

Vaccine Adverse Event Enrichment Tests

Shuoran Li MS¹ | Lili Zhao PhD^{2*}

¹Department of Statistics, University of Michigan, Ann Arbor, Michigan, 48109, USA.

²Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, 48109, USA

Correspondence

Lili Zhao PhD, Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, 48109, USA
Email: zhaolili@umich.edu

Funding information

Research reported in this publication was supported by the National Institute Of Allergy And Infectious Diseases of the National Institutes of Health under Award Number R01AI158543. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Vaccination safety is critical for individual and public health. Many existing methods have been used to conduct safety studies with the VAERS (Vaccine Adverse Event Reporting System) database. However, these methods frequently identify many adverse event (AE) signals and they are often hard to interpret in a biological context. The AE ontology introduces biologically meaningful structures to the VAERS database by connecting similar AEs, which provides meaningful interpretation for the underlying safety issues. In this paper, we develop rigorous statistical methods to identify "interesting" AE groups by performing AE enrichment analysis. We extend existing gene enrichment tests to perform AE enrichment analysis, while incorporating the special features of the AE data. The proposed methods were evaluated using simulation studies and were further illustrated on two studies using VAERS data. The proposed methods were implemented in R package *AEenrich* and can be installed from the Comprehensive R Archive Network, CRAN, and source code are available at <https://github.com/umich-biostatistics/AEenrich>.

KEYWORDS

Enrichment analysis, MedDRA, Vaccine adverse event, Vaccine safety, VAERS

Abbreviations: ABC, a black cat; DEF, doesn't ever fret; GHI, goes home immediately.

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: [10.1002/sim.9027](https://doi.org/10.1002/sim.9027)

This article is protected by copyright. All rights reserved.

1 | INTRODUCTION

The Centers for Disease Control and Prevention (CDC) and the U.S. Food and Drug Administration (FDA) conduct post-licensure vaccine safety monitoring using the Vaccine Adverse Event Reporting System (VAERS) [1, 2]. VAERS accepts spontaneous reports of suspected vaccine adverse events after administration of any vaccine licensed in the United States from 1990 to present. As a national public health surveillance resource, VAERS is a key component in ensuring the safety of vaccines. [The VAERS report form includes data fields for date of vaccination, vaccine\(s\) administered, date of onset and description of the adverse event, recovery status, and other relevant information.](#) [VAERS data from the primary reports, with identifying patient information removed, are publicly available on the VAERS website \(\[www.vaers.hhs.gov/data/index\]\(http://www.vaers.hhs.gov/data/index\)\); for a detailed description of the VAERS system and its limitations, see \[3\].](#)

Numerous methods have been used to conduct safety studies with the VAERS database [4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16]. In these methods, a contingency table is generally created to display counts for all vaccine and adverse event pairs during a specified time period. In this table, each row represents a vaccine and each column represents an adverse event (AE). Each cell in the table contains the number of VAERS reports that mention both that vaccine and that event for a defined period. A statistical measure is then calculated to quantify the association between an adverse event and a vaccine. A large value of the measure shows a strong association, which might indicate a vaccine safety problem (called "signal"). A signal is considered evidence that an adverse event might be caused by vaccination and warrants further investigation or action. However, these methods frequently identify many AE signals and they are often hard to interpret in a biological context.

Adverse events are naturally related; for example, events of retching, dysphagia and reflux all belong to an AE at a higher-level(i.e., the abnormal digestive system). [AE group analysis \(i.e., enrichment analysis\) is able to provide meaningful interpretation for the underlying safety issues.](#) [For example, strong signals of muscle weakness, loss of sensation and seizures might indicate a potential problem in the nervous system.](#) [Moreover, enrichment analysis is able to detect AE groups containing weak/moderate but concordant AE signals.](#) [For example, the AE group of respiratory system includes five AEs: pneumonia, sinusitis, asthma, bronchitis, and rhinorrhea, which all carry weak to moderate signals.](#) [Each individual AE might not be detected due to lack of statistical power, but the respiratory system as a whole might be detected by the enrichment analysis.](#) [In this paper, we use the phrase "enriched AE groups" to describe AE groups containing unexpectedly large number safety signals.](#)

The largest resource for AE ontology is MedDRA (Medical Dictionary for Regulatory Activities) [17]. It has a five level hierarchy. VAERS uses the second lowest term, "Preferred Terms" (PT), which is a distinct descriptor for a symptom, sign and disease. Related PTs are grouped into higher-level AE terms, including "High Level Group Terms" (HLGT) and "System Organ Classes" (SOC). Higher layers of HLGT and SOC represent biologically and clinically meaningful categories for the AEs observed on the lower PT level. The AE ontology has been used to classify AE signals [18, 19, 20, 21]. For example, [18] showed that most AE signals identified on the PT level were found to be in behavior/neurological AEs on the SOC level. However, these findings are based on an *ad-hoc* strategy of comparing proportions of signaled AEs between AE groups. In this paper, we present rigorous statistical methods to identify AE group that are associated with a vaccine of interest and quantify AE group uncertainty in the enrichment analysis.

