METHODOLOGY ARTICLE

Vaccine Adverse Event Enrichment Tests

Shuoran Li[1] and Lili Zhao[2*]

*Correspondence:
zhaolili@umich.edu
[2]Department of Biostatistics, University of Michigan, Ann Arbor, USA
Full list of author information is available at the end of the article

Abstract

Background: : 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.

Results: We extend existing gene enrichment tests to perform AE enrichment analysis. Unlike the continuous gene expression data, AE data are counts. Therefore, AE data has many zeros and ties. We propose two enrichment tests, AEFisher and AEKS. AEFisher is a modified Fisher's exact test based on pre-selected significant AEs, while AEKS is based on a modified Kolmogorov–Smirnov statistic. Both tests incorporate 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.

Conclusion: 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. 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; Vaccine adverse event; VAERS; MedDRA

Background

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.

Numerous methods have been used to conduct safety studies with the VAERS database [3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15]. 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 are all related to an abnormal digestive system. The AE ontology introduce biologically meaningful structures to the VAERS database by connecting similar AEs, which provides meaningful interpretation for the underlying safety issues. The largest resource for describing AE relationships is MedDRA (Medical Dictionary for Regulatory Activities) [16]. 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 [17, 18, 19, 20]. For example, [17] 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 groups of AEs 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 basic idea in the traditional strategy 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 [21]. A more recent approach is the Gene Set Enrichment Analysis (GSEA) method [22, 23], which uses gene ranks based on a difference measure, such as fold change, rather than a "cut-off" strategy based on gene significance. In GSEA, the distribution of gene ranks from the gene set is compared against the distribution for the rest of the genes by using the enrichment score (ES) based on a Kolmogorov–Smirnov statistic.

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.

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_{1i} is the number of VAERS reports that mention both the target vaccine and the i^{th} AE in a defined period, $n_{\cdot i}$ is the total number of reports that mention the i^{th} AE, $n_{1\cdot}$ is the total number of reports that mention the target vaccine, and $n_{\cdot\cdot}$ 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

Vaccine \ AE	AE_1	AE_2	$\cdots$	AE_N	Total
Yes	n_{11}	n_{12}	$\cdots$	n_{1N}	$n_{1\cdot}$
No	$n_{\cdot 1} - n_{11}$	$n_{\cdot 2} - n_{12}$	$\cdots$	$n_{\cdot N} - n_{1N}$	$n_{\cdot\cdot} - n_{1\cdot}$
Total	$n_{\cdot 1}$	$n_{\cdot 2}$	$\cdots$	$n_{\cdot N}$	$n_{\cdot\cdot}$

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

In this section, we extend the current GSEA [22, 23] to handle AE data with ties and excessive zero. Poisson distribution has been commonly used to model the n_{1i} [3, 10, 24],

$$n_{1i} \sim Poisson(n_{\cdot i} \times \lambda_i), \quad \text{for } i = 1, \cdots, N, \tag{1}$$

where λ_i is the reporting ratio (RR) for the i^{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 $\frac{n_{1i}}{n_{\cdot i}}$'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 $L = \{\tilde{AE}_1, \tilde{AE}_2, \ldots, \tilde{AE}_J\}$, where $\tilde{AE}_j = \{AE_{j1}, \ldots AE_{jn_j}\}$ 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.

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

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

where $\mathbb{1}(\cdot)$ 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 $\frac{n_{1i}}{n_{\cdot i}}$. Thus, a zero count will produce a zero RR. Let p_0^G denote the proportion of AEs with RR equal to 0 in group G, and let $p_0^{G^c}$ denote the proportion not in group G. We consider Group G not enriched only 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

