Are All Vaccines Created Equal?
Using Electronic Health Records to Discover Vaccines Associated With Clinician-Coded Adverse Events

Mary Regina Boland, MA[1,4] and Nicholas P Tatonetti, PhD[1-4]
[1]Department of Biomedical Informatics, [2]Department of Medicine, [3]Department of Systems Biology, [4]Observational Health Data Sciences and Informatics, Columbia University

Abstract

Adverse drug events (ADEs) are responsible for unnecessary patient deaths making them a major public health issue. Literature estimates 1% of ADEs recorded in Electronic Health Records (EHRs) are reported to federal databases making EHRs a vital source of ADE-related information. Using Columbia University Medical Center (CUMC)'s EHRs, we developed an algorithm to mine for vaccine-related ADEs occurring within 3 months of vaccination. In phase one, we measured the association between vaccinated patients with an ADE (cases) against those vaccinated without an ADE. To adjust for healthcare-process effects, phase two compared cases against those who returned to CUMC within 3 months without an ADE. We report 7 results passing multiplicity correction after demographic confounder adjustment. We observed an association, having some literature support, between swine flu vaccination and ADEs (H1N1v-like, OR=9.469, p<0.001; H1N1/H3N2, OR=3.207, p<0.001). Our algorithm could inform clinicians of the risks/benefits of vaccinations towards improving clinical care.

1. Introduction

1.1 Adverse Drug Events Are Important for Public Health

Adverse drug events (ADEs) are a major cause of death in the United States of America [1]. To address this serious public health issue, the Federal Drug Administration (FDA) developed an adverse event reporting system. Since this reporting system began, more than 75 drugs or drug products have been removed from public use [2]. The number of ADEs occurring between 1998 and 2005 increased 2.6 fold illustrating the increasing importance of ADE prevention in clinical care [3]. Many ADE detection methods rely on adequate physician, pharmacist, or nurse reporting of the ADE to federal reporting systems. Realizing that 1% of ADEs recorded in Electronic Health Records (EHRs) are reported on the federal level [4], we chose to harness the large set of clinician-reported ADEs available in EHRs to find novel vaccine-ADE associations.

1.2 Informatics Methods Enable Harnessing of Data Within EHRs

The widespread adoption of EHRs enables meaningful use [5] of data recorded during the clinical encounter. Appropriate use of EHR data requires overcoming definition discrepancies [6], data sparseness and quality [7], bias [8], and healthcare process effects [9]. Informatics methods overcome these challenges by employing standardized ontologies to minimize definition discrepancies [10-12], measuring concordance across integrated datasets for data sparseness and quality assessment [7], and minimizing bias and healthcare process effects using statistical methods [13]. Informatics methods applied to EHRs [14] have been successful in diverse areas [15-17] including pharmacovigilance [18, 19]. They are also useful in predicting ADEs using chemical and molecular structures of compounds [20]. Approaching the problem from a different angle, our method investigates ADEs occurring and recorded during routine clinical care.

1.3 ADE Detection and Prevention Feasible Using EHRs

Multiple algorithms have shown the usefulness of EHRs for ADE detection. Haerian et al. developed a method for identifying drugs associated with two serious ADEs, rhabdomyolysis and agranulocytosis, after adjusting for patient comorbidities [19]. Luo et al. developed a pattern mining method for detecting ADEs from clinical trials data [21]. Linder et al. found that only 1% of EHR recorded ADEs are reported to the federal government, demonstrating that EHRs are a rich data source for ADE detection [4].

2. Materials and Methods

2.1 Columbia University Medical Center Dataset

We used EHR data from Columbia University Medical Center (CUMC), previously converted to the Common Data Model (CDM) [22] developed by the Observational Medical Outcomes Partnership (OMOP). This dataset contains patients' drug-related and diagnosis information. The CUMC Institutional Review Board approved this study.

2.2 An Algorithm to Mine for Vaccines Associated with Adverse Events

We mapped all *International Classification of Diseases, version 9* (ICD-9) codes to the *Systemized Nomenclature of Medicine – Clinical Terms* (SNOMED-CT) using the OMOP CDM v.4 [22], which was proven useful by a number of prior research studies [23, 24]. By taking advantage of the medication-terminology mapping in the CDM (which

includes both RxNORM and NDF-RT)[22] we are able to map many different vaccines from different manufacturers to the same core ingredient set. Others obtained high quality results when using this same CDM mapping for medications [24]. Using the CDM also helps minimize terminology mapping issues common when using EHRs for medication information [25].

