Investigating safety profiles of human papillomavirus vaccine across group differences using VAERS data and MedDRA

Yuxi Jia[1,2,*], Cong Zhu[3,*], Jingcheng Du[2], Yang Xiang[2], Yong Chen[4], Wei Wang[1] and Cui Tao[2]

[1] Department of Medical Informatics, School of Public Health, Jilin University, Changchun, Jilin Province, China
[2] School of Biomedical Informatics, University of Texas Health Center at Houston, Houston, TX, USA
[3] Department of Epidemiology, Human Genetics, and Environmental Sciences, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, USA
[4] Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, PA, USA
* These authors contributed equally to this work.

Corresponding authors
Wei Wang, w_w@jlu.edu.cn
Cui Tao, cui.tao@uth.tmc.edu

ABSTRACT

Background: The safety of vaccines is a critical factor in maintaining public trust in national vaccination programs. This study aimed to evaluate the safety profiles of human papillomavirus (HPV) vaccines with regard to the distribution of adverse events (AE) across gender and age, and the correlations across various AEs using the Food and Drug Administration/Centers for Disease Control and Prevention Vaccine Adverse Event Reporting System (VAERS).

Methods: For analyses, 27,348 patients aged between 9 and 25 years old with at least one AE reported in VAERS between the year of 2006 and 2017 were included. AEs were summarized into two levels: the lower level preferred term (PT) and higher level system organ classes (SOCs) based on the structure of Medical Dictionary for Regulatory Activities (MedDRA). A series of statistical analyses were applied on both levels of AEs. Zero-truncated Poisson regression and multivariate logistic regression models were first developed to assess the rate and risk of SOCs across age groups and genders. Pairwise Pearson correlation analyses and hierarchical clustering analyses were then conducted to explore the interrelationships and clustering pattern among AEs.

Results: We identified 27,337 unique HPV vaccine reports between 2006 and 2017. Disproportional reporting of AEs was observed across age and gender in 21 SOCs ($p < 0.05$). The correlation analyses found most SOCs demonstrate weak positive correlations except for five pairs which were negatively correlated: skin and subcutaneous tissue disorders + injury poisoning and procedural complications; skin and subcutaneous tissue disorders + nervous system disorders; Skin and subcutaneous tissue disorders + pregnancy, puerperium and perinatal conditions; nervous system disorders + pregnancy, puerperium and perinatal conditions; pregnancy, puerperium and perinatal conditions + general disorders and administration site conditions. Nervous system disorders had the most AEs which contributed to 12,448 (46%) cases. In the further analyses of correlations between

PT in nervous system disorders, the three most strongly correlated AEs were psychiatric disorders ($r = 0.35$), gastrointestinal disorders ($r = 0.215$), and musculoskeletal and connective tissue disorders ($r = 0.261$). We observed an inter-SOCs correlation of the PTs among AE pairs by nervous system disorders/psychiatric disorders/gastrointestinal disorders/musculoskeletal and connective tissue disorders.

Conclusions: The analyses revealed a different distribution pattern of AEs across gender and age subgroups in 21 SOC level AEs. Correlation analyses and hierarchical clustering analyses further revealed several correlated patterns across various AEs. However, findings from this study should be interpreted with caution. Further clinical studies are needed to understand the heterogeneity of AEs reporting across subgroups and the biological pathways among the statistically correlated AEs.

BACKGROUND

Based on statistics collected by the *Centers for Disease Control and Prevention (CDC) (2017b)*, approximately 33,700 incidences of cancer per year are associated with human papillomavirus (HPV) infection. HPV is considered a leading cause of cervical cancers (*Walboomers et al., 1999*), respiratory papilloma, as well as of several types of cancers of the vagina, vulva, penis, anus, rectum and oropharynx (*Hoory et al., 2008*; *Lele et al., 2002*). To prevent HPV infection and HPV related cancers (*Centers for Disease Control and Prevention (CDC), 2016a*; *Cutts et al., 2007*; *Sundaram et al., 2017*), an HPV vaccine was first introduced in 2006 and has demonstrated its effectiveness based on post-market surveillance data. Studies showed that HPV vaccines contributed to a 64% reduction in vaccine-type HPV infections among teenage girls in the U.S. (*Centers for Disease Control and Prevention (CDC), 2016c*) and potentially prevents 31,200 (90%) HPV related cancers per year (*Centers for Disease Control and Prevention (CDC), 2017b*). According to CDC guidelines, HPV vaccines can be given to people beginning at age 9 years and is recommended for ages 11 through 26 (*Centers for Disease Control and Prevention (CDC), 2016b*). Ideally people should receive the HPV vaccines before they become sexually active and are exposed to HPV (*Centers for Disease Control and Prevention (CDC), 2016b*). In 2018 the U.S. Food and Drug Administration (FDA) expanded the approved use of the HPV vaccine to include women and men aged 27 through 45 years (*Food and Drug Administration (FDA), 2018*). With the increase in HPV vaccination coverage (*Centers for Disease Control and Prevention (CDC), 2018*; *Curtis et al., 2013*), pharmacovigilance should be further strengthened.

The distribution and cause of adverse events (AEs) has been an integral topic of pharmacovigilance (*Du et al., 2017*; *Harpaz et al., 2011*; *Tao et al., 2015*; *Wilson et al., 2018*). Post market surveillance is one of the common approaches to investigate vaccine AEs (*McPhillips & Marcuse, 2001*). The Vaccine Adverse Event Reporting System

(VAERS), a national early warning system that detects potential safety concerns in U.S. licensed vaccines (*Vaccine Adverse Event Reporting System (VAERS), 2018*), is one of the popular resources explored by various post-market vaccine safety studies (*Haber et al., 2016*; *Iskander, Haber & Herrera, 2005*; *Moro et al., 2015*; *Slade et al., 2009*; *Xie et al., 2016*). The topics of these studies ranged from validating results of clinical research (*Block et al., 2010*; *Gold, Buttery & McIntyre, 2010*; *Li et al., 2016*; *Stokley et al., 2014*), general description of AE types (*Levi et al., 2013*), identifying safety signals and causal relationships (*Debeer et al., 2008*; *Gee et al., 2011*, *2016*; *Lu et al., 2011*; *Pomfret, Gagnon & Gilchrist, 2011*), AEs and differences in categories of AE across races (*Huang et al., 2018*). However, none of these studies investigated the association between the variation in HPV vaccination related AE categories and age or gender. In addition, the correlation or clustering among AE categories is underexplored, though it has been studied using a data source other than VAERS (*Chandler et al., 2017*).

This article reports on a retrospective observational study aimed to address the major two knowledge gaps above. In addition to identify differences among AEs, the novelty of this study was that it explored the underlying relationship among AEs, which could potentially advance future investigation into the biological or clinical relevance of these relationships. The investigation was based on the HPV vaccine AE records extracted from the VAERS database between 2006 and 2017.