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

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

Estimate statistical significance The distribution of $ES(G)$ under the null is not analytically tractable and is obtained using Monte Carlo hypothesis testing[24]. Under the null hypothesis, H_0: $\lambda_1 = \lambda_2 = \cdots = \lambda_N = \lambda_0$. Under this hypothesis, $n_{1i} \sim Poisson(n_{\cdot i} \times \lambda_0)$, for $i = 1, \ldots, N$. Based on the relationship between Poisson and Multinonial distributions, the joint distribution of $(n_{11}, \ldots, n_{1N})$, conditioning on $\sum_{i=1}^{N} n_{1i} = n_{1\cdot}$ and $(n_{\cdot 1}, \ldots, n_{\cdot N})$ is

$$(n_{11}, n_{12}, \ldots, n_{1N}) | n_{1\cdot}; n_{\cdot 1}, \ldots, n_{\cdot N} \sim \text{multinomial}\,(n_{1\cdot}, \{r_1, \ldots, r_N\}) \tag{3}$$

where $r_i = \frac{n_{\cdot i}\lambda_0}{n_{\cdot\cdot}\lambda_0} = \frac{n_{\cdot i}}{n_{\cdot\cdot}}$.

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 [25] converting p-values into q-values to control the false discovery rate. Moreover, we can 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.

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_{1i}	$n_{\cdot i} - n_{1i}$
No	$n_{\cdot i} - n_{1i}$	$(n_{\cdot\cdot} - n_{1\cdot}) - (n_{\cdot i} - n_{1i})$

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

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

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, $\mathbb{1}(p_0^G \leq p_0^{G^c})$ ensures that an enriched group has the proportion of zeros smaller or equal to remaining groups.

Estimate statistical significance 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.

Results and Discussion

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 (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.

As shown in Figure 1, AEKS and the AEFisher performed significantly better than the GSEA and the conventional Fisher's exact test, respectively.

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 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 significantly enriched if q-value < 0.1.

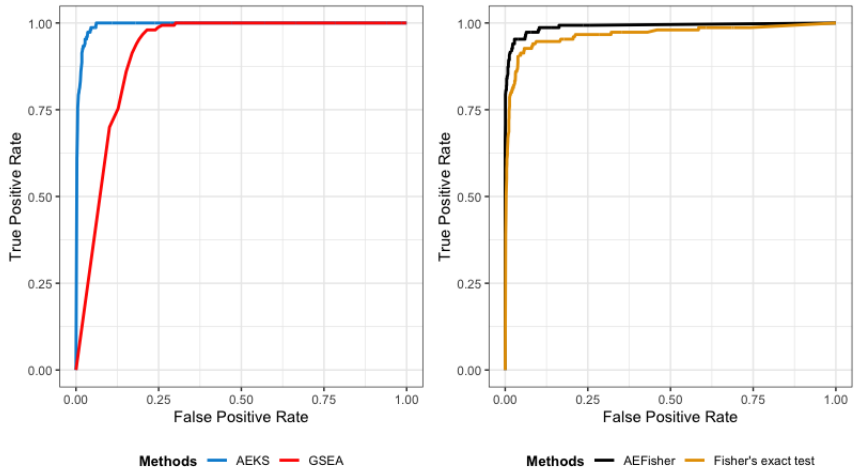

Figure 1: ROC curves using K-S statistics (AEKS vs GSEA) (left) and Fisher's exact tests (AEFisher vs Fisher's exact test) (right). In the Fisher exact test, a signaled AE is defined if q-value < 0.1 and OR > 1.5. Results were summarized over 10 simulated datasets.

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 139353 and 21820 reports for IIV and LAIV, respectively. We compared AE profiles with LAIV relative to IIV.

As shown in Table 3, AEKS and AEFisher identified the same 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 [26, 27, 28], New signals, such as epistaxis, is clinically interesting, and it might be true signals that need to be validated in large healthcare databases.

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 53415, 33087, 10356 reports with Havrix, Engerix-B, and Twinrix, respectively. In this study, AEKS and AEFisher identified different AE groups. As shown in Table 4, AEKS identified peripheral neuropathies, while AEFisher identified musculoskeletal and connectivetissue disorders. Peripheral neuropathies were also mentioned in [29] as an important AE group associated with the combination hepatitis vaccine.

Table 3: The enriched AE groups and significant AEs using AEKS and AEFisher to study LAIV relative to IIV.

