Project Submission

Zhihua Jin (zjin80)
zjin80@gatech.edu

Introduction

HPV (Human papillomavirus) infection is a common viral infection that can be sexually transmitted. In most occasions, HPV can go away on its own without causing health problems. However, it is the major cause of cervical cancer around the world. In over 100 types of HPV, 15 types are “high-risk” cancer-causing ones. Among them, HPV-16 and HPV-18 cause nearly 70% of all the cervical cancers, while some other types can lead to genital warts.

There are two kinds of vaccines approved by FDA in 2006 that can reduce young people (from 9 to 26 years old)’s chance of getting HPV infected ("NIH Fact Sheets - Cervical Cancer", 2019). They are Gardasil® and Cervarix®, both of which have been proved effective for the prevention of cervical cancer caused by HPV-16 and HPV-18. In 2015, nine-valent Gardasil 9 was approved to protect against HPV-31, HPV-33, HPV-45, HPV-52, and HPV-58, in addition to the previous two types. For both men and women, Gardasil 9 is 88% effective against HPV infections, despite the fact that most HPV-vaccinated people are women ("Gardasil", 2019). According to patients’ age, the vaccine may be given as 2 or 3-dose through arm muscle injection.

Inevitably, using either of these vaccines may lead to several side effects. The most common ones include fainting, swelling, redness, pain, headache, nausea, and fever, etc. ("GARDASIL®9 (Human Papillomavirus 9-valent Vaccine, Recombinant) | Official Site", 2019). These symptoms are usually temporal and mild, so they should disappear in a few minutes or a few days. However, there have been concerns about vaccines’ long-term and major side effects. For example, several patients in Japan reported chronic pain and other symptoms which lead to the suspension of HPV vaccination; while patients in Latin America showed marked decrease of immunization rates (Cervantes & Doan, 2018).

Some of the severe ones are rheumatoid arthritis, type 1 diabetes and multiple sclerosis ("Autoimmune Disease List – AARDA", 2019). The classic sign of autoimmune disease is inflammation and patients may undergo redness, heat, pain and swelling. In a review

paper (Genovese, LA Fauci, Squeri, Trimarchi & Squeri, 2018), a meta-analysis was conducted to find correlation between HPV and autoimmune diseases. The author concluded that no correlation was found for bivalent and quadrivalent HPV vaccines. However, these kinds of side effect have been constantly reported as adverse events by patients.

In this study, I investigate the relationship between HPV vaccine and autoimmune diseases, which are frequently mentioned in HPV-relevant researches. There are more than 100 types of autoimmune diseases, in which people’s immune system attack their own tissues. Some of the severe ones are rheumatoid arthritis, type 1 diabetes and multiple sclerosis ("Autoimmune Disease List – AARDA", 2019). The classic sign of autoimmune disease is inflammation and patients may undergo redness, heat, pain and swelling. In a review paper (Genovese, LA Fauci, Squeri, Trimarchi & Squeri, 2018), a meta-analysis was conducted to find correlation between HPV and autoimmune diseases. The author concluded that no correlation was found for bivalent and quadrivalent HPV vaccines. However, these kinds of side effect have been constantly reported as adverse events by patients.

Since nine-valent Gardasil 9 was introduced to the market in recent years, I conduct my own research and discover whether HPV vaccination of the 3 vaccines on the market increases the risk of autoimmune diseases. I present results obtained using data from SynPUF and reveal the effect of HPV vaccine concerning a few autoimmune diseases across diverse ethnic groups, populations and age groups. I expect to determine if HPV vaccine increase patients’ risk for autoimmune diseases compared with those who have not been vaccinated against HPV.

Methods

Patients are found in CMS 2008-2010 Data Entrepreneurs’ Synthetic Public Use File (DE-SynPUF) ("CMS 2008-2010 Data Entrepreneurs’ Synthetic Public Use File (DE-SynPUF) - Centers for Medicare & Medicaid Services", 2019) to conduct my analysis. The data source was created using a set of metadata to allow researchers do data mining as well as conduct timely and less expensive research. It has 5 types of data and can be used partially. In this study, one dataset used contains synthetic SynPuf files for 100,000 persons while another contains 23M patients.