Over the last few decades, bioinformatics methods have used gene ontology to systematically dissect large gene lists in order to assemble a summary of the most enriched and pertinent biology. The first idea for gene enrichment analysis is to take the user's pre-selected significant genes, and then compare the difference between the proportion of significant genes that fall into the gene set and the proportion of significant genes that do not [22]. [There are several drawbacks of this method as discussed in \[23, 24\].](#) [For example, it depends on an arbitrary threshold to select the significant genes, and it is incapable of detecting low but concordant signals \(below the used threshold\) from](#)

genes within a gene set. Also, it relies on the gene-gene independence assumption that is known to be biologically invalid. A more recent approach is the Gene Set Enrichment Analysis (GSEA) method [25, 26], which assumes that few major gene expression changes have a considerable effect on pathways function, and the sum of several weaker and concurrent changes in pathways' genes impact the general functioning as well [27]. GSEA uses gene ranks based on a difference measure, such as fold change, and compares the distribution of gene ranks from the gene set to the distribution for the rest of the genes based on a Kolmogorov–Smirnov statistic. Compared to the “cut-off” strategy based on gene significance, GSEA is more statistical powerful by using more data information of the ranks.

The enrichment analyses for gene sets can be applied to study AE groups. However, important issues exist when gene enrichment analysis is applied to adverse event enrichment analysis. Unlike continuous gene expression data, adverse event data are counts, and a large amount of AEs have a zero count. For example, in the VAERS dataset, approximately 40% AEs were never mentioned with the “FLU4” vaccine, resulting in 40% AEs with a zero count. The current gene enrichment tests can not handle excessive zeros. Additionally, we encountered 20% ties in a ratio measure (as defined in the Methods section below) with the count data. The current GSEA assigns random ranks to the tied statistics, which can lead to inaccurate results. In this work, we extend the current enrichment tests to appropriately address these two issues to perform AE enrichment analysis. The proposed method was implemented as R package `AEenrich` <https://CRAN.R-project.org/package=AEenrich>.

2 | METHODS

Data Structure

For a particular vaccine (denoted as the target vaccine), we create a $2 \times N$ contingency table (see Table 1), with two rows for the target vaccine (Yes/No) and N columns for the AEs reported in the VAERS database during a study period. In this table, n_{1j} is the number of VAERS reports that mention both the target vaccine and the j^{th} AE in a defined period, $n_{.j}$ is the total number of reports that mention the j^{th} AE, $n_{1.}$ is the total number of reports that mention the target vaccine, and $n_{..}$ is the total number of reports in the study period.

TABLE 1 AE count data in a $2 \times N$ table for a target vaccine

VaccineAE	AE_1	AE_2	\dots	AE_N	Total
Yes	n_{11}	n_{12}	\dots	n_{1N}	$n_{1.}$
No	$n_{.1} - n_{11}$	$n_{.2} - n_{12}$	\dots	$n_{.N} - n_{1N}$	$n_{..} - n_{1.}$
Total	$n_{.1}$	$n_{.2}$	\dots	$n_{.N}$	$n_{..}$

2.1 | AEKS: AE enrichment analysis based on modified K-S statistic

In this section, we extend the current GSEA [25, 26] to handle AE data with ties and excessive zero. Poisson distribution has been commonly used to model the n_{1j} [4, 11, 28],

$$n_{1j} \sim \text{Poisson}(n_{.j} \times \lambda_j), \text{ for } j = 1, \dots, N, \quad (1)$$

where λ_j is the reporting ratio (RR) for the j^{th} AE with the target vaccine, with a large value indicating a strong safety signal. RRs are the statistics of interest and we will use their maximum likelihood estimates $\frac{n_{1j}}{n_{.j}}$'s as observed values.

Our goal is to determine whether members of a AE group tend to have higher RRs.

