



A Model for COVID-19 Vaccine Adverse Reaction

Suling Lee, Singapore Management University

Motivation

- The COVID-19 vaccines are crucial to ending the global pandemic that is causing surges of infections and deaths globally.
- However, the unprecedented rate at which it was developed and administered had raised doubts in the community regarding its safety.
- Data from the United States Vaccine Adverse Event Reporting System, VAERS, has the potential to help determine if the safety concerns of the vaccines are founded.
- This paper uses the combination of both structured and unstructured text data from VAERS to model the adverse reactions to COVID-19 vaccines.



Vaccine Adverse Event Reporting System
www.vaers.hhs.gov

Vaccine Adverse Event Reporting System (VAERS)

- A passive surveillance system for adverse events reporting
- Doctors, vaccine manufacturers, patients, caregivers etc use this platform
- Used by CDC and FDA to guide vaccine recommendations and regulatory action
- The link between adverse event and vaccination is not established

Why use VAERS data?

- Data is available to the public
- Data is up to date
- Not all adverse events are likely to be captured during clinical trials
- Key patterns in reporting trends

For more information:

[Safety monitoring in the Vaccine Adverse Event Reporting System \(VAERS\)](#)

PMC 2015 November 04.

Tom T. Shimabukuro, Michael Nguyen, David Martin and Frank DeStefano

Datasets

VAERSDATA

- 35 variables
- Patient's profile (age, state, sex, etc)
- Text narratives of Symptoms, Medical History, Allergies, Medications
- Patient's outcomes for the adverse event (died, extended stay in hospital, life threatening illness, disabled as a result of vaccination, etc)

For more information:
[VAERS data use guide](#)

SYMPTOM_TEXT	V_ADMINBY	VAERS_ID	VAERS_ID 2	VAX_DOSE_SERIE S	VAX_MANU	X_STAY	acute respiratory failure Binary	bell's palsy Binary	cardiac arrest Binary	cerebrovascular acci
27,647 unique values	PHM	29,475 unique values	29,475 unique values	1	PFIZER\BIONTECH	0	1	1	1	
	PVT			2	MODERNA	1				
	UNK				JANSSEN					
	OTH									
PUB										
4 others										
Pt contracted Co...	SEN	1414645	1414645	1	PFIZER\BIONTECH	0	0	0	0	
Pt contracted Co...	SEN	1414645	1414645	2	PFIZER\BIONTECH	0	0	0	0	
Pt received his 2n...	UNK	1413776	1413776	2	PFIZER\BIONTECH	0	0	0	0	