The cohorts defined and generated are:

1) Target cohort [zjin80] HPV vaccine receivers

Patients who got HPV vaccination, have continuous observation period of 365 days prior to initial event and at least 60 days after event index date.

http://gt-health-analytics.us-east-1.elasticbeanstalk.com/#/cohortdefinition/1510

Generated Patient Counts (for 100k): 53
Generated Patient Counts (for 2.3M): 1,120

2) [zjin80] HPV vaccine non-receivers

Patients who did not get HPV vaccination, have continuous observation period of 365 days prior to initial event and at least 60 days after event index date

http://gt-health-analytics.us-east-1.elasticbeanstalk.com/#/cohortdefinition/1512

Generated Patient Counts (for 100k): 83,057
Generated Patient Counts (for 2.3M): 1,931,044

3) [zjin80] Patients with autoimmune diseases

Patients who have a diagnosis of autoimmune diseases

http://gt-health-analytics.us-east-1.elasticbeanstalk.com/#/cohortdefinition/1515

Generated Patient Counts (for 100k): 49,424

Generated Patient Counts (for 2.3M): 1,148,848

For concepts, HPV vaccines include “Gardasil 9”, “Gardasil” & “Cervarix”. When it comes to autoimmune diseases, I referenced to the common ICD-10 codes related to autoimmune ("ICD-10 Common Codes Related to Autoimmune", 2019), a few research papers on the correlation of HPV vaccines (Cervantes & Doan, 2018) and autoimmune diseases and the discovered side-effects of HPV vaccines ("Human papillomavirus vaccine Side Effects: Common, Severe, Long Term - Drugs.com", 2019). So “Type 1 diabetes mellitus”, “rheumatoid arthritis” and a few other terms that have available RCs are added. Those symptoms of autoimmune disorders like “joint pain” and “pain in limb” that may also be caused by other diseases are not included. Descendants of

“Autoimmune disease” are also referred to. For those with RC available (about 30 concepts), those with at least 1 paper or report discussing their relationship with HPV vaccines (Cervantes & Doan, 2018; Genovese et al., 2018; Liu et al., 2018) are added, resulting in 15 diseases in total.

Using Atlas supported by OHDSI, I characterized my patient cohorts using following strategy (Figure 1) and calculated the Incidence Rates. The time-at-risk is defined here to start on start day and end on 60 days after cohort end.

Figure 1. Incidence calculation components. (Informatics, 2019)

And incidence is defined as:

$$Incidence\ Rate = \frac{\#\ persons\ in\ cohort\ with\ new\ outcome\ during\ TAR}{person\ time\ at\ risk\ contributed\ by\ persons\ in\ cohort}$$

For Characterization, feature selected are Gender, Age Group, Condition Era any time prior, Drug Era any time prior.

ID	Name	Description	Actions
10	Condition Era Any Time Prior	One covariate per condition in the condition_era table overlapping with any time prior to index.	Remove
71	Demographics Age Group	Age of the subject on the index date (in 5 year age groups)	Remove
74	Demographics Gender	Gender of the subject.	Remove
89	Drug Era Any Time Prior	One covariate per drug in the drug_era table overlapping with any time prior to index.	Remove

Figure 2. Characterization design tab - feature selection.

For Stretch Assignment, OHDSI RStudio package is used to do the a retrospective, observational, comparative cohort design (Ryan et al., 2013) . As mentioned above, the patients have continuous observation in the database for at least 365 days prior to treatment initiation. Any patient who had autoimmune diseases prior to getting HPV vaccination are not be excluded because these kinds of diseases are very common. Other settings are the same as above. Similar to the paper given (Schuemie et al., 2013), I present PS (propensity scores) and covariate balance metrics to assess confounding control, provide (HR hazard ratios) estimates and Kaplan-Meier survival plots for the outcome of autoimmune disease.