Calculate enrichment score for each AE group

1. Rank order the N AEs based on the statistic RR. Assume that there are J distinct RRs ($J \leq N$) and we order the N AEs from the highest to the lowest rank as $L = \{\hat{A}E_1, \hat{A}E_2, \dots, \hat{A}E_J\}$, where $\hat{A}E_j = \{AE_{j1}, \dots, AE_{jn_j}\}$ is a set of n_j AEs with same RR.
2. Extend GSEA to handle tied RRs. Given position i in L , evaluate the fraction of AEs in group G ("hits") and the fraction of AEs not in G ("misses"). N_G denotes the number of distinct AE terms in group G .

$$P_{hit}(G, i) = \sum_{j \leq i} \sum_{k \leq n_j} \frac{1}{N_G} 1(AE_{jk} \in G),$$

$$P_{miss}(G, i) = \sum_{j \leq i} \sum_{k \leq n_j} \frac{1}{N - N_G} 1(AE_{jk} \notin G),$$

where $1(\cdot)$ is the indicator function. We then compute a running sum across all N AEs. The K-S statistic for AE group G is defined as $KS(G) = \max_{1 \leq i \leq J} (P_{hit}(G, i) - P_{miss}(G, i))$, which is the maximum value that P_{hit} is above P_{miss} . When many members of G appear at the top of the list, $KS(G)$ is high.

3. Handle zero counts. The maximum likelihood estimate for λ_i is $\frac{n_{ij}}{n_j}$. Thus, a zero count will produce a zero RR. The above $KS(G)$ treats zero RRs as ties and can not efficiently handle excessive zeros. For example, it may identify a group as enriched if the group has majority of AEs with zero counts and only a few AEs with large signals. To include the number of zero counts in the enrichment test, we define group G as enriched only if it contains a smaller proportion of zero counts compared to the rest of groups. To this end, we let p_0^G denote the proportion of zero counts in group G , $p_0^{G^c}$ denote the proportion not in group G , and create an indicator function by comparing these two zero proportions. The function is zero if p_0^G is larger than $p_0^{G^c}$.
4. Combine the statistics in 2 and 3, we propose a composite enrichment score (ES)

$$ES(G) = KS(G) \times 1(p_0^G \leq p_0^{G^c}), \quad (2)$$

where $ES(G) \in [0, 1]$, and $ES(G)$ is zero if the proportion of zero counts in group G is larger than the proportion of zero counts in other groups. $ES(G)$ is large if group G has a smaller proportion of zero counts than the remaining groups and the non-zero counts in group G are concentrated at the top of the list L .

Estimate statistical significance for an AE group

The distribution of $ES(G)$ under the null is not analytically tractable and is obtained using Monte Carlo hypothesis testing[28]. Under the null hypothesis, $H_0: \lambda_1 = \lambda_2 = \dots = \lambda_N = \lambda_0$. Under this hypothesis, $n_{1i} \sim Poisson(n_i \times \lambda_0)$, for $i = 1, \dots, N$. Based on the relationship between Poisson and Multinomial distributions, the joint distribution of (n_{11}, \dots, n_{1N}) , conditioning on $\sum_{i=1}^N n_{1i} = n_1$ and (n_1, \dots, n_N) is

$$(n_{11}, n_{12}, \dots, n_{1N}) | n_1; n_1, \dots, n_N \sim \text{multinomial}(n_1, \{r_1, \dots, r_N\}) \quad (3)$$

where $r_i = \frac{n_i \lambda_0}{n \cdot \lambda_0} = \frac{n_i}{n}$.

Given this multinomial distribution, we generate the AE count data and compute $ES^*(G)$ using formula (2). We

repeat this process M times (M is generally large; say 5000) to create a null distribution of $ES^*(G)$. The p -value is the proportion of $ES^*(G)$ that is greater than or equal to the observed $ES(G)$. Finally, we apply the Benjamini-Hochberg procedure [29] converting p -values into q -values to control the false discovery rate.

Estimate statistical significance for individual AEs within the AE group

Based on the null distribution generated by the Monte Carlo hypothesis testing method, we can also estimate the statistical significance of each individual AE by computing the proportion of RR^* in the null distribution that is greater than or equal to the observed RR . The Benjamini-Hochberg procedure is then used to adjust for the false discovery rate.

2.2 | AEFisher: AE Enrichment test based on modified Fisher's exact test

This approach first assesses the significance of the association between each AE and the vaccine and then uses a "cutoff" strategy to classify the AEs into signaled and unsignaled AEs. To test the significance of the association, we apply the Fisher's exact test to data in a 2 by 2 table (see Table 2) and then use the Benjamini-Hochberg procedure to convert p -values into q -values for controlling the false discovery rate. A signaled AE is defined based on both the strength of the signal and statistical significance, such as q -value < 0.1 and odds ratio (OR) > 1.5 .