MATERIALS AND METHODS

Methods overview

We implemented a retrospective observational study based on the VAERS data to address the following research questions: (1) Are the frequencies of AEs different across different gender and age groups? If significant differences are observed, it is essential to develop a more precise vaccine information statement for these targeted subgroups. (2) Are there any correlations among the AEs? Specifically, we explored whether some AEs were more likely to occur together. This study included subjects aged between 9 and 25 years old who received the following vaccine types: HPVX/HPV2/HPV4/HPV9 (HPV vaccine with no brand name/HPV Cervarix/HPV Gardasil/HPV Gardasil 9). We excluded records with any missing information regarding age, gender or symptoms. The final eligible sample size was 26,023 including 21,647 females and 4,376 males.

Data source mapping symptoms in VAERS

Vaccine Adverse Event Reporting System database included nearly 47,000 reports associated with a HPV vaccine as of December 2017. AE symptoms were annotated following the guidelines of Medical Dictionary for Regulatory Activities (MedDRA) (*Food and Drug Administration (FDA), 2016*). MedDRA is a standardized medical terminology for medical products. MedDRA consists of a five-level structural hierarchy: system organ classes (SOC), high level group term, high level term, prefer term (PT), lowest level term (LLT) (*International Council for Harmonisation, 2018*). The AE symptoms in the VAERS dataset were annotated using the PT level terms. To map PT level AE to SOC level, we followed the method from a previous study and developed a dictionary-like lookup table by using

Table 1 The SOC list. The MedDRA Terminology SOC List—Internationally Agreed Order. MedDRA Management Board endorsed the creation of an additional 27th SOC category named "product issue" in 2016. It has no major impact on this research.

No.	SOC
1	Infections and infestations
2	Neoplasms benign, malignant and unspecified (including cysts and polyps)
3	Blood and lymphatic system disorders
4	Immune system disorders
5	Endocrine disorders
6	Metabolism and nutrition disorders
7	Psychiatric disorders
8	Nervous system disorders
9	Eye disorders
10	Ear and labyrinth disorders
11	Cardiac disorders
12	Vascular disorders
13	Respiratory, thoracic and mediastinal disorders
14	Gastrointestinal disorders
15	Hepatobiliary disorders
16	Skin and subcutaneous tissue disorders
17	Musculoskeletal and connective tissue disorders
18	Renal and urinary disorders
19	Pregnancy, puerperium and perinatal conditions
20	Reproductive system and breast disorders
21	Congenital, familial and genetic disorders
22	General disorders and administration site conditions
23	Investigations
24	Injury, poisoning and procedural complications
25	Surgical and medical procedures
26	Social circumstances
27*	Product issues

the internationally agreed upon order of the SOC list (Table 1) (*Du et al., 2017*). The sort of the list is based upon the relative importance of each SOC, which was determined by the Expert Working Group (*Medical Dictionary for Regulatory Activities (MedDRA), 2016*). The "dictionary" converts the symptoms at the PT level in VAERS to a corresponding unique SOC. Based on the data update, the "dictionary" has been expanded and refined in a new round (Fig. 1).

There are 27 SOC categories defined by MedDRA. However, since the "product issue" category does not describe AEs in humans, any AE that belonged to this category was excluded from the analyses. Therefore, 26 SOCs was included in this study.

Statistical methods

Our analyses consisted of two major inter-coherent parts based on the retrospective cohort study design. The first part was to address the hypothesis whether individual
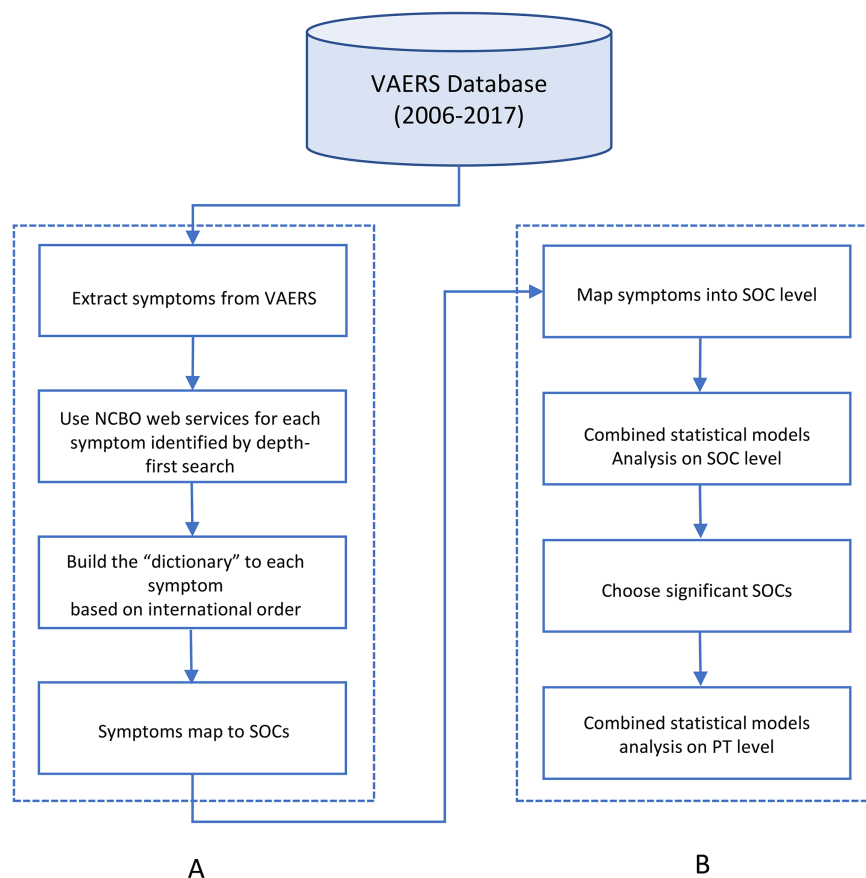
Figure 1 The flowchart of project. (A) was to map AEs from MedDRA PT level to MedDRA SOC level. (B) was to apply statistical models on mapped AEs.

SOC level AEs contributed to the total AE reports differently, and whether age and gender profiles vary across these individual SOC AEs. In addition to investigate the difference, in the second part of the analyses we explored the hypothesis that correlations or clustering patterns might exist among some AE categories, and examined the lower AEs term (PT level) that contributed to these potential observed correlations.

Corresponding to the analyses scheme part one, an independent t-test was performed to assess the age distribution across gender and specific SOCs. A multivariate logistic regression model was developed to evaluate the risk of developing a specific SOC, accounting for potential confounding effects by age and gender. Odds ratio (OR) greater than 1 indicates an increased odd of getting a specific SOC, whilst a less than 1 OR suggests decreased odds. A zero truncated Poisson regression model was used to model the number of events because patients need to have at least one AE in order to be captured by the VAERS system. Zero truncated Poisson regression model was performed to assess the number of individual SOC cases per year adjusted for gender and age.

For any SOCs that were statistically significant from either the logistic regression model or zero truncated Poisson regression model, the lower-level symptoms (PT) attributed to

these SOCs were further investigated using the multivariate logistic regression model adjusted for age and gender.