Results

1) Characterization

Due to page length, only prevalence covariates for all four conditions in 100k dataset are shown here. It could be seen that covairates with biggest difference (std diff) between HPV receivers and non-receivers are all drugs. Age group and Condition Era Any Time Prior have smaller std diff. Considering gender, male has a negative std diff. Since HPV is mostly vaccinated by female, there is not many male in target cohort, so this analysis has its limitation. Details can be found: http://rstudio.gt-health-analytics.us-east-1.elasticbeanstalk.com/#/cc/characterizations/592

Covariate	Explore	Concept ID	[zjin80] HPV vaccine non-receivers		[zjin80] HPV vaccine receivers		Std diff
			Count	Pct	Count	Pct	
Simvastatin	Explore	1539403	26,920	32.41%	44	74.58%	0.4053
Omeprazole	Explore	923645	14,816	17.84%	30	50.85%	0.3958
Captopril	Explore	1340128	12,026	14.48%	25	42.37%	0.3676
Oxygen	Explore	19025274	21,287	25.63%	35	59.32%	0.3634
carvedilol	Explore	1346823	7,331	8.83%	18	30.51%	0.3435
Metformin	Explore	1503297	19,238	23.16%	31	52.54%	0.3357
Hydralazine	Explore	1373928	5,306	6.39%	15	25.42%	0.3352
Enalapril	Explore	1341927	14,897	17.94%	26	44.07%	0.3299
atorvastatin	Explore	1545958	15,932	19.18%	27	45.76%	0.3279
Imipramine	Explore	778268	4,084	4.92%	13	22.03%	0.3275

Figure 3. Characterization results showing top prevalence covariates for each cohort.

2) Incidence Rate

As shown in Figure 4, the incidence rate is expressed in the number of cases per 1000 person-years. The difference between two cohorts are not significant in both datasets.
http://rstudio.gt-health-analytics.us-east-1.elasticbeanstalk.com/#/iranalysis/677

Target cohort	Outcome cohort	Rate per 1k yrs (100k)	Rate per 1k yrs (23m)
Target: [zjin80] HPV vaccine receivers	[zjin80] Patients with autoimmune diseases	250.00	205.02
Comparator: [zjin80] HPV vaccine non-receivers	[zjin80] Patients with autoimmune diseases	246.03	245.72

Figure 4.Incidence Rate for each cohort .

3) Stretch Assignment

Population characteristics

Figure 5 (left) diagrams the inclusion of study subjects after several stages of analysis using 100k SynPUF. Figure 6 compares base-line characteristics between patient cohorts.