In the OMOP CDM [22], one code for "adverse effect due to correct medicinal substance properly administered" maps to 75 ICD-9 codes (each with relatively low prevalence). We used this mapping and extracted a population of 16,296 patients with a coded ADE from an appropriately prescribed and administered drug. Because we were interested in vaccines, we extracted all patients who were vaccinated in our medical system (N=70,050). Subsequently, we recorded patients as having a vaccine-related ADE if the ADE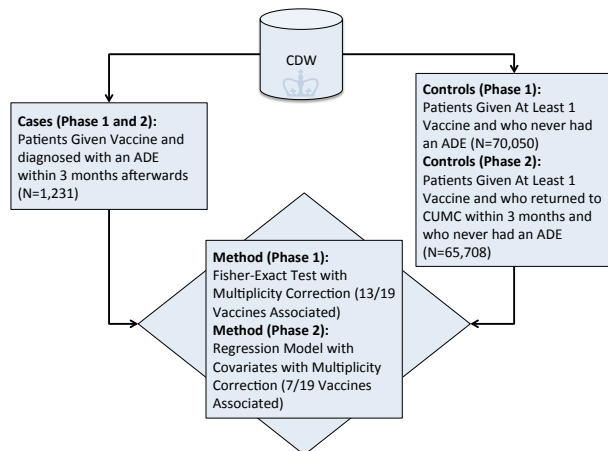 occurred within a 3-month window (i.e., 90 days) after the vaccination date. We selected a 3-month window because there is literature suggesting that over 8 weeks time may be necessary to appropriately capture a vaccine-related ADE [26]. If several ADEs occurred within the 3-month time frame (e.g., one 2 days, and another 7 days after vaccination) then both were included in the analysis. This was done because both clinician-coded ADEs could be the result of the vaccination.

Figure 1. **Algorithm Schema to Detect Vaccines Associated with Clinician-Coded Adverse Events**

2.2.1 Phase One: Mining Vaccine-ADE Associations Across All Vaccinated Patients

The first part of our algorithm (**Figure 1**) calculates the association between each vaccine and an ADE within 3 months by comparing each individual vaccine (case) to all other vaccines in our dataset (as controls). Controls include all patients who were vaccinated regardless of whether they returned to the hospital for a follow-up visit. Associations are measured using the fisher-exact test with multiplicity correction using Bonferroni's method (R v.3.1.0).

2.2.2 Phase Two: Mining Vaccine-ADE Associations Adjusting for Health-Care Process and Demographic Effects
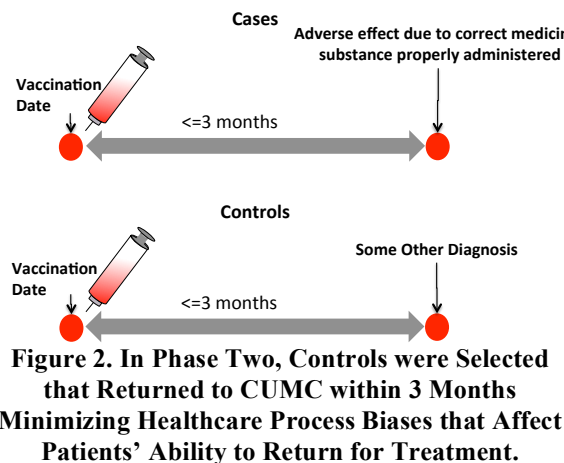

Figure 2. In Phase Two, Controls were Selected that Returned to CUMC within 3 Months Minimizing Healthcare Process Biases that Affect Patients' Ability to Return for Treatment.

To adjust for various health-care process effects [8, 9] that may affect whether or not a patient returns to CUMC within 3 months, we decided to use as controls all patients who were vaccinated and were subsequently diagnosed with some other medical condition (not an ADE) within 3 months of the vaccination date. Our cases remained unchanged and consisted of all vaccinated patients with an ADE diagnosis within 3 months. Therefore, in this second phase of the algorithm both cases and controls returned to CUMC within 3 months. For this analysis, we had 65,708 controls and the same 1,231 cases (**Figure 1**). We measured the association between each vaccination and an ADE diagnosis using logistic regression. Specifically, each potential confounder (i.e., ethnicity, race, sex, age (at time of vaccination)) was modeled as a covariate in the logistic regression equation with the binary response (outcome) variable indicating the presence or absence of an ADE within 3 months of vaccination and the predictor variable denoting presence or absence of the vaccine of interest (R v.3.1.0). An association is reported as significant if the Bonferroni adjusted p-value is <=0.05. We further illustrate phase two's control selection method in **Figure 2**.