Stroke; sitting on the floor totally disoriented; Cardiac arrest; This is a spontaneous report from a contactable consumer (patient's Daughter). A 101-year-old female patient received first dose BNT162B2 (PFIZER-BIONTECH COVID-19 VACCINE, Batch/Lot number was not reported), via an unspecified route of administration at the age of 101-year-old on 03Mar2021 at single dose for covid-19 immunisation. Medical history included glaucoma, blood pressure high, Mini stroke, her blindness was an infarct to her optic nerve, diagnosed allergies, compromised immune status, respiratory illness, genetic/chromosomal abnormalities, endocrine abnormalities, and obesity. Family history included Patient's sister got the vaccination and either 3 or 5 days later she passed away. There was none history of all previous immunization with the Pfizer vaccine considered as suspect. Concomitant medication included metoprolol taken for high blood pressure; levothyroxine; amlodipine taken for high blood pressure; sertraline to keep her spirits up because she is old and blind; latanoprost (XALATAN); dorzolamide hydrochloride, timolol maleate (COSOPT); acetylsalicylic acid (BABY ASPIRIN); calcium (CALCIUM); vitamin C [ASCORBIC ACID]; tocopherol (VITAMIN E). Prior Vaccinations (within 4 weeks), there was no any other vaccinations within four weeks prior to the first administration date of the suspect vaccine(s). The patient received the Flu shot every year and patient also received the shingles shot and the pneumonia shot. Reporter stated patient had lunch and was sitting in the chair so it was probably somewhere between 12 and 3pm on 09Mar2021 when she had her stroke. Reporter states when she got back home patient was on the floor totally disoriented. Above events results in Emergency Room and patient was hospitalized from 10Mar2021 to 12 Mar2021. Reporter stated that patient died in her sleep so early in the morning probably between 3am and 6am on 06Jun2021. Reported cause of death was cardiac arrest. Reporter was calling about the Pfizer COVID vaccination. Reporter was calling on behalf of her mother. Reporter believed that her mother had a reaction to the first shot. Reporter stated that her mother had a stroke 6 days after receiving the first Pfizer COVID vaccination and has since passed away. Reporter thought this was a contributing factor. Reporter states that her mother was 101 years old but was an incredibly lucid and healthy individual. Reporter states that her mother used to be 5 foot 1 inch but she was probably 4 foot 11 inches at the time of this event (09Mar2021). Reporter states at the time of this event her mother probably weighed about 133 pounds. Reporter states that she came home on 09Mar2021 and found her mother on the floor. She took her mom to the hospital on 10Mar2021 and she was diagnosed with a stroke. Her mother was hospitalized 10Mar2021 through 12Mar2021. Reporter stated that her mother went to rehab after being discharged from the hospital. Stated that her mother never came back home. Reporter states that her mother lived with her. Reporter states she looked at her labs on 01Jun2021, her labs were always good. She stated that her mother also had a cat scan with contrast in 2021, no further details provided. Reporter states that she was not saying that this product caused her death but the timing was very suspicious. The patient underwent lab tests included Fasting blood glucose: 99 on 01Jun2021, cat scan with contrast with unknown results in 2021. The outcome of events was fatal. The patient died on 06Jun2021. An autopsy was not performed. Information on the lot/batch number has been requested.; Reported Cause(s) of Death: Stroke; sitting on the floor totally disoriented; Cardiac arrest

Sample of symptoms text

The data is quite dirty, a lot of useful information but in narrative text, expect spelling error, typos, excessively long or brief statements...

Datasets

VAERSVAX

- 8 variables
- Vaccine information (type, manufacturer, vaccine lot, etc)
- VAERS_ID (unique identification number of patient)

VAERS_ID	VAX_MANU	VAX_DOSE_SERIES	VAX_ROUTE	VAX_SITE
1.55M	PFIZER\BIONTECH	1	IM	LA
	MODERNA	2	SYR	RA
	JANSSEN		OT	AR
			UN	UN
917k				OT
				5 others
916600	MODERNA	1	IM	LA
916601	MODERNA	1	IM	RA
916602	PFIZER\BIONTECH	1	IM	LA
916604	MODERNA	1	IM	LA
916606	MODERNA	1	IM	LA