TABLE 2 A 2 by 2 contingency table for a vaccine- AE_i pair

Vaccine AE_i	Yes	No
Yes	n_{1j}	$n_{.j} - n_{1j}$
No	$n_{.j} - n_{1j}$	$(n_{..} - n_{1.}) - (n_{.j} - n_{1j})$

To conduct the enrichment analysis for a particular AE group G , a conventional approach is to compare proportions of the signaled AEs in group G and not in group G . If there are significantly more signaled AEs in group G , then group G is enriched. To incorporate the excessive zero RRs in the test, we propose a composite enrichment score

$$ES(G) = OR^G \cdot 1(p_0^G \leq p_0^{G^c}),$$

where OR^G is the odds ratio estimating the association between signaled AEs and group G . A large OR^G ($OR^G > 1$) indicates more signaled AEs in group G than in the remaining groups. As in the AEKS test, $1(p_0^G \leq p_0^{G^c})$ ensures that an enriched group has the proportion of zeros smaller or equal to remaining groups.

Estimate statistical significance for an AE group

We perform a permutation test to assess the significance of the enrichment score for group G by randomly reshuffling the signaled/unsignaled labels. This in spirit is the same as the Fisher's exact test of fixing the row and column margins (here, the group size and the total number of signaled and unsignaled AEs are fixed), while considering the zero proportions.

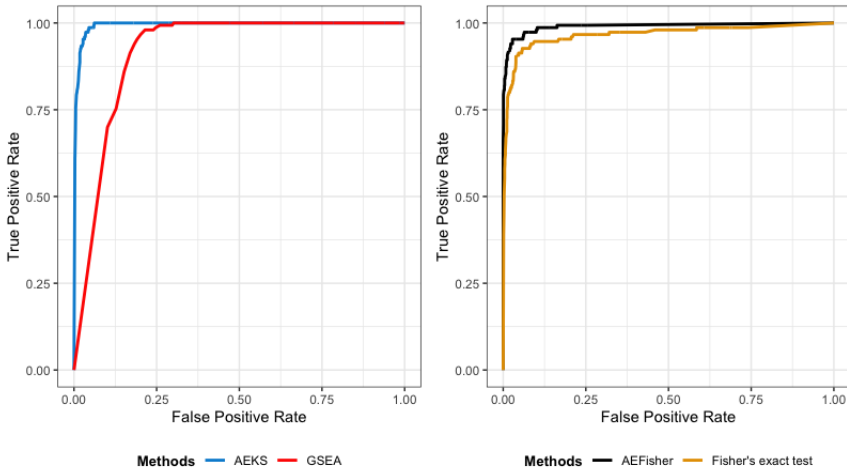


FIGURE 1 ROC curves using K-S statistics (AEKS vs GSEA) (left) and Fisher's exact tests (AEFisher vs Fisher's exact test) (right). The ROC curve is constructed using p-values from the 10 simulated datasets. By comparing the enrichment decision from the test to the true enrichment status in the data generating step, the true and false positive rates were estimated and connected to make the ROC curve.

3 | SIMULATION STUDIES

We ran simulation studies to investigate our proposed methods and compared them to existing enrichment tests. To make simulation studies more realistic, data in simulated datasets were made similar to the real dataset. We first created the AE group structure using the AE groups defined on the HLGT level in MedDRA. In each simulated dataset, we set the number of AE groups to be 150 and determined the group size, N_G , by randomly sampling the group size data in MedDRA under the constraint of $N_G \geq 10$. Similarly, the total count of each AE was determined by randomly sampling the AE total count data in VAERS. Then we randomly selected 10% of the AE groups as enriched and the remaining groups as un-enriched. In VAERS, the proportion of zero AEs per group, p_0 , is between 10-60%, therefore, we used p_0 in this range in simulations. We generated a non-zero AE count from a Poisson distribution in formula (1) with the rate parameter randomly sampled from the estimated λ 's in VAERS. Specifically, in an enriched group, p_0 was sampled uniformly from 0.1 to 0.3 and the rate parameter was constrained to be larger than 0.3. In an un-enriched AE group, we either set the range of p_0 between 0.4 to 0.6 without constraining the rate parameter, or set the rate parameter smaller than 0.4 with p_0 in the range of 0.1 to 0.6.