Methods	Significant groups	Significant AEs
AEKS & AEFisher	Respiratory tract infections	Croup infectious
		Influenza
		Nasopharyngitis
		Pneumonia
		Rhinitis
		Sinusitis
		Upper respiratory tract infection
	Upper respiratory tract disorders (excl infections)	Anosmia
		Epistaxis
		Nasal congestion
		Nasal oedema
		Seasonal allergy
		Sinus disorder
		Stridor
		Tonsillar hypertrophy

Table 4: The enriched AE groups and significant AEs using AEKS and AEFisher to study hepatitis A and B combination relative to monovalent Hepatitis A and B vaccines

Methods	Significant AE group	Significant AEs
AEKS	Peripheral neuropathies	Guillain-Barre syndrome
		Miller Fisher syndrome
		Neuropathy peripheral
AEFisher	Musculoskeletal and connective tissue disorders NEC	Back pain
		Mobility decreased
		Musculoskeletal disorder
		Musculoskeletal pain
		Neck pain
		Pain in extremity

Conclusions

In this article, we develop new methods for vaccine adverse event enrichment analysis. We extend the existing gene enrichment tests by incorporating the special features of the AE count data. We have demonstrated the advantage of our methods over the existing methods in simulation studies. 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.

Abbreviations

AE: adverse event **RR:** reporting ratio
OR: odds ratio **HLGT:** high level group terms
GSEA: gene set enrichment analysis **ES:** enrichment score
VAERS: vaccine adverse event reporting system **MedDRA:** medical dictionary for regulatory activities

Declarations

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 Michigan Institute for Clinical and Health Research (grant UL1TR002240) and by the Propelling Original Data Science pilot grant from the Michigan Institute for Data Science. The funding body was not involved in the design of the study and collection, analysis, and interpretation of data or in writing the manuscript.

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

Author's contributions
SL and LZ developed the enrichment tests. SL performed simulation and real data analysis. SL and LZ participated in writing of the manuscript. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Consent for publication
Not applicable.

Ethics approval and consent to participate
Not applicable.

Author details
[1]Department of Statistics, University of Michigan, 1085 S University Ave, Ann Arbor, USA. [2]Department of Biostatistics, University of Michigan, Ann Arbor, USA.

References

1. Varricchio, F., Iskander, J., DeStefano, F., Ball, R., Pless, R., Braun, M.M., Chen, R.T.: Understanding vaccine safety information from the vaccine adverse event reporting system. Pediatr Infect Dis J. **23**, 287–294 (2004)
2. Shimabukuro, T.T., Nguyen, M., Martin, D., DeStefano, F.: Safety monitoring in the vaccine adverse event reporting system (vaers). Vaccine **33**(36), 4398–4405 (2015)
3. DuMouchel, W.: Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. The American Statistician **53**, 177–190 (1999)
4. Evans, S.J., Waller, P.C., Davis, S.: Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. Pharmacoepidemiology and Drug Safety **10**, 483–486 (2001)
5. van Puijenbroek, E.P., Bate, A., Leufkens, H.G., Lindquist, M., R, R.O., Egberts, A.C.: A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions. Pharmacoepidemiology and Drug Safety **11**, 3–10 (2002)
6. Bate, A., Lindquist, M., Edwards, I.R., Olsson, S., Orre, R., Lansner, A., Freitas, R.M.D.: A Bayesian neural network method for adverse drug reaction signal generation. European Journal of Clinical Pharmacology **54**, 315–321 (1998)
7. Lansner, R.O.A., Bate, A., M. Lindquist, I.R.E., Olsson, S., Orre, R., Lansner, A., Freitas, R.M.D.: Bayesian neural networks with confidence estimations applied to data mining. Comput Stat Data Anal **34**, 473–493 (2000)
8. Nóren, G.N., Bate, A., Orre, R., Edwards, I.R.: Extending the methods used to screen the WHO drug safety database towards analysis of complex associations and improved accuracy for rare events. BCPNN **25**, 3740–3757 (2006)
9. 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, 67–76 (2001)
10. Szarfman, A., Machado, S.G., O'Neill, R.T.: 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 **25**, 381–392 (2002)
11. Kulldorff, M., Davis, R.L., Kolczak, M., Lewis, E., Lieu, T., Platt, R.: A maximized sequential probability ratio test for drug and vaccine safety surveillance. Sequential Analysis **30**, 58–78 (2011)
12. Davis, R.L., Kolczak, M., Lewis, E., Nordin, J., Goodman, M., Shay, D.K., Platt, R., Black, S., Shinefield, H., Chen, R.T.: Active surveillance of vaccine safety: a system to detect early signs of adverse events. Epidemiology **16**, 336–341 (2005)
13. Li, L., Kulldorff, M.: A conditional maximized sequential probability ratio test for pharmacovigilance. Statistics in Medicine **29**, 284–295 (2009)
14. Li, R., Stewart, B., Weintraub, E., , McNeil, M.M.: Continuous sequential boundaries for vaccine safety surveillance. Statistics in Medicine **33**, 3387–3397 (2014)
15. Kulldorff, M., Dashevsky, I., Avery, T.R., Chan, K.A., Davis, R.L., Graham, D., Platt, R., Andrade, S.E., Boudreau, D., Gunter, M.J., Herrinton, L.J., Pawloski, P., Raebel, M.A., Roblin, D., Brown, J.S.: Drug safety data mining with a tree-based scan statistic. Pharmacoepidemiology and Drug Safety **22**, 517–523 (2013)
16. Mozzicato, P.: Meddra: an overview of the medical dictionary for regulatory activities. Pharm Med **23**, 65–75 (2009)
17. Zhang, Y., Tao, C., He, Y., Kanjamala, P., H, H.L.: Network-based analysis of vaccine-related associations reveals consistent knowledge with the vaccine ontology. Journal of Biomedical Semantics **4**, 33 (2013)
18. 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 **4**, 40 (2013)