Sample of VAERSVAX dataset

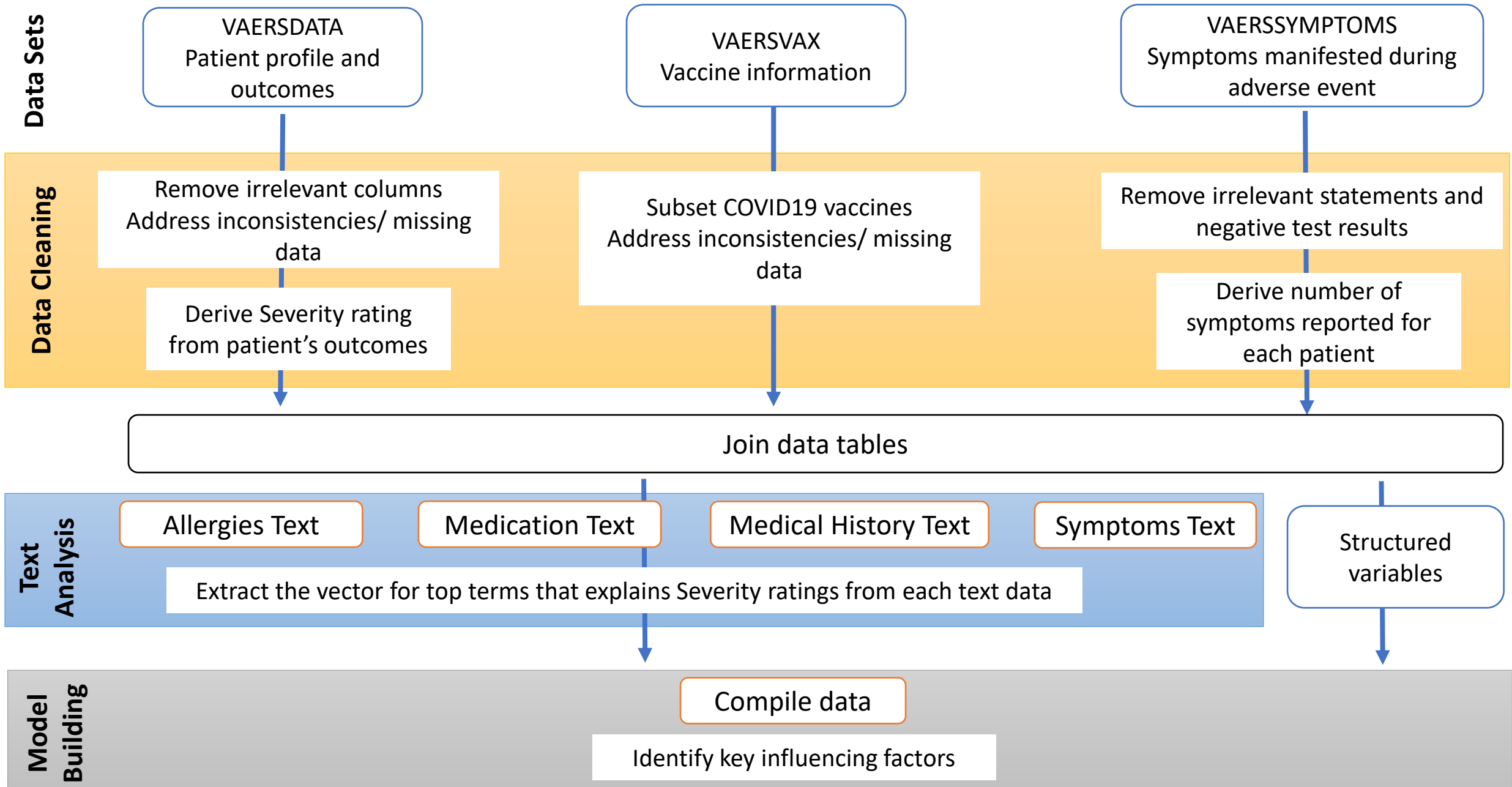
VAERSSYMPTOMS

- 11 variables
- Terms describing the symptoms of the adverse event
- There could be multiple rows for each VAERS_ID

VAERS_ID	SYMPTOM1	SYMPTOM VERSION1	SYMPTOM2	SYMPTOM VERSION2	
1.55M	Chills	24	Headache	24	Headach
	Arthralgia		Fatigue		Fatigue
	Dizziness		Chills		Pain
	Fatigue		Dizziness		Pyrexia
	Headache		Pain		Nausea
917k	6,312 others	23.1	6,025 others	23.1	5,591 ot
916603	Dizziness	23.1	Fatigue	23.1	Mobili
916604	Injection site erythema	23.1	Injection site pruritus	23.1	Injetic
916605	Chills	24	Confusional state	24	Eye infl
916605	Pyrexia	24	White blood cell count decreased	24	
916606	Pharyngeal swelling	23.1			•
916607	Abdominal pain	23.1	Chills	23.1	Sleep c
916608	Diarrhoea	23.1	Nasal congestion	23.1	

Sample of VAERSSYMPTOMS dataset

Overview of Processes



Severity Rating

Patient Outcomes

- Based on reporter's assessment of vaccine recipient
- 12 variables, mostly binary (Y), describes health status of patient and hospitalisation