To compare performance of different methods, we constructed the ROC curve using p-values from the 10 simulated datasets. By using a threshold on the p-values, an AE group is identified as enriched if it has a p-value smaller than the threshold. By comparing the enrichment decision from the test to the true enrichment status in the data generating step, we estimated the true positive rate (i.e., the percentage of AE groups that are correctly identified as enriched) and false positive rate (i.e., the percentage of AE groups that are incorrectly identified as enriched). By varying the threshold, we obtained a series of false and true positive rates and then connected them to make an ROC curve. As shown in Figure 1, AEKS and the AEFisher performed significantly better than the GSEA and the conventional Fisher's exact test, respectively.

In order to study the effect of sample size on the results of the analysis, we repeated the simulation studies with different sample sizes. Specifically, we let the total count, n_j , for the i^{th} AE ($i = 1, \dots, N$) equal to 10%, 20%, 40%,

60% or 80% of the original count, such that the total sample size is reduced proportionally. All the parameters in the data generating model remain the same. For each sample size, we calculated the area under the ROC curves (AUC) (see Figure 2). As shown in this Figure, both AEKS and AEFisher improve its performance in distinguishing enriched from un-enriched AE groups as sample size increases, and AEKS performs better than AEFisher for different sample sizes. This result is expected as AEKS is more statistical powerful by using ranks of individual AEs rather than the binary (significant vs non-significant) data from the single-AE analysis in AEFisher. Moreover, the advantage of AEKS over AEFisher is more obvious when the sample size is small. One reason is that a small sample size produces inaccurate estimations on statistical significance for individual AEs in AEFisher, thereby leading to under-performed enrichment test.

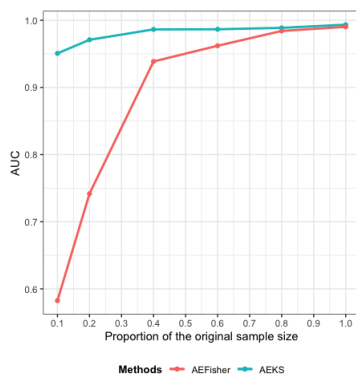


FIGURE 2 Area under the ROC curves for AEKS and AEFisher with the reduced sample sizes. The ROC curve is constructed using p-values from the 10 simulated datasets. By comparing the enrichment decision from the test to the true enrichment status in the data generating step, the true and false positive rates were estimated and connected to make the ROC curve.

4 | APPLICATION TO VAERS DATASETS

We applied AEKS and AEFisher to VAERS dataset to study flu and hepatitis vaccines. In both studies, we used the HLGT level of MedDRA to define AE groups. In AEKS, a signaled AE is defined if the q -value < 0.1 . In AEFisher, a signaled AE is defined if the q -value < 0.1 and OR > 1.5 . In both AEKS and AEFisher, an AE group is enriched if q -value < 0.1 .

4.1 | Study flu vaccines

Influenza vaccine is given in large quantities and it prevents millions of illnesses and flu-related doctor's visits each year. CDC recommends the appropriate vaccine during the flu season. Options include inactivated influenza vaccine (IIV) ("FLU3" or "FLU4" in VAERS) or live attenuated influenza vaccine (LAIV) ("FLUN3" or "FLUN4" in VAERS). By restricting the age of the vaccine recipients between 2 and 49, there were 139,353 and 21820 reports for IIV and LAIV, respectively. To compare LAIV relative to IIV, we created a contingency table with two rows representing vaccine (LAIV vs IIV) and 3,534 columns representing AEs. These AEs were classified into 287 AE groups. As in the gene set enrichment analysis, we removed AE groups containing less than 5 AEs, resulting in 132 AE groups with a total of

TABLE 3 The enriched AE groups (bolded) and significant AEs using AEKS and AEFisher to study LAIV relative to IIV.

Method	AE (group size)	q value (AEKS)	q value (AEFisher)	Odds ratio
	Upper respiratory tract disorders (excl infections) (19)	3.30×10^{-2}	6.60×10^{-2}	
	Epistaxi	0.00	8.79×10^{-47}	12.10
	Nasal congestion	0.00	1.99×10^{-58}	6.36
	Stridor	1.13×10^{-2}	2.44×10^{-3}	4.61
	Tonsillar hypertrophy	2.38×10^{-2}	6.61×10^{-3}	7.30
	Nasal oedema	2.99×10^{-2}	7.30×10^{-3}	12.78
	Sinus disorder	5.07×10^{-2}	2.07×10^{-2}	6.39
	Anosmia	6.79×10^{-2}	2.50×10^{-2}	25.54
	Seasonal allergy	6.79×10^{-2}	2.86×10^{-2}	10.65
AEKS & AEFisher	Respiratory tract infections (17)	6.60×10^{-2}	6.60×10^{-2}	
	Epistaxi	0.00	8.79×10^{-47}	5.82
	Nasal congestion	0.00	1.74×10^{-19}	6.11
	Stridor	4.25×10^{-9}	7.03×10^{-12}	5.69
	Tonsillar hypertrophy	2.93×10^{-8}	3.40×10^{-13}	115.05
	Nasal oedema	3.88×10^{-6}	1.04×10^{-7}	3.88
	Sinus disorder	5.06×10^{-5}	3.36×10^{-6}	2.95
	Anosmia	1.35×10^{-3}	1.63×10^{-4}	3.95
	Seasonal allergy	2.46×10^{-2}	8.13×10^{-3}	3.78