Finally, corresponding to the analyses scheme part two, pairwise Pearson correlation and un-supervised hierarchy clustering analyses were conducted to explore the interrelationships and clustering patterns across AEs. The analyses were first based on SOC level to identify the general pattern of the high-level AEs. Then PT level clustering analyses were performed on the SOC level AE that contributed to the majority of the AE cases and its most correlated three SOC level AEs. The hierarchical clustering method is based on the Euclidean distance between the vectors of Pearson correlation using a complete linkage algorithm. The optimal number of clusters was determined by the average silhouette method (*Rousseeuw, 1987*). Correlation network figures were then developed to summarize findings regarding correlation strength and clustering patterns of the PT level AEs. Only those with statistically significantly weak or above correlation ($r > 0.2$) are included in the Figures. Statistical significance level was defined as adjusted *p*-value < 0.05, after accounting for multiple testing using the Holm method (*Aickin & Gensler, 1996*). The details of the Holm method have been discussed elsewhere, in general, it controls type-I error (false positive) in a less conservative way than Bonferroni correction (*Bland & Altman, 1995*) with lowered risks in type II error (false negative).

All statistical analyses were based on R version 3.5.1 (*R Core Team, 2018*).

RESULTS

Table 2 lists the age distribution of SOCs by gender evaluated through independent *t*-test. Overall, the first vaccination age of females was relatively older than males according to the VAERS data (16.318 vs 14.14, 95% CI [2.073–2.283]), the difference was statistically significant ($p < 0.0019$). Similar distribution was observed across all individual SOC subgroups. The age difference was statistically significant among all SOCs except for neoplasms benign, malignant and unspecified (including cysts and polyps), ear and labyrinth disorders, hepatobiliary disorders and social circumstances.

Table 3 illustrates the association between the rate of developing any SOCs across gender and age. The interaction term between gender and age from the pooled model was statistically significant ($p < 0.0019$), which indicated the effect of age on developing any SOCs varied by gender. The subsequent stratification analyses shed further light on the difference of the effect by age variable. As shown in the male-only Poisson Regression model, with 1-year increase in age, the rate of developing SOCs among the observed cohort increased by 0.01 cases per person-year. In contrast, in the female-only model, each 1-year increase in age was associated with 0.004 decrease case per person-year on average.

Table 4 presents the odds of developing specific SOCs adjusting for gender, age and interaction term between gender and age. This table only includes top 10 SOCs with which contributed to the most AE. A complete table of 24 SOCs can be found in Table S1. Using nervous system disorders as an example for the interpretation of effects of the three variables: Among females, each 1 year increase in age is associated with 0.055 ($1 - \exp(-0.057 \times 1 + 0.131 \times 1 \times 0)$) decrease in the odds of developing nervous system disorders.

Table 2 Age distribution of SOCs by gender, VAERS 2006–2017[*].

SOC	Female mean age	Male mean age	Mean difference (95% CI)		p-value[**]	Sample size
ALL	16.318	14.14	2.073	2.283	<0.001	26,023
Nervous system disorders	15.853	14.506	1.202	1.492	<0.001	12,369
General disorders and administration site conditions	16.327	13.762	2.402	2.727	<0.001	10,537
Skin and subcutaneous tissue disorders	15.935	13.342	2.4	2.785	<0.001	6,688
Investigations	16.792	14.991	1.519	2.08	<0.001	5,790
Gastrointestinal disorders	16.277	14.248	1.773	2.284	<0.001	4,706
Musculoskeletal and connective tissue disorders	16.554	14.769	1.471	2.098	<0.001	4,241
Immune system disorders	16.169	14.144	1.742	2.307	<0.001	3,802
Injury, poisoning and procedural complications	15.941	14.74	0.91	1.49	<0.001	2,791
Psychiatric disorders	15.846	14.301	1.227	1.864	<0.001	2,711
Vascular disorders	15.776	14.243	1.229	1.837	<0.001	2,650
Infections and infestations	16.793	14.347	1.868	3.022	<0.001	1,962
Cardiac disorders	16.497	14.376	1.694	2.547	<0.001	1,803
Metabolism and nutrition disorders	16.113	13.331	2.245	3.32	<0.001	1,178
Respiratory, thoracic and mediastinal disorders	15.913	14.113	1.276	2.324	<0.001	1,083
Surgical and medical procedures	17.42	14.8	1.95	3.286	<0.001	979
Blood and lymphatic system disorders	16.779	15.176	0.782	2.423	0.002	773
Social circumstances	16.181	15.173	0.072	1.945	0.106	625
Eye disorders	15.511	14.505	0.281	1.731	0.034	621
Reproductive system and breast disorders	17.574	14.357	1.661	4.772	0.004	565
Endocrine disorders	16.55	14.5	0.751	3.35	0.032	537
Ear and labyrinth disorders	16.069	15.808	−1.381	1.904	0.847	228
Neoplasms benign, malignant and unspecified	16.769	13.583	0.873	5.499	0.044	220
Renal and urinary disorders	16.324	13.882	1.037	3.847	0.011	205
Hepatobiliary disorders	16.917	16.2	−1.182	2.615	0.847	53

Notes:
[*] SOCs congenital, familial and genetic disorders and pregnancy, puerperium and perinatal conditions were not included because their cases were exclusively females.
[**] p-value was adjusted for multiple testing using Holm method.

Table 3 Zero truncated Poisson regression of total population and by gender. The association between the rate of developing any SOCs across gender and age.

		Coefficient	p-value
Pooled	Age (Years)	−0.004	<0.001
	Male (Ref: Female)	−0.482	<0.001
	Age (Years) * Male	0.014	<0.001
Male	Age (Years)	0.010	<0.001
Female	Age (Years)	−0.004	<0.001

In contrast, among males, each 1 year increase in age is associated with ($\exp(-0.057 \times 1 + 0.131 \times 1 \times 1) - 1$) 0.077 increased odds of developing nervous system disorders.

As for PT level analyses, nervous system disorders was targeted because it contributed to the most adverse events (12,378) in VAERS data. As shown in Table 5, a disproportional

Table 4 Logistic Regression by Individual SOCs adjusted for gender, age and interaction term between gender and age. The odds of developing specific SOCs adjusting for gender, age and interaction term between gender and age. This table only includes top 10 SOCs with which contributed to the most AE. A complete table of 24 SOCs can be found in Table S1.

