drug Saf* 2002; **25**: 381–392.
- Hauben M. A brief primer on automated signal detection. *Ann Pharmacother* 2003; **37**: 1117–1123.
- Hauben M, Zhou X. Quantitative methods in pharmacovigilance: focus on signal detection. *Drug Saf* 2003; **26**: 159–186.
- Niu MT, Erwin DE, Braun MM. Data mining in the US Vaccine Adverse Event Reporting System (VAERS): early detection of intussusception and other events after rotavirus vaccination. *Vaccine* 2001; **19**: 4627–4634.
- Vaccine Safety Committee, Institute of Medicine. *Adverse Events Associated with Childhood Vaccines: Evidence Bearing on Casuality*, Stratton KR, Howe CJ, Johnston RB (eds). National Academy Press: Washington, DC, 1994. see also <http://www.hrsa.gov/osp/vicp/table.htm>.
- Church KW, Hanks P. Word association norms, mutual information, and lexicography. *Computational Linguistics* 1990; **16**: 22–29.
- Friedman C, Hripcsak G, DuMouchel W, Johnson SB, Clayton PD. Natural language processing in an operational clinical information system. *Natural Lang Eng* 1995; **2**(1): 83–108.
- Dumouchel W, Friedman C, Hripcsak G, Johnson SB, Clayton PD. Two applications of statistical modeling to natural language processing. In *AI and Statistics V*, Fisher D, Lenz H (eds). Springer-Verlag: New York, 1996: 413–422.
- Little RJA, Wu M-M. Models for contingency tables with known margins when target and sampled populations differ. *J Am Stat Assoc* 1991; **86**: 87–95.
- Bate A, Lindquist M, Orre R, Edwards IR, Meyboom RHB. Data-mining analyses of pharmacovigilance signals in relation to relevant comparison drugs. *Eur J Clin Pharmacol* 2002; **58**: 483–490.
- DuMouchel W. Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. *Am Statistician* 1999; **53**: 177–190.
- Finney DJ. Systemic signalling of adverse reactions to drugs. *Methods Inf Med* 1974; **13**: 1–10.
- Evans SJ, Waller PC, Davis S. Use of proportional reporting ratios (PRRs) for signal from spontaneous adverse drug reaction reports. *Pharmacoepidemiol Drug Saf* 2001; **10**: 483–486.
- Yates F. Contingency tables involving small numbers and the chi-square test. *Suppl J Roy Stat Soc* 1934; **1**: 217–235.
- Carlin BP, Louis TA. *Bayes and Empirical Bayes Methods for Data Analysis*. Boca Raton: Chapman & Hall/CRC: 2000.
- DuMouchel W, Pregibon D. Empirical Bayes screening for multi-item associations. *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2001, pp. 67–76.
- Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning*. Springer-Verlag: New York, 2001 (see chapter 7).