1,828 AEs. Finally, we applied both AEKS and AEFisher to identify enriched AE groups based on the data contained in this $2 \times 1,828$ contingency table.

As shown in Table 3, of the 132 AE groups we studied, both AEKS and AEFisher identified the same two enriched AE groups: respiratory tract infections and upper respiratory tract disorders. Relative to IIV, LAIV is associated with increased risk of respiratory system disorders. Individual AE identified in each group include rhinitis, nasal congestion, sinus disorder, which have been reported before [30, 31, 32], New signals, such as epistaxis, is clinically interesting, and it might be true signals that need to be validated in large healthcare databases.

4.2 | Study hepatitis A and B combination vaccines

In this study, we were interested in identifying safety problems that are likely due to interactions of two vaccines when they are administered to an individual at the same time. Specifically, we compared AE profiles induced by the hepatitis A and B combination vaccine ("Twinrix" in VAERS) to monovalent hepatitis A and B vaccines ("Havrix" for hepatitis A and "Engerix-B" for hepatitis B in VAERS). We selected vaccine reports from 2002 to 2018. There were 53,415, 33,087, 10,356 reports with Havrix, Engerix-B, and Twinrix, respectively. To compare the combination vaccine to the two monovalent vaccines, we created a contingency table with one row for Twinrix, one row for Havrix or Engerix-B, and 1,528 AEs (after removing AE groups containing less than 5 AEs). These AEs are grouped into 127 AE groups. Finally, we applied both AEKS and AEFisher to identify enriched AE groups based on the data contained

TABLE 4 The enriched AE groups (bolded) and significant AEs using AEKS and AEFisher to study hepatitis A and B combination relative to monovalent Hepatitis A and B vaccines

Method	AE (group size)	q value	Odds ratio
AEKS	Peripheral neuropathies (10)	3.18×10^{-2}	
	Guillain-Barre syndrome	7.08×10^{-7}	
	Neuropathy peripheral	1.83×10^{-3}	
	Miller Fisher syndrome	3.48×10^{-3}	
AEFisher	Musculoskeletal and connective tissue disorders NEC (19)	6.35×10^{-2}	
	Pain in extremity	3.37×10^{-4}	1.57
	Back pain	8.25×10^{-4}	2.24
	Neck pain	2.04×10^{-3}	2.03
	Musculoskeletal pain	2.42×10^{-2}	2.17
	Mobility decreased	5.98×10^{-2}	2.14
	Musculoskeletal disorder	8.98×10^{-2}	11.14

in this $2 \times 1,528$ contingency table. In this study, AEKS and AEFisher identified different AE groups. As shown in Table 4, AEKS identified peripheral neuropathies, while AEFisher identified musculoskeletal and connective tissue disorders. Peripheral neuropathies were also mentioned in [33] as an important AE group associated with the combination hepatitis vaccine.

5 | DISCUSSION

AEKS and AEFisher have inherited pros and cons from the GSEA and Fisher's exact test in the gene enrichment analysis, respectively. For example, AEFisher depends on the threshold to select the significant AEs, and it can't be applied if no AEs are found to be significant based on the threshold. Compared to AEFisher, AEKS is more statistical powerful by using ranks of individual AEs rather than a "cut-off" strategy based on an arbitrary threshold. AEKS is capable of detecting AE groups with a few large AE signals or several weaker and concordant AE signals. In simulation studies, we have demonstrated the advantage of AEKS over AEFisher. Moreover, similar to the gene enrichment analysis, both AEKS and AEFisher can not efficiently deal with AE group overlap, where some AEs may belong to several AE groups. There have been several attempts to alleviate the effect of gene set overlap [24]. However, these methods suffer from low sensitivity. Developing methods dealing AE group overlap in the context of adverse events would be an interesting future work.