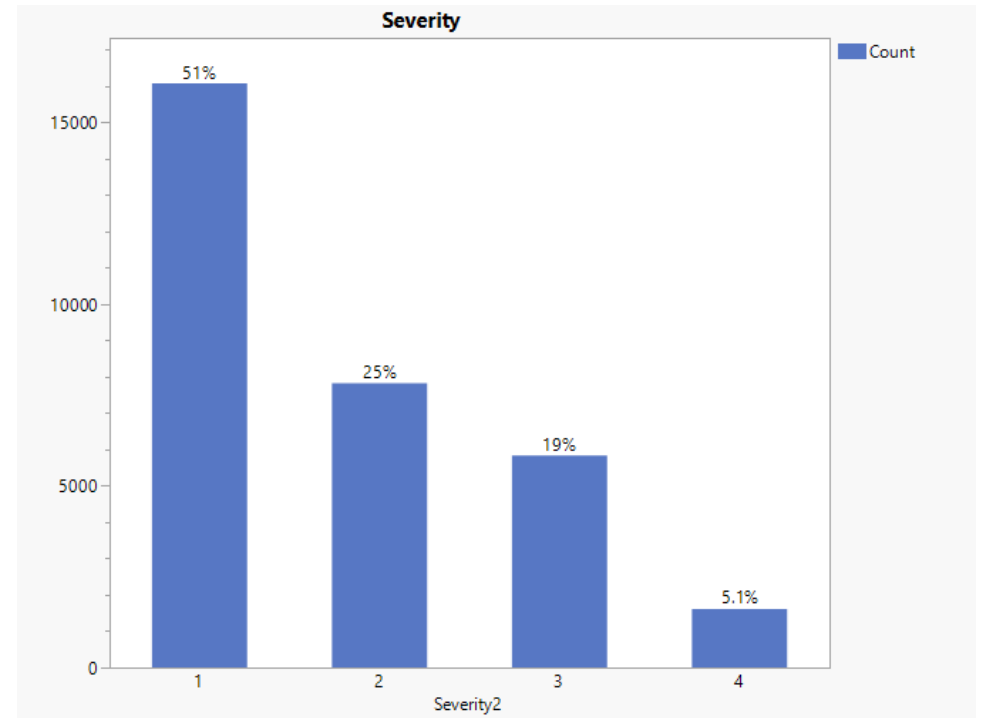
Severity ratings were based on Patient Outcomes

Level 4 = DIED or DISABLE or BIRTH_DEFECT = 'Y'

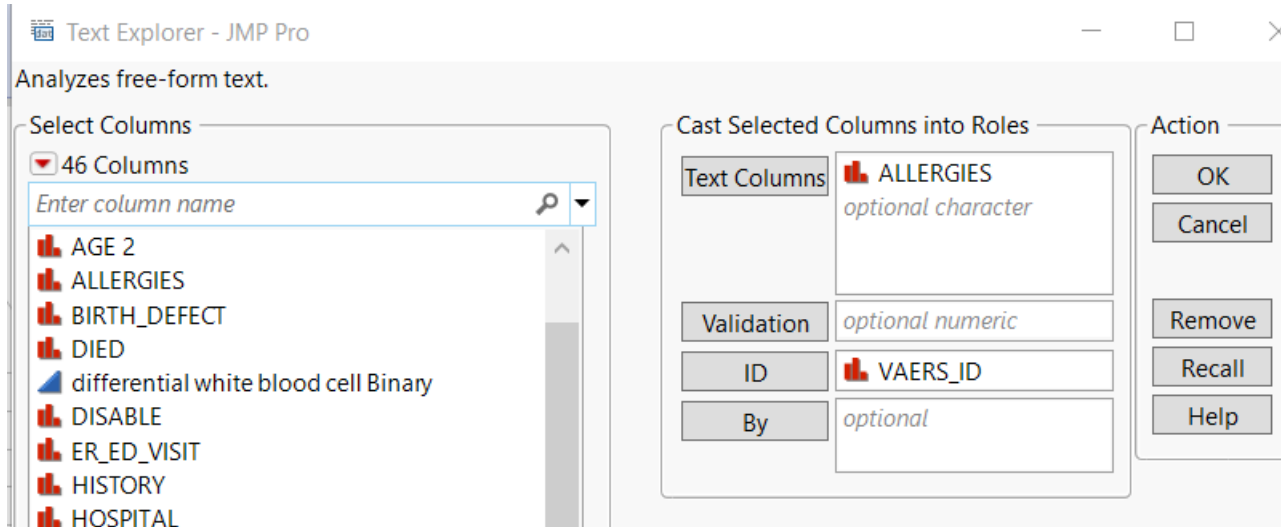
Level 3 = non-level 4 and (ER_ED_VISIT or L_THREAT or X_STAY = 'Y')