No.			Coefficient	Standard error	p-value
4	Immune system disorders	Age (in Year)	−0.011	0.005	0.225
		Gender (Male:1, Female:0)	−0.35	0.227	1.000
		Age * Gender	0.012	0.015	1.000
7	Psychiatric disorders	Age (in Year)	−0.034	0.006	<0.001
		Gender (Male:1, Female:0)	−0.951	0.259	0.004
		Age * Gender	0.053	0.018	0.043
8	Nervous system disorders	Age (in Year)	−0.057	0.003	<0.001
		Gender (Male:1, Female:0)	−2.018	0.157	<0.001
		Age * Gender	0.131	0.011	<0.001
12	Vascular disorders	Age (in Year)	−0.039	0.006	<0.001
		Gender (Male:1, Female:0)	−0.737	0.249	0.043
		Age * Gender	0.052	0.017	0.039
14	Gastrointestinal disorders	Age (in Year)	−0.003	0.004	1.000
		Gender (Male:1, Female:0)	−0.534	0.213	0.148
		Age * Gender	0.017	0.014	1.000
16	Skin and subcutaneous tissue disorders	Age (in Year)	−0.033	0.004	<0.001
		Gender (Male:1, Female:0)	1.725	0.181	<0.001
		Age * Gender	−0.106	0.013	<0.001
17	Musculoskeletal and connective tissue disorders	Age (in Year)	0.018	0.004	0.001
		Gender (Male:1, Female:0)	−1.134	0.226	<0.001
		Age * Gender	0.054	0.015	0.006
22	General disorders and administration site conditions	Age (in Year)	0.001	0.003	1.000
		Gender (Male:1, Female:0)	1.236	0.159	<0.001
		Age * Gender	−0.075	0.011	<0.001
23	Investigations	Age (in Year)	0.039	0.004	<0.001
		Gender (Male:1, Female:0)	−1.465	0.215	<0.001
		Age * Gender	0.058	0.014	0.001
24	Injury, poisoning and procedural complications	Age (in Year)	−0.027	0.006	<0.001
		Gender (Male:1, Female:0)	−1.011	0.22	<0.001
		Age * Gender	0.097	0.014	<0.001

reporting of PT level terms was found between the genders. For example, males experienced fewer headaches and hypoesthesia than females, 19.40% vs 24.20% and 3.97% vs 7.9%, respectively.

Figure 2 illustrates the overall correlation among the 26 SOCs. The majority of the SOCs seem to be positively connected except five pairs which demonstrate weak negative correlation (skin and subcutaneous tissue disorders + injury poisoning and procedural complications; skin and subcutaneous tissue disorders + nervous system disorders;

Table 5 Chi-square test of PT level symptoms from nervous system disorders*.

		Female		Male		p-value
		N	%	N	%	
Dizziness	No	6,982	67.50	1,359	66.70	0.454
	Yes	3,357	32.50	680	33.30	
Syncope	No	7,348	71.10	1,470	72.10	0.365
	Yes	2,991	28.90	569	27.90	
Headache	No	7,838	75.80	1,644	80.60	<0.001
	Yes	2,501	24.20	395	19.40	
Loss of consciousness	No	8,866	85.75	1,716	84.16	0.067
	Yes	1,473	14.25	323	15.84	
Hypoaesthesia	No	9,522	92.10	1,958	96.03	<0.001
	Yes	817	7.90	81	3.97	

Note:
* Only top five PT-level terms were demonstrated.
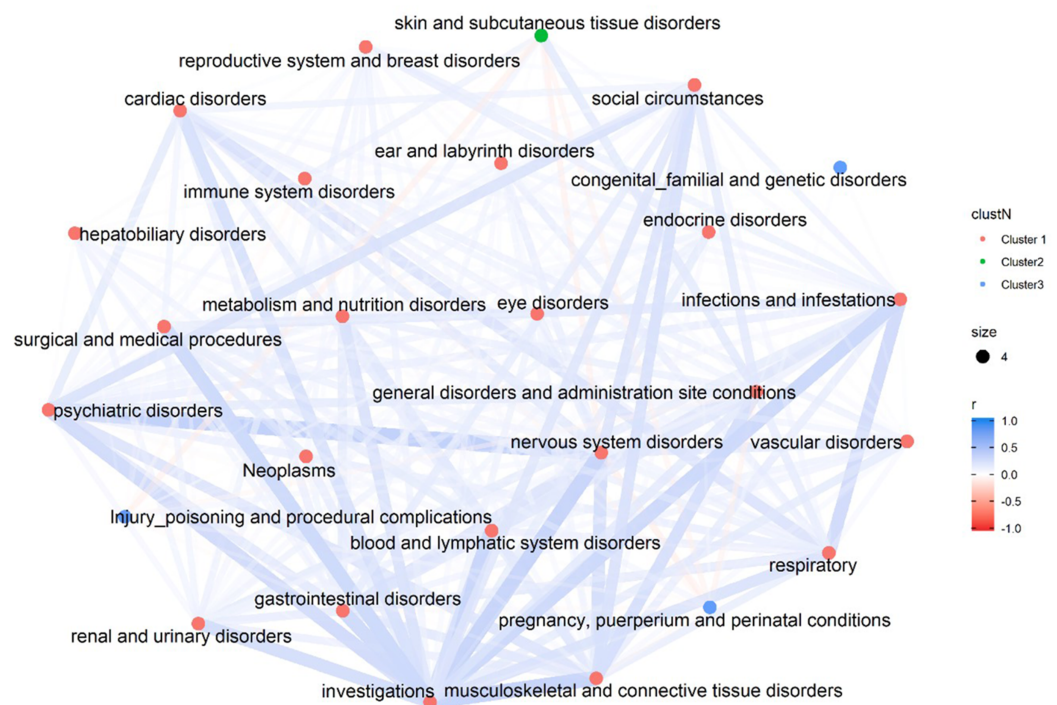
Figure 2 Correlation Network of 26 SOCs. The overall correlation among the 26 SOCs. The colors of the nodes differentiate the clusters of each SOC, which is determined by the hierarchical clustering analyses using a complete linkage algorithm.