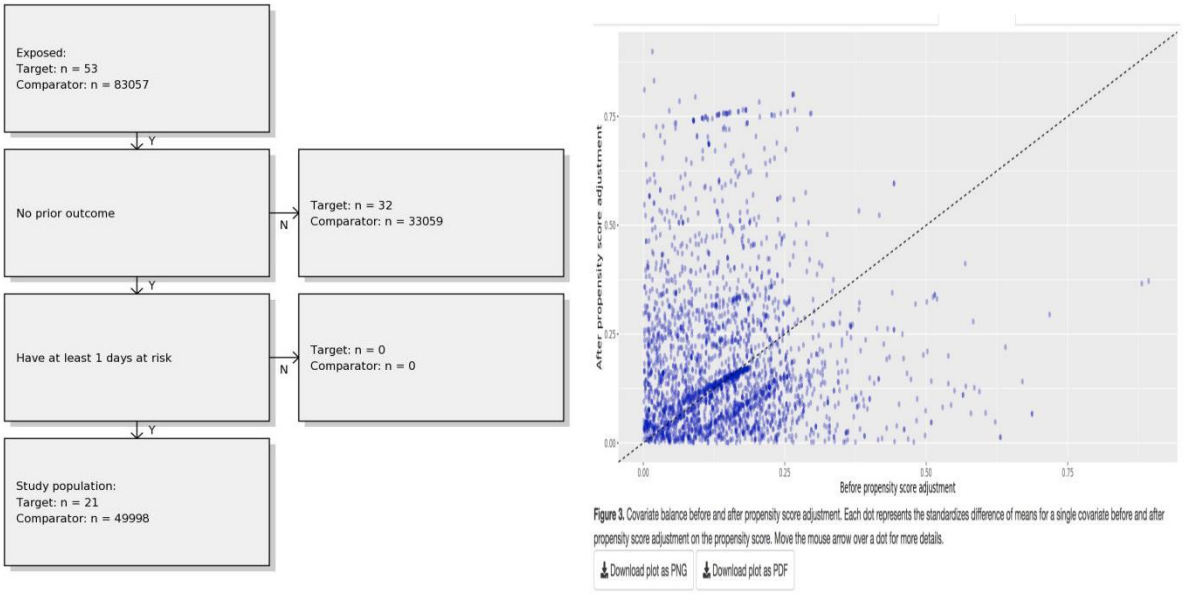

Figure 5. Left. Attrition diagram showing number of subjects in target (HPV-receivers) and comparator (HPV non-receivers); Right. Patient characteristics balance before and after stratification

Patient characteristics balance

Figure 5 (right) plots StdDiff for all patient features that serve as input for the PS model. However, after stratification the StdDiff of features is still not low enough, meaning it is not well-balanced enough. Figure 9 plots the preference score distributions, re-scalings of PS estimates to adjust for differential treatment prevalences, for patients treated with HPV vaccination and those who don't. We assess characteristics balance achieved through PS adjustment by comparing all characteristics' standardized difference (StdDiff) between treatment group means before and after PS trimming and stratification (Figure 6).

Outcome assessment

Figure 8 details the time to first autoimmune disease distributions for patients in the HPV-receiver and non HPV-receiver cohorts. Figure 10 plots Kaplan-Meier survival curves for patients under the intent-to-treat design. The target curve (HPV vaccine receivers) shows the actual observed survival. The comparator curve (non HPV-receivers) applies reweighting to approximate the counterfactual of what the target survival would look like had the amlodipine cohort been exposed to the comparator instead. The shaded area denotes the 95% CI.

Power | Attrition | Population characteristics | Propensity model | Propensity scores | Covariate balance | Systematic error

Table 2. Select characteristics before and after propensity score adjustment, showing the (weighted) percentage of subjects with the characteristics in the target (*[zjin80] H receivers*) group, as well as the standardized difference of the means.

Characteristic	Before PS adjustment		
	Target	Comparator	Ta
	%	%	Std. diff
Age group			
55-59	<9.4	3.4	0.11
60-64	<9.4	3.5	-0.10
65-69	9.4	17.7	-0.24
70-74	15.1	19.8	-0.12
75-79	22.6	16.8	0.15
85-89	9.4	9.7	-0.01
90-94	9.4	3.7	0.23
95-99	<9.4	2.6	-0.05
Gender: female	56.6	57.6	-0.02
Race			
race = Black or African American	9.4	9.9	-0.02
race = White	77.4	83.8	-0.16
Ethnicity			
ethnicity = Hispanic or Latino	<9.4	2.3	0.24
ethnicity = Not Hispanic or Latino	92.5	97.7	-0.24
Medical history: General			
Acute respiratory disease	35.8	33.4	0.05
Depressive disorder	17.0	17.1	0.00
Diabetes mellitus	64.2	62.1	0.04
Gastroesophageal reflux disease	28.3	24.6	0.08
Hyperlipidemia	37.7	43.2	-0.11
Hypertensive disorder	<9.4	8.7	-0.04
Lesion of liver	<9.4	6.2	0.05
Osteoarthritis	58.5	56.2	0.05
Renal impairment	32.1	33.9	-0.04
Schizophrenia	18.9	8.2	0.31
Urinary tract infectious disease	26.4	32.7	-0.14
Visual system disorder	56.6	54.3	0.05

Figure 6. Patient demographics (data after PS trimming and stratification not presented here due to page limit).