Level 2 = non-level 3 and non-level 4 and (HOSPITAL or OFC_VISIT = 'Y')

Level 1 = non-level 2, 3 and 4



Text Explorer – data cleaning

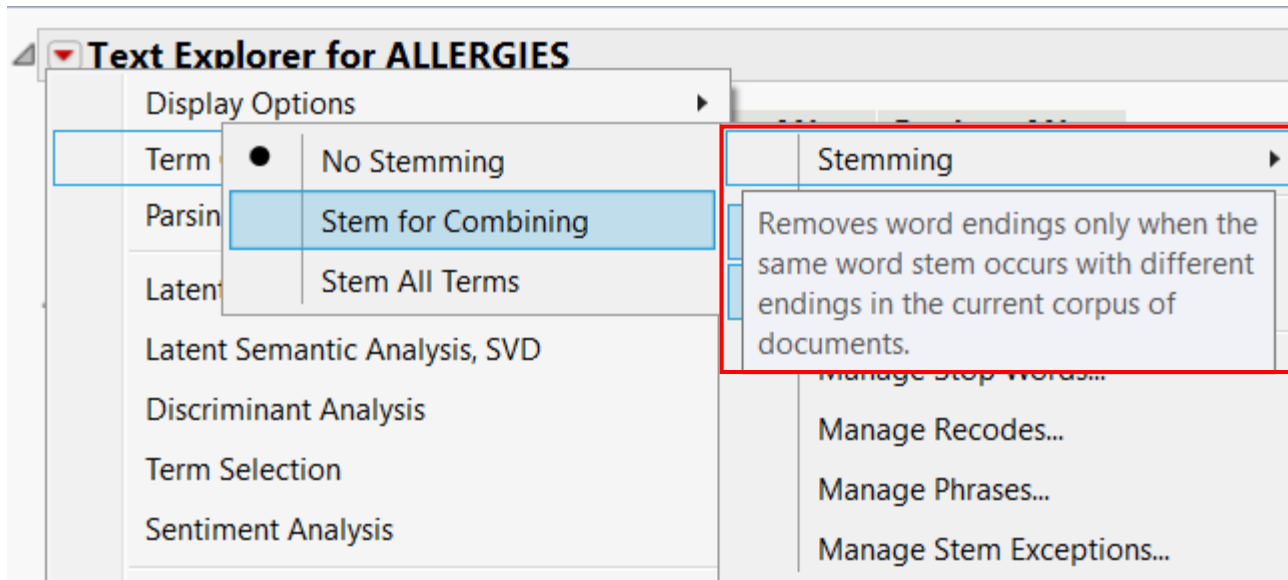


Data contains many spelling errors and typos

Term and Phrase Lists

Term	Count
ant	4
anthrax	1
anti	21
antianflamitoris	1
antibiotic	1
antibiotics	1
antibiots	1
antibody	1
antibiotic	1
antibiotics	1
anticholinergics	1
antidepressants	3
antifungal	3
antifungals	1
antihistamine	5
antihistamines	6
antibiotics	1
antiperspirant	1
antiphallic	1
antiphylaxis	2
antitriplene	1
ants	5
	13
	2
	25

Sort list
alphabetically and
recode terms



Stem terms based on
word endings only

Text Explorer – data cleaning

Term and Phrase Lists

Term	Count	Phrase	Count	N
bedtime	2	sulfa drugs	291	2
bee-	80	known drug	87	2
bee- sting-	94	bee stings	78	2
bee- venom	34	tree nuts	77	2
beeanaphylaxis	1	sulfa antibiotics	64	2
beef	16	contract dye	50	2

Recode - Term

Count	Old Values (6)	New Values (6)
15	bee	bee
16	bee sting	bee sting
78	bee stings	bee stings
34	bee venom	bee venom
1	beeanaphylaxis	beeanaphylaxis
65	bees	bees

Filter:

Group controls

- View Groups
- Show Only Grouped
- Show Only Ungrouped

Group

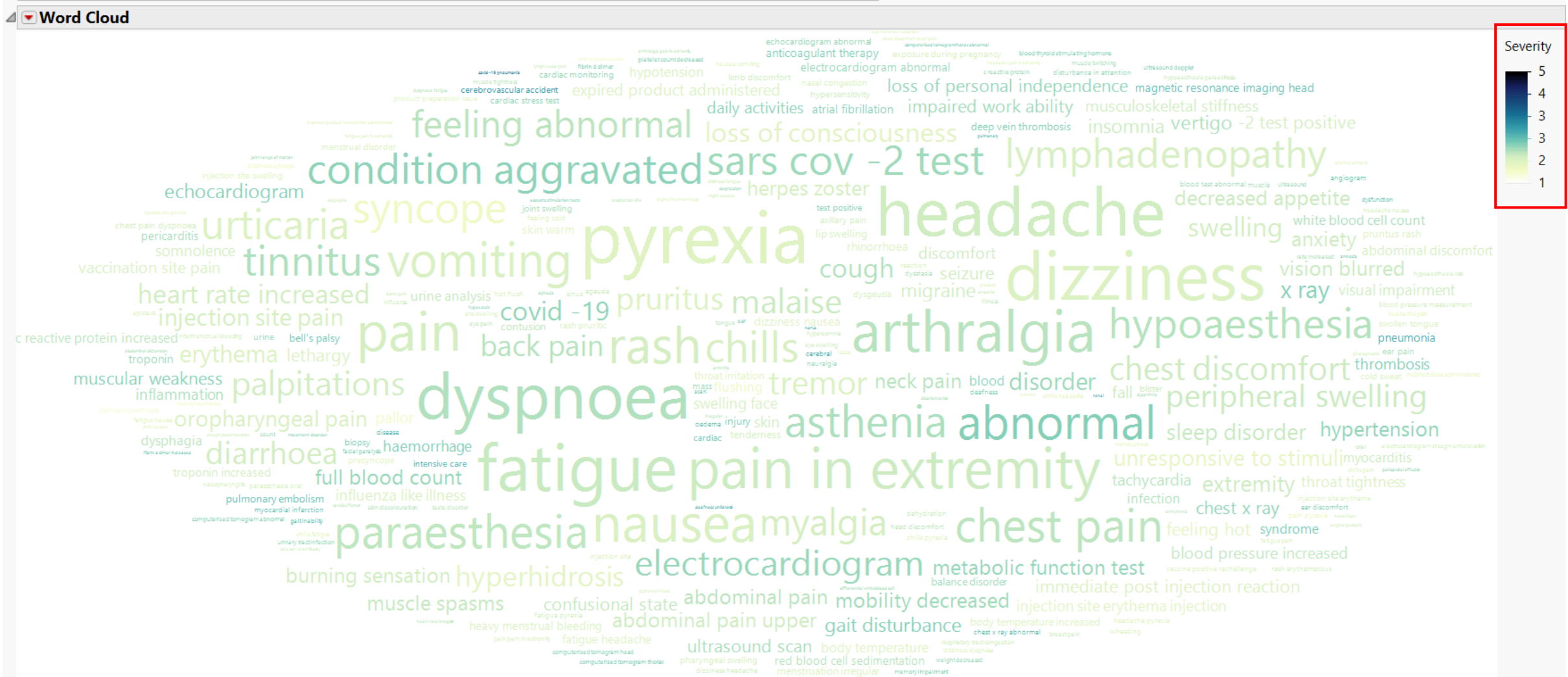
All
 Only Modified
 Only Unmodified

Changes

Add phrases and
recode terms with
similar meaning

Text Explorer – word cloud

Useful for stop word exclusion and seeing the effect of target variable on terms.



Text Explorer – Term Selection

Text Explorer for ALLERGIES

- Display Options
- Term Options
- Parsing Options
- Latent Class Analysis
- Latent Semantic Analysis, SVD
- Discriminant Analysis
- Term Selection**
- Sentiment Analysis
- Save Document Term Matrix

Portion of Non-Empty Cases: 9378

Portion of Non-Empty Cases: 0.2992

Phrase: sulfa drugs

Determines which terms explain different responses. Also useful for sentiment analysis when the responses are ratings or emotions.