skin and subcutaneous tissue disorders + pregnancy, puerperium and perinatal conditions; nervous system disorders + pregnancy, puerperium and perinatal conditions; pregnancy, puerperium and perinatal conditions + general disorders and administration site conditions).
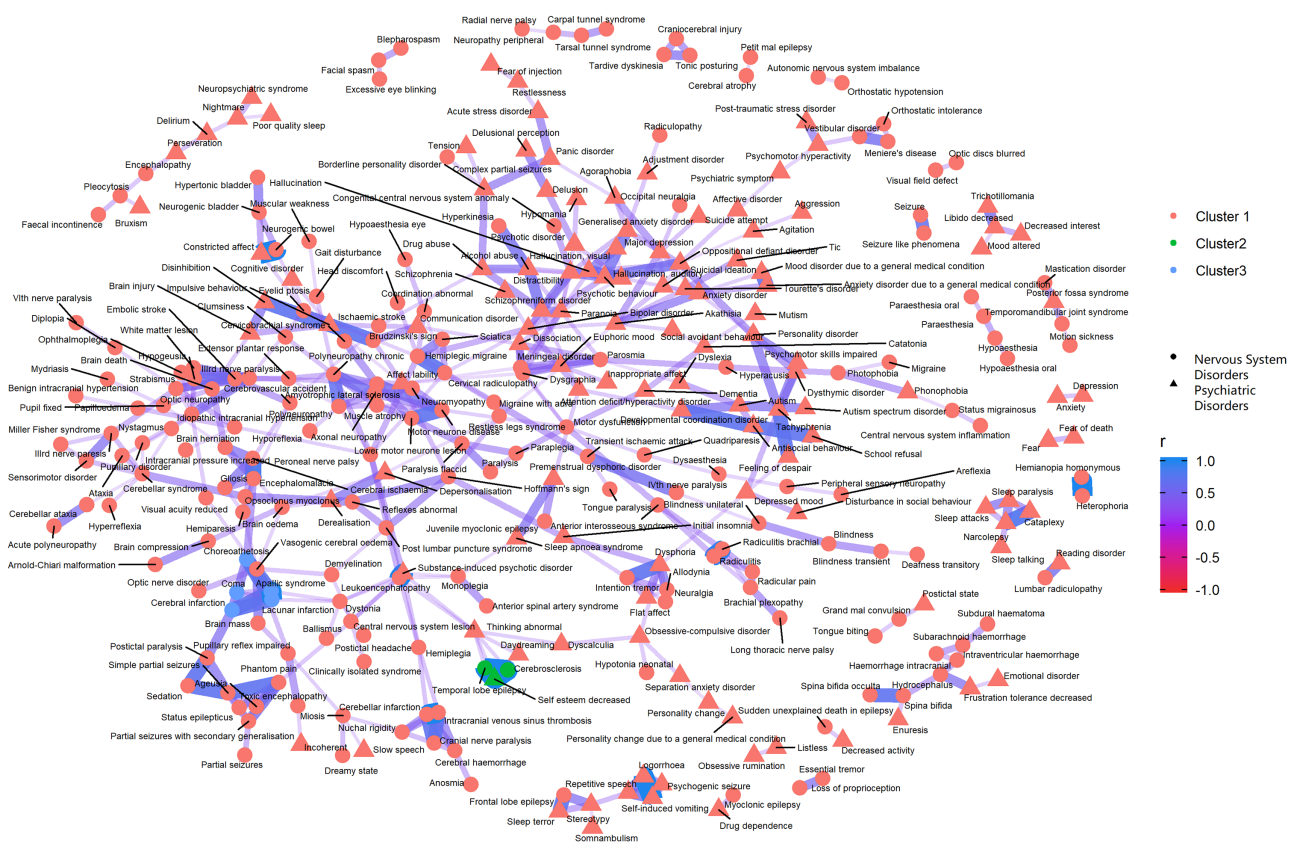
Figure 3 Correlation network of pairwise Pearson correlation of PT-term level AEs from nervous system disorders and psychiatric disorders. The PT-level interrelationship among AEs from nervous system disorders and psychiatric disorders.

Figures 3–5 show the PT-level interrelationship among AEs from four SOC groups: nervous system disorders, psychiatric disorders, gastrointestinal disorders and musculoskeletal and connective tissue disorders. The first one contributed the most AEs while the other three demonstrate clinical relevance and strongest correlation with nervous systems disorders ($r = 0.357, 0.215, 0.261$, respectively). The figures only display those AEs with Pearson correlation $r > 0.2$, and adjusted p-value < 0.05 (Holm method). An inter-SOCs correlation of the PTs was observed among nervous system disorders + psychiatric disorders and nervous system disorders + musculoskeletal and connective tissue disorders; the relatively strong correlations were not limited to PTs attributed to the same SOC category. For example, the correlation between Photophobia (nervous system disorders) and phonophobia (psychiatric disorders) was 0.38 ($p < 0.05$). This frequently observed inter-SOCs correlation may be due to the close clinical relationship among these three SOCs. However, this pattern was much less common in pairs of nervous system disorders and gastrointestinal disorders. Most of the weak or above correlations ($r > 0.2$) were found among PTs from the same SOC category. Only a few inter-SOCs correlation were identified, such as the pair of Intestinal obstruction and Migraine without aura ($r = 0.41$).

As for the clustering analyses, most of the AEs were recommended to group into one large cluster, with the rest categorized into two small clusters. The PTs within the same
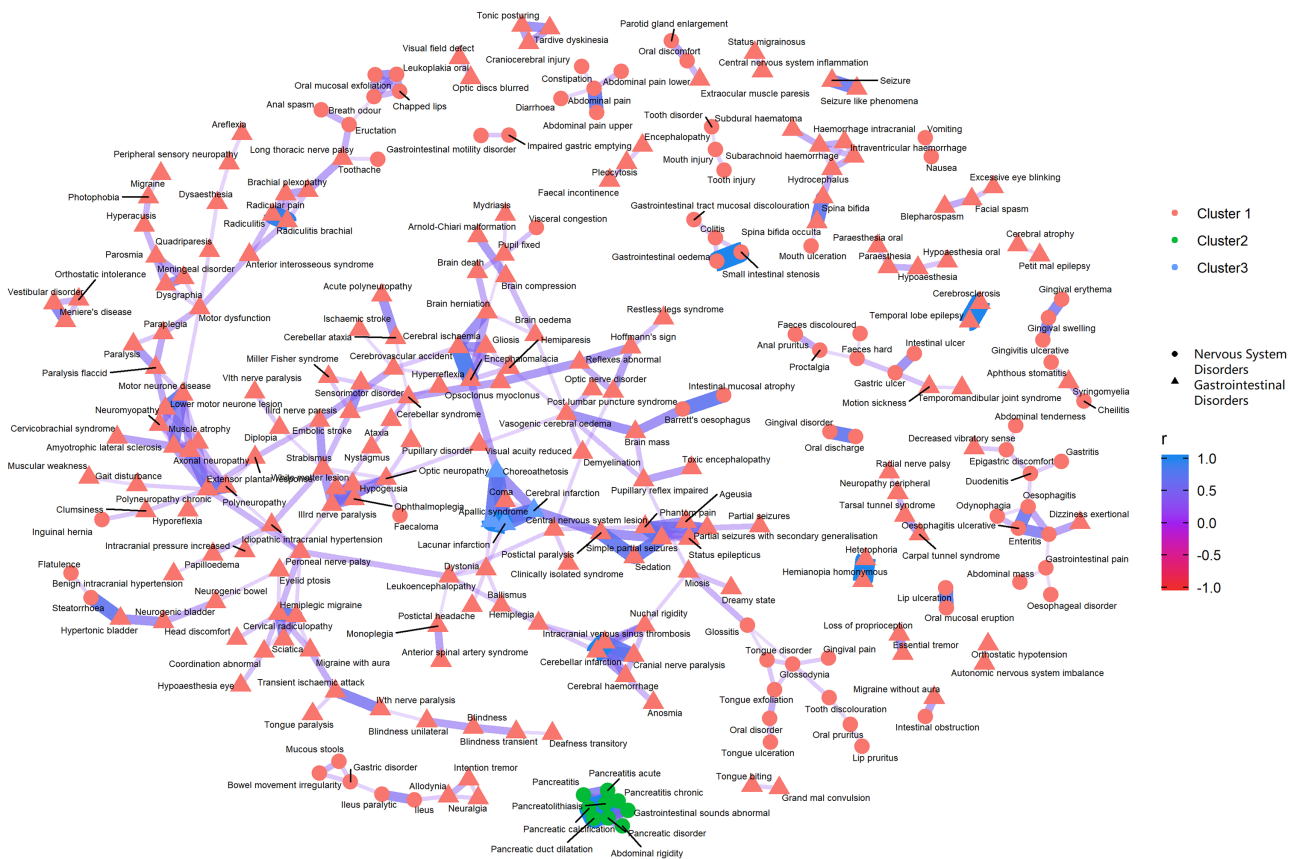
Figure 4 Correlation network of pairwise Pearson correlation of PT-term level AEs from nervous system disorders and gastrointestinal disorders. The PT-level interrelationship among AEs from nervous system disorders and gastrointestinal disorders.

cluster that demonstrated the most distinct clinical relevance were pancreatitis related AEs, abdominal rigidity, and gastrointestinal sounds abnormal (green cluster, Fig. 4).