Characteristic	Before PS adjustment		
	Target	Comparator	Ta
	%	%	Std. diff
Medical history: Cardiovascular disease			
Atrial fibrillation	45.3	37.6	0.16
Coronary arteriosclerosis	45.3	45.4	0.00
Heart disease	69.8	69.1	0.02
Heart failure	30.2	35.5	-0.11
Ischemic heart disease	26.4	24.5	0.04
Peripheral vascular disease	34.0	42.2	-0.17
Venous thrombosis	15.1	13.0	0.06
Medical history: Neoplasms			
Hematologic neoplasm	18.9	14.1	0.13
Malignant neoplasm of anorectum	<9.4	2.6	-0.05
Malignant neoplastic disease	45.3	37.3	0.16
Malignant tumor of breast	11.3	5.5	0.21
Malignant tumor of colon	15.1	6.0	0.30
Malignant tumor of lung	<9.4	4.6	0.12
Malignant tumor of urinary bladder	9.4	4.8	0.18
Medication use			
Agents acting on the renin-angiotensin system	84.9	57.3	0.64
Antibacterials for systemic use	66.0	48.5	0.36
Antidepressants	71.7	46.5	0.53
Antiepileptics	52.8	30.7	0.46
Antiinflammatory and antirheumatic products	50.9	33.5	0.36
Antineoplastic agents	20.8	13.8	0.18
Antithrombotic agents	54.7	40.1	0.30
Beta blocking agents	83.0	52.7	0.69
Calcium channel blockers	62.3	45.0	0.35
Diuretics	86.8	56.3	0.72
Drugs for acid related disorders	60.4	37.6	0.47
Drugs for obstructive airway diseases	49.1	30.6	0.38
Drugs used in diabetes	66.0	41.8	0.50
Immunosuppressants	11.3	6.0	0.19
Lipid modifying agents	84.9	57.7	0.63
Opioids	43.4	33.9	0.20
Psycholeptics	67.9	44.1	0.50
Psychostimulants, agents used for adhd and nootropics	18.9	14.3	0.12

Figure 7. Patient demographics (continued).

Cohort	Min	P10	P25	Median	P75	P90	Max
Target	38	62	143	274	538	581	632
Comparator	1	102	256	473	642	664	737

Figure 8. Time-at-risk distributions as percentiles in the target and comparator cohorts after stratification.

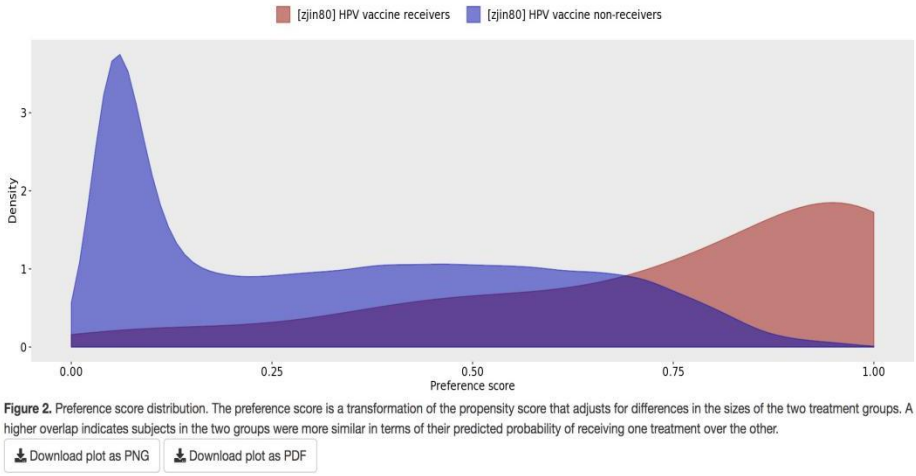

Figure 9. Preference score distribution for HPV vaccine receivers and non-receivers.