Why use Generalised Regression model:

- Vectors in DTM are expected to be highly correlated and non-normally distributed
- Each row is a patient, all rows are independent of one another
- Variable selection

Term Selection

Settings

Choose Response Column

45 Columns

Enter column name

- PRIOR_VAX
- RECOVD
- RECVDATE
- Regions
- Severity**
- SEX

Target Levels

Local Data Filter

Clear

Inverse

Severity (5)	
0	10429
1	15205
2	4144
3	955
4	607

Model Settings

Run

Allow Early Stopping

Estimation Method: **Elastic Net**

Validation Method: AICc

Term Settings

Weighting: **Binary**

Maximum Term Count: 400

Lasso tend to select one term from a group of correlated factors while Elastic Net will select the group of terms

Binary weights for DTM made more sense than TF-IDF

Text Explorer – Term Selection

Top symptoms factors found by term selection

1 > Generalized Regression for Severity (Weighting=Binary, Model=Elastic Net/AICc) X

Model	DTM Weighting	Response	Response Distribution	Estimation Method	Validation Method	Relevant Terms	AICc	BIC	Generalized RSquare	RASE
1	Binary	Severity	Normal	Elastic Net	AICc	72	79894	80512	0.012	0.864

Term	Coefficient	LogWorth	Count
cardiac arrest	1.673	18.263	70
covid -19 pneumonia	0.873	6.627	98
cerebrovascular accident	0.795	16.322	168
deafness unilateral	0.672	7.539	92
lumbar puncture	0.616	7.838	87
acute respiratory failure	0.584	2.682	69
bell's palsy	0.570	22.681	189
pneumonia	0.481	7.009	226
myocardial infarction	0.401	2.829	138
intensive care	0.399	4.197	185
nerve	0.391	4.095	99

Term	Coefficient	LogWorth	Count
dizziness hyperhidrosis	-0.365	11.418	130
chills fatigue headache	-0.301	7.669	91
syncope	-0.294	19.158	780
dizziness syncope	-0.291	4.337	109
virus test	-0.279	1.573	77
dizziness fall	-0.270	3.119	81
headache pain	-0.262	5.857	130
pallor	-0.260	7.781	271
appendicitis	-0.240	1.549	78
pain pyrexia	-0.221	4.648	142
underdose	-0.215	3.858	85

Text Explorer – Save term vectors to table

Text Explorer for SYMPTOM

- Display Options
- Term Options
- Parsing Options
- Latent Class Analysis
- Latent Semantic Analysis, SVD
- Discriminant Analysis
- Term Selection
- Sentiment Analysis
- Save Document Term Matrix**
- Save Stacked DTM for Association
- Save DTM Formula
- Save Term Table
- Score Terms by Column
- Local Data Filter
- Redo
- Save Script

Number of Non-Empty Cases: 28782

Portion of Non-Empty Cases: 0.9184

Phrase

tomogram magn
upper respirator
abnormal paraes
aggravated pain

disorder nausea
dizziness unresp
erythema skin
general physical
headache tinnit
incomplete cour
lactate dehydrog
nausea dyspnoe
nucleic acid

Saves a column to the data table for each term, indicating or scoring the document by each term.

Specifications

Specifications for Terms and Weights

Maximum Number of Terms: 500

Minimum Term Frequency: 69

Weighting: Binary

OK Cancel Help

pyrexia Binary	headache Binary	dizziness Binary	fatigue Binary	dyspnoea Binary	arthralgia Binary	pain in extremity Binary
1	1	1	1	1	1	1
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0

Save the DTM to the data table and extract the required columns of top terms

Modelling – Validation column

The screenshot shows the JMP Pro interface with the 'Analyze' menu open. The 'Make Validation Column' option is highlighted with a red box. The background shows a data table with columns 'Binary', 'aortic Binary', and 'care'.

Binary	aortic Binary	care
1	1	

The screenshot shows the 'Make Validation Column' dialog box. The 'Specify rates or relative rates' section is highlighted with a red box. The 'Options' section shows the 'New Column Name' set to 'Validation' and the 'Validation Column Type' set to 'Fixed'.

Stratification Columns: Severity

	Adjusted Rates	Row Counts
Training Set	0.75	23505
Validation Set	0.25	7835
Test Set	0	0
Excluded Rows		0
Total Rows		31340

Options

New Column Name: Validation
Validation Column Type: Fixed
Random Seed: .