3. Results

3.1 Overview of CUMC Dataset

Our dataset contained 472,451 patients with both medication and diagnosis-related information. We found 19 vaccines prescribed at CUMC with at least one patient with a recorded ADE within 3 months after vaccination. In total, 1,231 vaccinated patients were diagnosed with an ADE within 3 months, and **Figure 3** depicts their characteristics.

3.2 Vaccine-ADE Algorithm

3.2.1 Phase One: Mining for Associations Across All Vaccinated Patients

We applied our algorithm to all 19 vaccines and measured the association between the vaccine's administration and an ADE within 3 months afterwards. After Bonferroni adjustment, we found 13 vaccines were associated with an ADE at this step. Characteristics of our case patients are shown in **Figure 3** (note that the case population did not change between methods). All results from both phases of the algorithm are provided in **Table 1** (following page).

3.2.2 Phase Two: Mining for Associations After Adjusting for Healthcare Process, and Demographics

During phase two, we constructed a logistic regression model with covariates for age, ethnicity, race and sex. We report results passing multiplicity correction (7 of 19 vaccines).
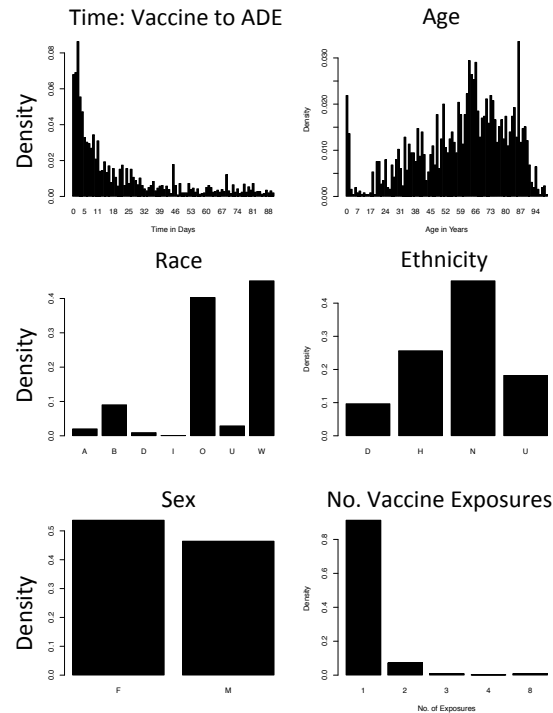

Figure 3. Characteristics of Patients That Developed an ADE within 3 months after Vaccination

Four vaccines were significantly associated with more ADEs and three vaccines were significantly associated with fewer ADEs compared with other vaccines. Two of the four vaccines associated with more ADEs were vaccines against flu originating in swine including: H1N1/H3N2/inactivated B-Brisbane-60-2008 strain and H1N1v-like virus vaccine. For the H1N1/H3N2 combo vaccine, 503 of 6,904 patients returning within 3 months experienced an ADE. Further, for the H1N1v-like vaccine, 103 of 423 patients returning within 3 months experienced an ADE. Both were significant after adjusting for demographic confounders (adjusted p<0.001 for both, **Table 1).** In several instances we found that all patients who returned to CUMC within 3 months had an ADE diagnosis. This includes five vaccines typically given to infants: pertussis/diphtheria/tetanus; hepatitis B surface antigen; tetanus; diphtheria/haemophilus B; Polio 3 types. While interesting, none of these vaccines were significant after adjusting for demographic confounders and multiplicity.

4. Discussion

4.1 Important Vaccine-ADE Associations

Vaccine-related ADEs can result from a number of different mechanisms important for achieving precision medicine [27]. Our two-phase algorithm was developed specifically for finding vaccines associated with clinician-coded ADEs in EHRs and was agnostic to the mechanism underlying the vaccine-ADE relationship.

Interestingly, two types of swine flu vaccines were positively associated with an increased risk of an ADE within 3 months of vaccination after adjustment for confounders (**Table 1),** namely the combo H1N1 / H3N2 / B-brisbane influenza (OR=3.207, p<0.001) vaccine and the influenza A-California-7-2009-(H1N1)v-like virus (OR=9.469, p<0.001) vaccine. Importantly, H1N1 originates in swine [28] and all swine flu vaccines in our study were associated with increased risk of ADEs. This fits well with prior literature supporting vaccine-related ADEs resulting from a different swine flu vaccine in the 1970s [26], which resulted in very serious ADEs including paralysis. Another study, found a similar result for H1N1 vaccination when compared to general influenza vaccination [29].