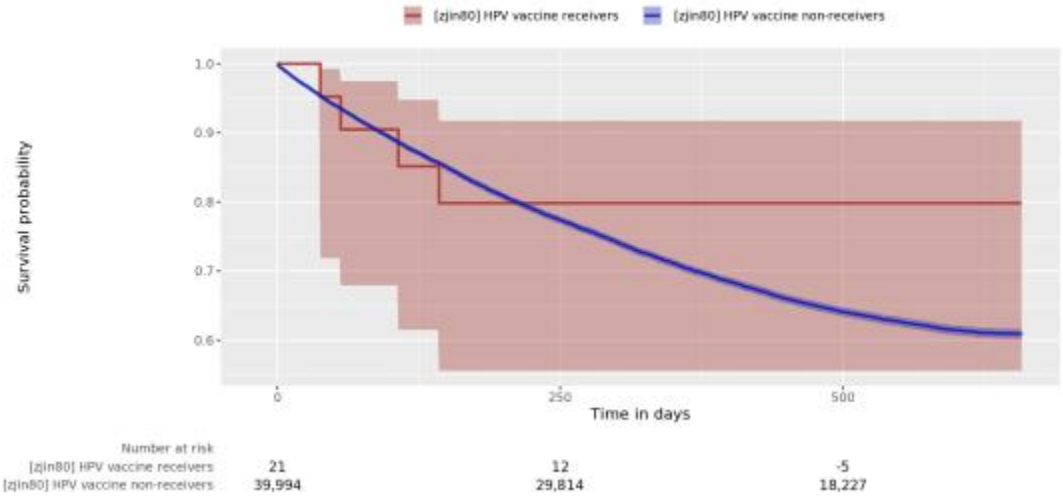

Figure 10. Kaplan Meier plot of autoimmune disease-free survival.

Discussion

We found the incidence of autoimmune diseases no greater for HPV vaccine receivers than HPV vaccine non-receivers. For HR value, it is 0.75 with lower bound 0.23 and upper bound 1.74. Since the value is smaller than 1, HPV vaccination is not likely to increase the risk autoimmune diseases. However, its upper bound is too high, so it is not 100% certain.

I think it is what I expected as there are other research reaching similar conclusion.

But this research has its limitation. First, in SynPUF dataset, there are only older patients, while HPV vaccine are usually received in a younger age. So there are not enough target patients available. And as an old dataset, it has very few Gardasil 9 receivers, as it was only popularized in recent years. Besides, with only limited year to observe patients, the longer-term hidden risk could not be discovered. Last but not least, the incidence rate does not control for differences in populations so I cannot make an assertion of statistical significance

For follow up, a younger and up-to-date dataset can be found to better address the question.

Conclusion

In summary, my study provides reassuring results regarding the risk of autoimmune diseases after HPV vaccination, since there is no apparent increased risk of these diseases detected. Further studies are required to confirm this finding.

Reference

1. Autoimmune Disease List – AARDA. (2019). Retrieved from https://www.aarda.org/diseaselist/
2. Cervantes, J., & Doan, A. (2018). Discrepancies in the evaluation of the safety of the human papillomavirus vaccine. *Memórias Do Instituto Oswaldo Cruz, 113*(8). doi: 10.1590/0074-02760180063
3. CMS 2008-2010 Data Entrepreneurs' Synthetic Public Use File (DE-SynPUF) - Centers for Medicare & Medicaid Services. (2019). Retrieved from https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/SynPUFs/DE_Syn_PUF.html
4. Genovese, C., LA Fauci, V., Squeri, A., Trimarchi, G., & Squeri, R. (2018). HPV vaccine and autoimmune diseases: systematic review and meta-analysis of the literature. *Journal of preventive medicine and hygiene, 59*(3), E194–E199. doi:10.15167/2421-4248/jpmh2018.59.3.998
5. Human papillomavirus vaccine Side Effects: Common, Severe, Long Term - Drugs.com. (2019). Retrieved from https://www.drugs.com/sfx/human-papillomavirus-vaccine-side-effects.html
6. ICD-10 Common Codes Related to Autoimmune. (2019). Retrieved from https://www.questdiagnostics.com/dms/Documents/Other/CPT-2015/ICD_10_Codes_Autoimmune-MI4953.pdf
7. Liu, E., Smith, L., Ellis, A., Whitaker, H., Law, B., & Kwong, J. et al. (2018). Quadrivalent human papillomavirus vaccination in girls and the risk of autoimmune disorders: the Ontario Grade 8 HPV

Vaccine Cohort Study. *Canadian Medical Association Journal*, *190*(21), E648-E655. doi: 10.1503/cmaj.170871

8. NIH Fact Sheets - Cervical Cancer. (2019). Retrieved from https://report.nih.gov/nihfactsheets/viewfactsheet.aspx?csid=76