DISCUSSION

Our analyses revealed a variation in the distribution of AEs across gender and age groups. Generally, males reported AEs at younger first vaccination ages than females. None of the SOC level AEs which demonstrated statistical significance with age and gender met FDA's standards of serious AEs. On average, each person experienced four SOC level AEs. Among males, older vaccination age was associated with a slight increase in reported SOC level AEs, while the trend was reversed among females.

In the clustering analyses which investigated the top SOC subgroup nervous system disorders and its three most correlated AE subgroups psychiatric disorders, musculoskeletal and connective tissue disorders and gastrointestinal disorders, we identified many weak and above ($r > 0.2$) inter subgroup correlation among PT pairs from nervous system disorders/psychiatric disorders and nervous system disorders/musculoskeletal and connective tissue disorders. However, the same level correlations among the pool of AEs by nervous system disorders and gastrointestinal disorders were mostly confined to
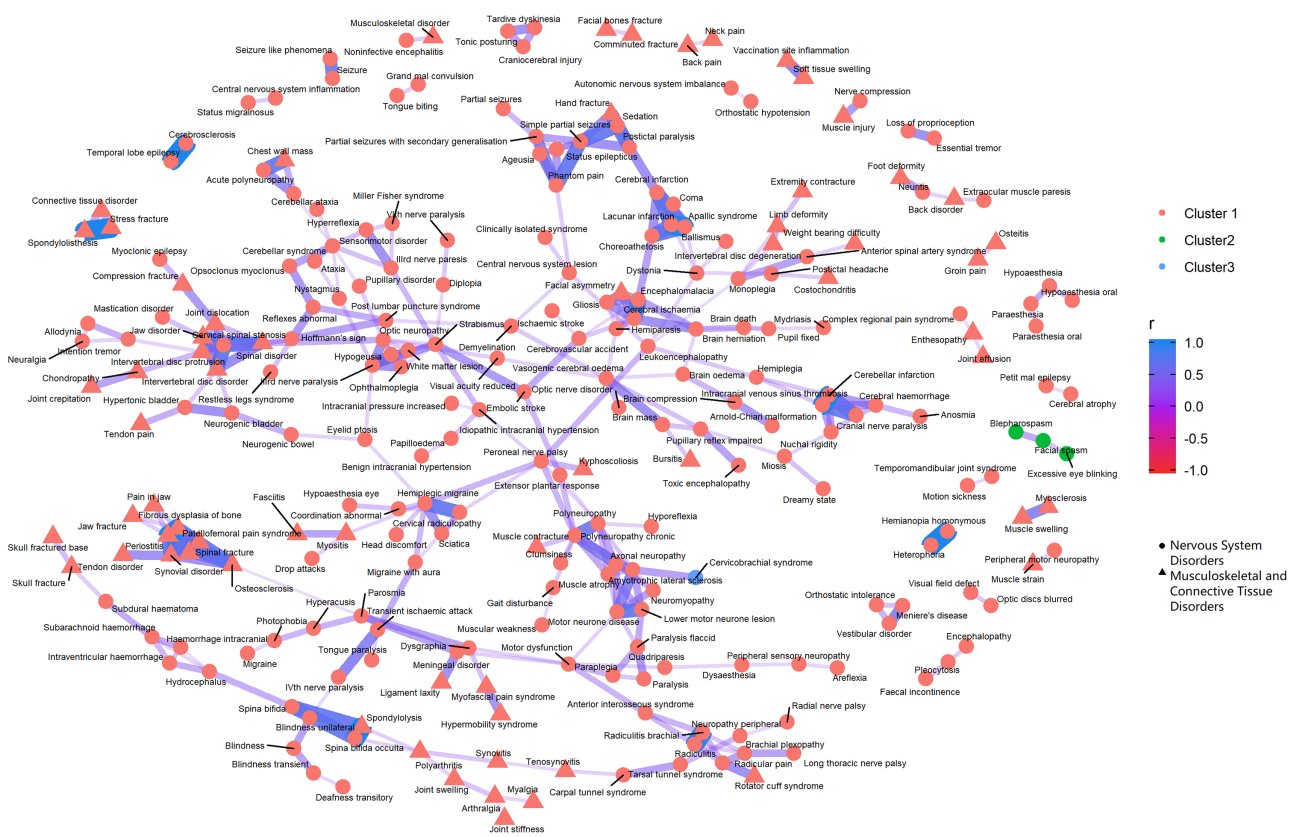
Figure 5 Correlation network of pairwise Pearson correlation of PT-term level AEs from nervous system disorders and musculoskeletal and connective tissue disorders. The PT-level interrelationship among AEs from nervous system disorders and musculoskeletal and connective tissue disorders.

PTs from the same SOC category. While the silhouette recommended that PT level AEs be divided into three clusters, whether such categorization is clinically informative demands further investigations.

The results of all of the statistical analyses from this study should be interpreted with caution. Since VAERS exclusively collected adverse reactions data (events post vaccination), the model based on its data only evaluates the risk or rate of developing specific adverse reactions among those who have already experienced adverse reactions. Our findings do not imply any real risk of developing adverse events after receiving HPV vaccination in the general population. Based on the CDC data, roughly half of adolescents received HPV vaccination by 2017 (*Centers for Disease Control and Prevention (CDC), 2017a*). Considering the 42 million population among the 10–19 years old group (*The Statista Portal, 2017*), we estimate the number of vaccinated adolescents to be approximately 21 million. This number should be even larger if considering the cohort effect (those who entered the adult cohort before 2017). However, only 26,023 adverse events were reported between 2006 and 2017 to VAERS, among which the non-serious AEs (dizziness, syncope, headache, etc. Table 4) contributed to the majority. In addition, events such as Syncope, fainting and short-term loss of consciousness are common

in adolescents regardless of the vaccination type. Therefore, our study didn't reveal any findings that contradict the current HPV vaccine safety profile.

The clustering pattern of the PT level terms was determined using the un-supervised learning methodology instead of clinical interpretations. Further studies are warranted to explore the biological pathways between the correlated adverse reactions. Due to the limitation as a passive surveillance system, issues such as under-reporting, over-reporting, inconsistent reporting quality etc. exist (*Moro et al., 2015*). Especially, VAERS accepts all reports without judging whether the event was caused by the vaccine (*Vaccine Adverse Event Reporting System (VAERS), 2019*). The observed statistical significance in the analyses does not imply cause-effect relationships and needs to be further validated in relevant clinical studies.