4.2 Value of Clinician-Coded ADE Associations

Early detection of ADEs is crucial for patient safety. Using our algorithm, we uncovered several vaccines that resulted in ADEs within 3 months for all patients who returned to CUMC within 3 months (**Table 1).** This was true for several vaccines given to infants. Although the results are not significant after covariate modeling (age is one confounder) it is suggestive of a relationship that may warrant further exploration. There are two main types of Hepatitis B vaccinations at CUMC: Hepatitis B (surface antigen) at a concentration of 0.04 mg/ml and Hepatitis B recombinant at a concentration of 0.01-0.02 mg/ml. Vaccination by the higher dose Hepatitis B vaccine resulted in 9 patients with ADEs out of 9 patients with the vaccine who returned to CUMC within 3 months (100% developed an ADE). Contrastingly, vaccination by the lower dose Hepatitis B recombinant vaccine (half to one-quarter the potency) resulted in ADEs among 1 of 1,617 patients seen at CUMC within 3 months after vaccination. Neither hepatitis vaccine was significantly associated with ADEs after adjusting for age, sex, ethnicity, and race. Patients receiving the higher dose Hepatitis B vaccine were less likely to return within 3 months (9/1474, **Table 1)** then those receiving the lower dose (1617/1659, **Table 1).** A likely explanation is that a higher proportion of infants received the lower dose (0.1-0.2 mg/ml) vaccine (94.45% of those vaccinated were <=0 years); whereas, both infants and toddlers received the higher dose (0.4 mg/ml) vaccination (65.94% of those vaccinated were <=0 years; 27.34% were one year olds).

Table 1. Vaccine-ADE Results for Phase 1 and 2 Trials.

Shortened Vaccine Name	Origin Organism	No. Cases[1]	Phase 1: Association Between Vaccine Administration and Adverse Effect				Phase 2: Adjusting for Demographics and Only Including Patients Returning Within 3 Months			
			No. Vaccinated	Prop.[2]	Odds Ratio (OR)	Adj. P[4]	No. Vaccinated who Returned within 3 Months	Prop.[3]	OR	Adj. P[4]
Associated with ADEs After Confounder Adjustment:										
Mumps	Human	5	785	0.006	0.451	1	8	0.625	106.793	9.54×10^{-9}
H1N1/ H3N2 / Inactivated B-Brisbane-60-2008 strain	Swine Virus (first 2), Human Virus	503	19517	0.026	2.119	3.10×10^{-41}	6904	0.073	3.207	5.86×10^{-103}
Pertussis / Diphtheria/ Haemophilus b / Polio / Tetanus	Bacteria (first 3), virus, bacteria	68	11575	0.006	0.399	1.83×10^{-16}	163	0.417	31.216	9.42×10^{-90}
H1N1v-like virus vaccine (0.25-0.5 mg/ml)	Swine Virus	103	3701	0.028	2.070	1.90×10^{-9}	423	0.243	9.469	7.70×10^{-79}
Associated with Fewer ADEs After Confounder Adjustment:										
Pneumococcal Type 1, 10A, 11A, 12F	Bacteria	1186	34983	0.034	4.149	5.68×10^{-230}	33976	0.035	0.406	7.20×10^{-66}
Rubella	Human	21	13568	0.002	0.101	1.21×10^{-56}	13565	0.002	0.055	2.01×10^{-37}
Pertussis / Diphtheria/ Hepatitis B Surface Antigen (0.02 mg/ml)	Bacteria (first 2), Primate virus	64	11406	0.006	0.381	2.30×10^{-17}	10660	0.006	0.208	8.65×10^{-33}
Insignificant After Confounder Adjustment:										
Hepatitis B (0.01 or 0.02 mg/ml)	Primate virus	1	1659	0.001	0.042	5.37×10^{-8}	1617	0.001	0.056	0.079
Varicella-Zoster Live (Oka-Merck)[Varivax]	Vertebrate Virus	3	99	0.030	2.208	1	30	0.1	5.618	0.103
Pertussis / Diphtheria / Tetanus [Infanrix]	Bacteria	59	11301	0.005	0.353	3.35×10^{-19}	59	1	8982647	1
Hepatitis B (0.04 mg/ml)	Primate virus	9	1474	0.006	0.431	0.133	9	1	135359422	1
Tetanus	Bacteria	57	11301	0.005	0.341	4.28×10^{-20}	57	1	89762087	1
Diphtheria / Haemophilus B	Bacteria	49	10242	0.005	0.325	2.11×10^{-19}	49	1	90639367	1
Polio Types 1-3	Virus	5	1453	0.003	0.242	0.002	5	1	82665283	1
Meningococcal Group A/C/W/Y	Bacteria	10	291	0.034	2.519	0.161	273	0.037	1.154	1
Measles / Mumps / Rubella	Human	21	13571	0.002	0.101	1.23×10^{-56}	21	1	129593962	1
Streptococcus Pneumonia	Bacteria	3	545	0.006	0.390	1	56	0.054	2.665	1
Haemophilus B	Bacteria	13	13	0.008	0.600	1	1545	1	104864930	1
Diphtheria / Tetanus	Bacteria	58	11301	0.005	0.347	1.42×10^{-19}	58	1	89078106	1