AEFisher uses a list of binary (signaled vs unsignaled AEs) data from the single-AE analysis to perform the enrichment test. In this paper, signaled AEs are determined using the Fisher's exact test. Alternatively, other tests for single-AE analysis can be used, including methods in [4, 7, 28]. On the other hand, AEKS uses ranks of reporting ratios of individual AEs, but other statistics can be used as alternatives to the reporting ratio, such as odds ratios or p-values. This is the first study on adverse event enrichment analysis, and we aim to set up a general framework for future research in this field.

In both AEKS and AEFisher, the enrichment score has two components, one for AEs with non-zero counts and one for AEs with zero counts. For the zero component, we used an indicator function, which is 0 when the proportion

of zero counts in group G is larger than the proportion in other groups. This is created based on our definition that an enriched AE group is required to have the zero count proportion less than or equal to other groups. However, this definition can be modified; for example, an AE group may be considered as enriched if it has many large AE signals from the non-zero component and slightly larger proportion of zero counts compared to other groups. In this case, a ratio of zero proportion not in the group versus in the group can replace the indicator function. However, an extremely large or small ratio will make the zero-component dominate the enrichment decision (i.e., ES is solely driven by the ratio). To mitigate this issue, a weight may be introduced for the ratio to achieve a good balance in decision between the zero and non-zero components. This topic will be an interesting work for future research.

6 | CONCLUSIONS

In this article, we develop two enrichment tests for vaccine adverse event enrichment analysis by incorporating the special features of the AE count data. AEFisher is a modified Fisher's exact test based on pre-selected significant AEs, while AEKS is based on a modified Kolmogorov–Smirnov statistic. By appropriately addressing the issues of ties and excessive zeros in AE count data, our enrichment tests performed well as demonstrated by simulation studies and analyses of VAERS data. While the proposed methods are developed for vaccine safety, they are broadly applicable to other safety surveillance projects. For example, they can be directly applied to the FDA Adverse Events Reporting System (FAERS) and the Adverse Drug Reactions (ADR) database for national and international drug safety.

Data Availability

VAERS data from the primary reports, with identifying patient information removed, are publicly available on the VAERS website (www.vaers.hhs.gov/data/index).

Acknowledgements

Thanks to Kirsten Herold for proofreading the article and thanks to Michael Kleinsasser for compiling the R package.

Funding

This work was supported by the National Institute Of Allergy And Infectious Diseases of the National Institutes of Health under Award Number R01AI158543. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

references

- [1] Varricchio F, Iskander J, DeStefano F, Ball R, Pless R, Braun MM, et al. Understanding vaccine safety information from the Vaccine Adverse Event Reporting System. *Pediatr Infect Dis J* 2004;23:287–294.
- [2] Shimabukuro TT, Nguyen M, Martin D, DeStefano F. Safety monitoring in the vaccine adverse event reporting system (VAERS). *Vaccine* 2015;33(36):4398–4405.
- [3] Zhao L, Lee S, Li R, Ong E, He Y, Freed G. Improvement in the Analysis of Vaccine Adverse Event Reporting System Database. *Statistics in Biopharmaceutical Research* 2020;12(3):303–310.