CONCLUSIONS

To the best of our knowledge, this is the first investigation that systematically reviewed post-market safety profiles of HPV vaccines using VAERS data. The analyses revealed a different distribution pattern of AEs among different gender and age subgroups. This research contributes to the notion that the benefits of vaccination outweigh the potential risks.

We found 13 SOCs with significant differences in first vaccination. Based on the clustering analyses of selected AEs from nervous system disorders, psychiatric disorders, musculoskeletal and connective tissue disorders and gastrointestinal disorders, we identified inter-dependences among AEs from different SOC level categories. VAERS data provide valuable HPV vaccines safety data. However, findings from this study should be interpreted with caution. Further clinical studies are warranted to understand the heterogeneity of AEs reporting across subgroups and the biology pathways among the statistically correlated AEs.

Funding

This work was supported by the Cancer Prevention Research Institute of Texas (CPRIT) Training Grant #RP160015. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosure

The following grant information was disclosed by the authors:
Cancer Prevention Research Institute of Texas (CPRIT) Training: #RP160015.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Yuxi Jia performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.

- Cong Zhu performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Jingcheng Du analyzed the data, contributed reagents/materials/analysis tools, approved the final draft.
- Yang Xiang contributed reagents/materials/analysis tools, approved the final draft.
- Yong Chen conceived and designed the experiments, approved the final draft.
- Wei Wang conceived and designed the experiments, approved the final draft.
- Cui Tao conceived and designed the experiments, approved the final draft.

Data Availability
Raw data is available in Table 4 and Table S1. Table 4 presents the odds of developing specific SOCs adjusting for gender, age and interaction term between gender and age. This table only includes SOCs with statistical significance in either age or gender parameters. A complete table of 26 SOCs can be found in Table S1.

Supplemental Information
Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.7490#supplemental-information.

REFERENCES

Aickin M, Gensler H. 1996. Adjusting for multiple testing when reporting research results: the Bonferroni vs Holm methods. *American Journal of Public Health* **86(5)**:726–728 DOI 10.2105/AJPH.86.5.726.

Bland JM, Altman DG. 1995. Multiple significance tests: the Bonferroni method. *BMJ* **310(6973)**:170 DOI 10.1136/bmj.310.6973.170.

Block SL, Brown DR, Chatterjee A, Gold MA, Sings HL, Meibohm A, Dana A, Haupt RM, Barr E, Tamms GM, Zhou H, Reisinger KS. 2010. Clinical trial and post-licensure safety profile of a prophylactic human papillomavirus (Types 6, 11, 16, and 18) L1 virus-like particle vaccine. *Pediatric Infectious Disease Journal* **29(2)**:95–101 DOI 10.1097/INF.0b013e3181b77906.

Centers for Disease Control and Prevention (CDC). 2016a. How well does HPV vaccine work? *Available at* https://www.cdc.gov/hpv/parents/questions-answers.html (accessed 1 March 2019).

Centers for Disease Control and Prevention (CDC). 2016b. HPV vaccine recommendations. *Available at* https://www.cdc.gov/vaccines/vpd/hpv/hcp/recommendations.html (accessed 1 March 2019).

Centers for Disease Control and Prevention (CDC). 2016c. Will sexually active females benefit from the vaccine? *Available at* https://www.cdc.gov/std/hpv/stdfact-hpv-vaccine-young-women.htm (accessed 1 March 2019).

Centers for Disease Control and Prevention (CDC). 2017a. HPV vaccination coverage data. *Available at* https://www.cdc.gov/hpv/hcp/vacc-coverage/index.html (accessed 1 March 2019).

Centers for Disease Control and Prevention (CDC). 2017b. Why is HPV vaccine important? *Available at* https://www.cdc.gov/hpv/hcp/hpv-important.html (accessed 1 March 2019).

Centers for Disease Control and Prevention (CDC). 2018. HPV vaccination coverage data. *Available at* https://www.cdc.gov/hpv/hcp/vacc-coverage/index.html (accessed 1 March 2019).

Chandler RE, Juhlin K, Fransson J, Caster O, Edwards IR, Norén GN. 2017. Current safety concerns with human papillomavirus vaccine: a cluster analysis of reports in VigiBase. *Drug Safety* **40(1)**:81–90 DOI 10.1007/s40264-016-0456-3.

Curtis CR, Yankey D, Jeyarajah J, Dorell C, Stokley S, MacNeil J, Hariri S. 2013. National and state vaccination coverage among adolescents aged 13–17 years—United States, 2012. *MMWR Morbidity and Mortality Weekly Report* **62**:685–693.

Cutts FT, Franceschi S, Goldie S, Castellsague X, De Sanjose S, Garnett G, Edmunds WJ, Claeys P, Goldenthal KL, Harperi DM, Markowitz L. 2007. Human papillomavirus and HPV vaccines: a review. *Bulletin of the World Health Organization* **85(9)**:719–726 DOI 10.2471/BLT.06.038414.

Debeer P, De Munter P, Bruyninckx F, Devlieger R. 2008. Brachial plexus neuritis following HPV vaccination. *Vaccine* **26(35)**:4417–4419 DOI 10.1016/j.vaccine.2008.06.074.

Du J, Cai Y, Chen Y, He Y, Tao C. 2017. Analysis of individual differences in vaccine pharmacovigilance using VAERS data and MedDRA system organ classes: a use case study with trivalent influenza vaccine. *Biomedical Informatics Insights* **9**:117822261770062 DOI 10.1177/1178222617700627.

Food and Drug Administration (FDA). 2016. What is a serious adverse event? *Available at* https://www.fda.gov/Safety/MedWatch/HowToReport/ucm053087.htm (accessed 1 March 2019).

Food and Drug Administration (FDA). 2018. FDA approves expanded use of Gardasil 9 to include individuals 27 through 45 years old. *Available at* https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm622715.htm (accessed 1 March 2019).

Gee J, Naleway A, Shui I, Baggs J, Yin R, Li R, Kulldorff M, Lewis E, Fireman B, Daley MF, Klein NP, Weintraub ES. 2011. Monitoring the safety of quadrivalent human papillomavirus vaccine: findings from the vaccine safety datalink. *Vaccine* **29(46)**:8279–8284 DOI 10.1016/j.vaccine.2011.08.106.

Gee J, Weinbaum C, Sukumaran L, Markowitz LE. 2016. Quadrivalent HPV vaccine safety review and safety monitoring plans for nine-valent HPV vaccine in the United States. *Human Vaccines & Immunotherapeutics* **12(6)**:1406–1417 DOI 10.1080/21645515.2016.1168952.

Gold MS, Buttery J, McIntyre P. 2010. Human papillomavirus vaccine safety in Australia: experience to date and issues for surveillance. *Sexual Health* **7(3)**:320–324 DOI 10.1071/SH09153.

Haber P, Moro PL, Lewis P, Woo EJ, Jankosky C, Cano M. 2016. Post-licensure surveillance of quadrivalent inactivated influenza (IIV4) vaccine in the United States, vaccine adverse event reporting system (VAERS), July 1, 2013–May 31, 2015. *Vaccine* **34(22)**:2507–2512 DOI 10.1016/j.vaccine.2016.03.048.