[1]Vaccinated and ADE within 3 months

[2]Cases / No. Vaccinated

[3]Cases / No. Vaccinated and Returned to CUMC Within 3 months

[4]Adjustment made using Bonferroni. Only Bonferroni-adjusted p-values <=0.05 were considered significant.

Infants have more wellness visits per year; therefore, vaccines given to a higher proportion of infants would be expected to have a higher return rate within 3 months (which we observed). This also demonstrates how the healthcare process can affect results of retrospective analyses using EHRs. Importantly, we adjusted for these types of biases in our algorithm by comparing patients receiving an ADE within 3 months to those who have returned to the hospital ADE-free within 3 months to help adjust for these biases. We also included age as a covariate in our regression model to adjust for age as well.

4.3 Limitations and Future Work

A limitation of our work includes our exclusive use of clinician recorded ADEs from EHRs. Some estimates suggest that only one-tenth of ADEs are clinician reported [30]. Therefore, we may be under-estimating the number of ADEs. We used only clinician-reported ADEs because we wanted to ensure that a clinician had validated the ADE as having occurred (i.e., a "true" ADE). Future work includes further exploration of dose-dependency effects for vaccine-related ADEs. Dosage data was only available for some vaccines at this stage. However, we hope to include clinical text and other data types in future to further tease out dosage effects and their relation to ADE risk.

5. Conclusion

We present an algorithm for discovering vaccines more likely to result in clinician-reported ADEs within 3 months of vaccination when compared to other vaccines. Our method found several interesting associations including two swine flu vaccinations that are positively associated with ADEs within 3 months of vaccination after confounder adjustment.