- [4] DuMouchel W. Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. *The American Statistician* 1999;53:177–190.
- [5] Evans SJ, Waller PC, Davis S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiology and Drug Safety* 2001;10:483–486.
- [6] van Puijenbroek EP, Bate A, Leufkens HG, Lindquist M, R RO, Egberts AC. A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions. *Pharmacoepidemiology and Drug Safety* 2002;11:3–10.
- [7] Bate A, Lindquist M, Edwards IR, Olsson S, Orre R, Lansner A, et al. A Bayesian neural network method for adverse drug reaction signal generation. *European Journal of Clinical Pharmacology* 1998;54:315–321.
- [8] Lansner ROA, Bate A, M Lindquist IRE, Olsson S, Orre R, Lansner A, et al. Bayesian neural networks with confidence estimations applied to data mining. *Comput Stat Data Anal* 2000;34:473–493.
- [9] Nören GN, Bate A, Orre R, Edwards IR. Extending the methods used to screen the WHO drug safety database towards analysis of complex associations and improved accuracy for rare events. *BCPNN* 2006;25:3740–3757.
- [10] 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, San Francisco, CA 2001*;p. 67–76.
- [11] 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. *PLoS ONE* 2002;25:381–392.
- [12] Kulldorff M, Davis RL, Kolczak M, Lewis E, Lieu T, Platt R. A maximized sequential probability ratio test for drug and vaccine safety surveillance. *Sequential Analysis* 2011;30:58–78.
- [13] Davis RL, Kolczak M, Lewis E, Nordin J, Goodman M, Shay DK, et al. Active surveillance of vaccine safety: a system to detect early signs of adverse events. *Epidemiology* 2005;16:336–341.
- [14] Li L, Kulldorff M. A conditional maximized sequential probability ratio test for pharmacovigilance. *Statistics in Medicine* 2009;29:284–295.
- [15] Li R, Stewart B, Weintraub E, McNeil MM. Continuous sequential boundaries for vaccine safety surveillance. *Statistics in Medicine* 2014;33:3387–3397.
- [16] Kulldorff M, Dashevsky I, Avery TR, Chan KA, Davis RL, Graham D, et al. Drug safety data mining with a tree-based scan statistic. *Pharmacoepidemiology and Drug Safety* 2013;22:517–523.
- [17] Mozzicato P. MedDRA: an overview of the Medical Dictionary for Regulatory Activities. *Pharm Med* 2009;23:65–75.
- [18] Zhang Y, Tao C, He Y, Kanjamala P, H HL. Network-based analysis of vaccine-related associations reveals consistent knowledge with the vaccine ontology. *Journal of Biomedical Semantics* 2013;4:33.
- [19] Marcos E, Zhao B, He Y. The Ontology of Vaccine Adverse Events (OVAE) and its usage in representing and analyzing adverse events associated with US-licensed human vaccines. *Journal of Biomedical Semantics* 2013;4:40.
- [20] Sarntivijai S, Xiang Z, Shedden KA, Markel H, Omenn GS, Athey BD, et al. Ontology-based combinatorial comparative analysis of adverse events associated with killed and live influenza vaccines. *PLoS ONE* 2012;7:e49941.
- [21] Guo A, Racz R, Hur J, Lin Y, Xiang Z, Zhao L, et al. Ontology-based collection, representation and analysis of drug-associated neuropathy adverse events. *J Biomed Semantics* 2016;7:29.
- [22] Zeeberg BR, Feng W, Wang G, Wang MD, Fojo AT, Sunshine M, et al. GoMiner: a resource for biological interpretation of genomic and proteomic data. *Genome biology* 2003;4(4):R28.

- [23] Maleki F, Ovens K, Hogan DJ, Kusalik AJ. Gene set analysis: challenges, opportunities, and future research. *Frontiers in genetics* 2020;11:654.
- [24] Tarca AL, Bhatti G, Romero R. A comparison of gene set analysis methods in terms of sensitivity, prioritization and specificity. *PLoS one* 2013;8(11):e79217.
- [25] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* 2005;102(43):15545–15550.
- [26] Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, et al. PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature genetics* 2003;34(3):267–273.
- [27] iuseppe A, Pastrello C, Jurisica I. Comprehensive pathway enrichment analysis workflows: COVID-19 case study. *Briefings in Bioinformatics*;
- [28] Huang L, Zalkikar J, Tiwari RC. A likelihood ratio test based method for signal detection with application to FDA's drug safety data. *Journal of the American Statistical Association* 2011;106(496):1230–1241.
- [29] Thissen D, Steinberg L, Kuang D. Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. *Journal of educational and behavioral statistics* 2002;27(1):77–83.
- [30] Baxtera R, Eatona A, Hansena J, Aukesa L, Caspardb H, Ambroseba CS. Safety of quadrivalent live attenuated influenza vaccinein subjects aged 2–49 years. *Vaccine* 2017;35:1254–1258.
- [31] Haber P, Moro PL, Cano M, Lewis P, Stewart B, Shimabukuro TT. Post-licensure surveillance of quadrivalent live attenuated influenza vaccine United States, Vaccine Adverse Event Reporting System(VAERS), July 2013–June 2014. *Vaccine* 2015;33:1987–1992.
- [32] Lambkin-Williams R, Gelder C, Broughton R, Mallett CP, Gilbert AS, Mann A, et al. An Intranasal Proteosome-Adjuvanted Trivalent Influenza Vaccine Is Safe, Immunogenic and Efficacious in the Human Viral Influenza Challenge Model. Serum IgG and Mucosal IgA Are Important Correlates of Protection against Illness Associated with Infection. *PLoS One* 2016;11:e0163089.
- [33] Xie J, Zhao L, Zhou S, He Y. Statistical and Ontological Analysis of Adverse Events Associated with Monovalent and Combination Vaccines Against Hepatitis A and B Diseases. *Scientific Report* 2016;6:3418.