19. Sarntivijai, S., Xiang, Z., Shedden, K.A., Markel, H., Omenn, G.S., Athey, B.D., He, Y.: Ontology-based combinatorial comparative analysis of adverse events associated with killed and live influenza vaccines. PLoS ONE **7**, 49941 (2012)
20. Guo, A., Racz, R., Hur, J., Lin, Y., Xiang, Z., Zhao, L., Rinder, J., Jiang, G., Zhu, Q., He, Y.: Ontology-based collection, representation and analysis of drug-associated neuropathy adverse events. J Biomed Semantics. **7**, 29 (2016)
21. Zeeberg, B.R., Feng, W., Wang, G., Wang, M.D., Fojo, A.T., Sunshine, M., Narasimhan, S., Kane, D.W., Reinhold, W.C., Lababidi, S., *et al.*: Gominer: a resource for biological interpretation of genomic and proteomic data. Genome biology **4**(4), 28 (2003)
22. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., *et al.*: Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences **102**(43), 15545–15550 (2005)
23. Mootha, V.K., Lindgren, C.M., Eriksson, K.-F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., *et al.*: Pgc-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nature genetics **34**(3), 267–273 (2003)
24. Huang, L., Zalkikar, J., Tiwari, R.C.: A likelihood ratio test based method for signal detection with application to fda's drug safety data. Journal of the American Statistical Association **106**(496), 1230–1241 (2011)
25. 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 **27**(1), 77–83 (2002)
26. Baxtera, R., Eatona, A., Hansena, J., Aukesa, L., Caspardb, H., Ambroseba, C.S.: Safety of quadrivalent live attenuated influenza vaccinein subjects aged 2–49 years. Vaccine **35**, 1254–1258 (2017)
27. Haber, P., Moro, P.L., Cano, M., Lewis, P., Stewart, B., T.Shimabukuro, T.: Post-licensure surveillance of quadrivalent live attenuated influenza vaccine united states,vaccine adverse event reporting system(VAERS), july 2013–june 2014. Vaccine **33**, 1987–1992 (2015)
28. Lambkin-Williams, R., Gelder, C., Broughton, R., Mallett, C.P., Gilbert, A.S., Mann, A., He, D., Oxford, J.S., Burt, D.: 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 **11**, 0163089 (2016)
29. 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 **6**, 3418 (2016)