Harpaz R, Perez H, Chase HS, Rabadan R, Hripcsak G, Friedman C. 2011. Biclustering of adverse drug events in the FDA's spontaneous reporting system. *Clinical Pharmacology & Therapeutics* **89(2)**:243–250 DOI 10.1038/clpt.2010.285.

Hoory T, Monie A, Gravitt P, Wu T. 2008. Molecular epidemiology of human papillomavirus. *Journal of the Formosan Medical Association* **107(3)**:198–217 DOI 10.1016/S0929-6646(08)60138-2.

Huang J, Du J, Duan R, Zhang X, Tao C, Chen Y. 2018. Characterization of the differential adverse event rates by race/ethnicity groups for HPV vaccine by integrating data from different sources. *Frontiers in Pharmacology* **9**:539 DOI 10.3389/fphar.2018.00539.

International Council for Harmonisation. 2018. Introductory Guide MedDRA Version 21.0. *Available at* https://www.meddra.org/sites/default/files/guidance/file/intguide_21_0_english.pdf (accessed 1 March 2019).

Iskander J, Haber P, Herrera G. 2005. Monitoring vaccine safety during an influenza pandemic. *Yale Journal of Biology and Medicine* **78**:265.

Lele SM, Pou AM, Ventura K, Gatalica Z, Payne D. 2002. Molecular events in the progression of recurrent respiratory papillomatosis to carcinoma. *Archives of Pathology & Laboratory Medicine* **126**:1184–1188.

Levi M, Bonanni P, Burroni E, Bechini A, Boccalini S, Sani C, Bonaiuti R, Indiani L, Azzari C, Lippi F, Carozzi F, Brandigi L, Canessa C, Di Pierro C, Iossa A, Lastrucci V, Sala A, Tanini T, Tiscione E. 2013. Evaluation of bivalent human papillomavirus (HPV) vaccine safety and tolerability in a sample of 25 year old Tuscan women. *Human Vaccines & Immunotherapeutics* **9(7)**:1407–1412 DOI 10.4161/hv.24337.

Li R, Stewart B, McNeil MM, Duffy J, Nelson J, Kawai AT, Baxter R, Belongia EA, Weintraub E. 2016. Post licensure surveillance of influenza vaccines in the vaccine safety datalink in the 2013-2014 and 2014-2015 seasons. *Pharmacoepidemiology and Drug Safety* **25(8)**:928–934 DOI 10.1002/pds.3996.

Lu B, Kumar A, Castellsagué X, Giuliano AR. 2011. Efficacy and safety of prophylactic vaccines against cervical HPV infection and diseases among women: a systematic review & meta-analysis. *BMC Infectious Diseases* **11(1)**:13 DOI 10.1186/1471-2334-11-13.

McPhillips H, Marcuse EK. 2001. Vaccine safety. *Current Problems in Pediatrics* **31(4)**:95–121 DOI 10.1016/S1538-5442(01)70025-9.

Medical Dictionary for Regulatory Activities (MedDRA). 2016. 27th system organ class. *Available at https://www.meddra.org/About%20MedDRA%20/%20Evolution%20/%2027th-system-organ-class* (accessed 1 March 2019).

Moro PL, Jankosky C, Menschik D, Lewis P, Duffy J, Stewart B, Shimabukuro TT. 2015. Adverse events following *Haemophilus influenzae* type b vaccines in the vaccine adverse event reporting system, 1990-2013. *Journal of Pediatrics* **166(4)**:992–997 DOI 10.1016/j.jpeds.2014.12.014.

Pomfret TC, Gagnon JM Jr, Gilchrist AT. 2011. Quadrivalent human papillomavirus (HPV) vaccine: a review of safety, efficacy, and pharmacoeconomics. *Journal of Clinical Pharmacy and Therapeutics* **36(1)**:1–9 DOI 10.1111/j.1365-2710.2009.01150.x.

R Core Team. 2018. *R: A language and environment for statistical computing*. Version 3.5.1. Vienna: R Foundation for Statistical Computing. *Available at https://www.R-project.org/*.

Rousseeuw PJ. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* **20**:53–65 DOI 10.1016/0377-0427(87)90125-7.

Slade BA, Leidel L, Vellozzi C, Woo EJ, Hua W, Sutherland A, Izurieta HS, Ball R, Miller N, Braun MM, Markowitz LE, Iskander J. 2009. Postlicensure safety surveillance for quadrivalent human papillomavirus recombinant vaccine. *JAMA* **302(7)**:750–757 DOI 10.1001/jama.2009.1201.

Stokley S, Jeyarajah J, Yankey D, Cano M, Gee J, Roark J, Robinette Curtis C, Markowitz L. 2014. Human papillomavirus vaccination coverage among adolescents, 2007–2013, and postlicensure vaccine safety monitoring, 2006–2014-United States. *MMWR Morbidity and Mortality Weekly Report* **63**:620–624.

Sundaram A, Vaughan B, Kost K, Bankole A, Finer L, Singh S, Trussell J. 2017. Contraceptive failure in the United States: estimates from the 2006-2010 national survey of family growth: contraceptive failure rates in the U.S. *Perspectives on Sexual and Reproductive Health* **49(1)**:7–16 DOI 10.1363/psrh.12017.

Tao C, Du J, Cai Y, Chen Y. 2015. Trivalent influenza vaccine adverse event analysis based on MedDRA system organ classes using VAERS data. *Studies in Health Technology and Informatics* **216**:1076 DOI 10.3233/978-1-61499-564-7-1076.

The Statista Portal. 2017. Resident population of the United States by sex and age as of July 1, 2017 (in millions). *Available at* https://www.statista.com/statistics/241488/population-of-the-us-by-sex-and-age/ (accessed 1 March 2019).

Vaccine Adverse Event Reporting System (VAERS). 2018. Background and public health importance. *Available at* https://vaers.hhs.gov/about.html (accessed 1 March 2019).

Vaccine Adverse Event Reporting System (VAERS). 2019. VAERS data limitations. *Available at* https://vaers.hhs.gov/data/dataguide.html (accessed 1 March 2019).

Walboomers JM, Jacobs MV, Manos MM, Bosch FX, Kummer JA, Shah KV, Snijders PJ, Peto J, Meijer CJ, Muñoz N. 1999. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *Journal of Pathology* **189**:12–19 DOI 10.1002/(SICI)1096-9896(199909)189:1<12::AID-PATH431>3.0.CO;2-F.

Wilson E, Zhu C, Galea S, Marko A, Victoria Urdaneta V, Straus W. 2018. Turning up the heat: effect of new vaccine for children's (VFC) program recommendations for use of temperature monitors upon incorrect product storage adverse event reporting. *Vaccine* **36(11)**:1516–1520 DOI 10.1016/j.vaccine.2017.10.059.

Xie J, Zhao L, Zhou S, He Y. 2016. Statistical and ontological analysis of adverse events associated with monovalent and combination vaccines against Hepatitis A and B diseases. *Scientific Reports* **6(1)**:34318 DOI 10.1038/srep34318.