References
1. Lazarou J, Pomeranz B, Corey P. Incidence of adverse drug reactions in hospitalized patients: A meta-analysis of prospective studies. JAMA. 1998;279(15):1200-5.
2. Wysowski DK, Swartz L. Adverse drug event surveillance and drug withdrawals in the united states, 1969-2002: The importance of reporting suspected reactions. Archives of Internal Medicine. 2005;165(12):1363-9.
3. Moore TJ, Cohen MR, Furberg CD. SErious adverse drug events reported to the food and drug administration, 1998-2005. Archives of Internal Medicine. 2007;167(16):1752-9.
4. Linder JA, Haas JS, Iyer A, Labuzetta MA, Ibara M, Celeste M, et al. Secondary use of electronic health record data: spontaneous triggered adverse drug event reporting. Pharmacoepidemiology and Drug Safety. 2010;19(12):1211-5.
5. Jha AK. Meaningful use of electronic health records: the road ahead. JAMA. 2010;304(15):1709-10.
6. Boland MR, Hripcsak G, Shen Y, Chung WK, Weng C. Defining a comprehensive verotype using electronic health records for personalized medicine. J Am Med Inform Assoc. 2013;20(e2):e232-8.
7. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. J Am Med Inform Assoc. 2013;20(1):144-51.
8. Hripcsak G, Knirsch C, Zhou L, Wilcox A, Melton G. Bias associated with mining electronic health records. J Biomed Discov Collab. 2011;6:48-52.
9. Hripcsak G, Albers DJ. Correlating electronic health record concepts with healthcare process events. J Am Med Inform Assoc. 2013;20(e2):e311-8.
10. Elkin PL, Brown SH, Husser CS, Bauer BA, Wahner-Roedler D, Rosenbloom ST, et al. Evaluation of the Content Coverage of SNOMED CT: Ability of SNOMED Clinical Terms to Represent Clinical Problem Lists. Mayo Clinic Proceedings. 2006;81(6):741-8.
11. Lin K, Hsieh A, Farzaneh S, Doan S, Kim H. Standardizing Phenotype Variables in the Database of Genotypes and phenotypes (dbGaP) based on Information Models. AMIA Jt Summits Transl Sci Proc 2013 2013;Mar 18:110.
12. Eilbeck K, Jacobs J, McGarvey S, Vinion C, Staes CJ, editors. Exploring the use of ontologies and automated reasoning to manage selection of reportable condition lab tests from LOINC. ICBO; 2013.
13. Hripcsak G, Knirsch C, Zhou L, Wilcox A, Melton GB. Using discordance to improve classification in narrative clinical databases: An application to community-acquired pneumonia. Computers in Biology and Medicine. 2007;37(3):296-304.
14. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. Nat Rev Genet. 2012;13(6):395-405.
15. Boland MR, Hripcsak G, Albers DJ, Wei Y, Wilcox AB, Wei J, et al. Discovering medical conditions associated with periodontitis using linked electronic health records. J Clin Periodontol. 2013;40(5):474-82.
16. Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, Brown-Gentry K, et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. Bioinformatics. 2010;26(9):1205-10.
17. Kohane IS. Using electronic health records to drive discovery in disease genomics. Nat Rev Genet. 2011;12(6):417-28.
18. Wang X, Hripcsak G, Markatou M, Friedman C. Active computerized pharmacovigilance using natural language processing, statistics, and electronic health records: a feasibility study. J Am Med Inform Assoc. 2009;16(3):328-37.
19. Haerian K, Varn D, Vaidya S, Ena L, Chase H, Friedman C. Detection of pharmacovigilance-related adverse events using electronic health records and automated methods. Clinical Pharmacology & Therapeutics. 2012;92(2):228-34.
20. Liu M, Wu Y, Chen Y, Sun J, Zhao Z, Chen X-w, et al. Large-scale prediction of adverse drug reactions using chemical, biological, and phenotypic properties of drugs. J Am Med Inform Assoc. 2012;19(e1):e28-e35.
21. Luo Z, Zhang G, Xu R. Mining patterns of adverse events using aggregated clinical trial results. AMIA Jt Summits Transl Sci Proc 2013. 2013;Mar 18:112-6.
22. Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. J Am Med Inform Assoc. 2012;19(1):54-60.
23. Boland MR, Tatonetti NP, Hripcsak G. CAESAR: a Classification Approach for Extracting Severity Automatically from Electronic Health Records. Intelligent Systems for Molecular Biology Phenotype Day. 2014;Boston, MA(In Press):1-8.
24. Ryan PB, Madigan D, Stang PE, Schuemie MJ, Hripcsak G. Medication-wide association studies. CPT: pharmacometrics & systems pharmacology. 2013;2:e76.
25. Pathak J, Chute CG. Analyzing categorical information in two publicly available drug terminologies: RxNorm and NDF-RT. J Am Med Inform Assoc. 2010;17(4):432-9.
26. Langmuir AD, Bregman DJ, Kurland LT, Nathanson N, Victor M. An Epidemiologic and Clinical Evaluation of Guillain-Barre Syndrome Reported in Association with the Administration of Swine Influenza Vaccines. American Journal of Epidemiology. 1984;119(6):841-79.
27. Peterson TA, Doughty E, Kann MG. Towards Precision Medicine: Advances in Computational Approaches for the Analysis of Human Variants. Journal of Molecular Biology. 2013;425(21):4047-63.
28. Trifonov V, Khiabanian H, Rabadan R. Geographic dependence, surveillance, and origins of the 2009 influenza A (H1N1) virus. New England Journal of Medicine. 2009;361(2):115-9.
29. Vellozzi C, Broder KR, Haber P, Guh A, Nguyen M, Cano M, et al. Adverse events following influenza A (H1N1) 2009 monovalent vaccines reported to the Vaccine Adverse Event Reporting System, United States, October 1, 2009–January 31, 2010. Vaccine. 2010;28(45):7248-55.
30. Classen DC, Resar R, Griffin F, Federico F, Frankel T, Kimmel N, et al. 'Global Trigger Tool' Shows That Adverse Events In Hospitals May Be Ten Times Greater Than Previously Measured. Health Affairs. 2011;30(4):581-9.