Buttons: Go, Cancel, Help

Validation
Training
Validation
Training
Training
Training
Training
Training
Training
Validation
Validation
Training
Training
Training
Training
Validation
Training

Modelling – Fit Model

Fit Model - JMP Pro

Model Specification

Select Columns: 87 Columns

Pick Role Variables: Y: Severity

Personality: **Generalized Regression**

Distribution: Ordinal Logistic

Buttons: Help, Run, Recall, Remove, Keep dialog open

Construct Model Effects:

- Add: AGE 2, differential white blood cell Binary, NUMDAYS 2
- Cross: Regions, SEX
- Nest: V_ADMINBY
- Macros: VAX_DOSE_SERIES, VAX_MANU
- Degree: 2
- Attributes: acute respiratory failure Binary
- Transform: bell's palsy Binary

JMP Fit Model
Generalised regression model
with Lasso estimates has the
smallest AICc and largest
Rsquare values

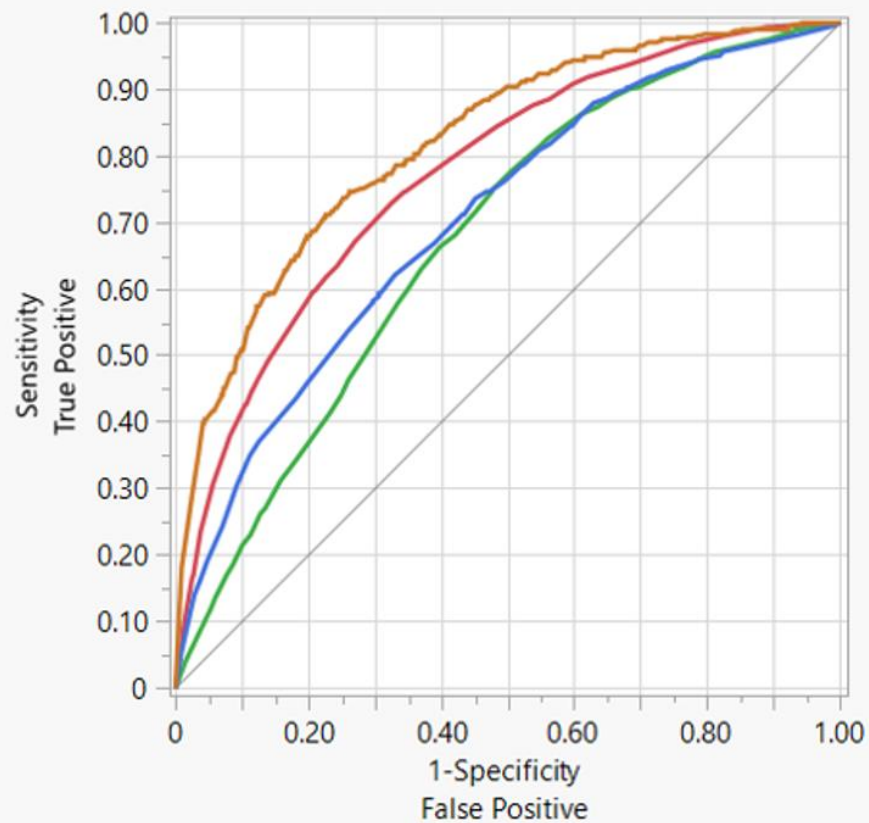
Generalized Regression for Severity2

Model Comparison

Show	Response Distribution	Estimation Method	Validation Method	Nonzero Parameters	AICc	BIC	Generalized RSquare	Validation Generalized RSquare
<input checked="" type="checkbox"/>	Multinomial	Lasso	Validation Column	251	46311.491	48330.358	0.3346117	0.3024193
<input checked="" type="checkbox"/>	Multinomial	Double Lasso	Validation Column	250	46303.369	48314.214	0.3348124	0.302395
<input checked="" type="checkbox"/>	Multinomial	Elastic Net	Validation Column	258	46312.591	48387.604	0.3350477	0.302048
<input checked="" type="checkbox"/>	Multinomial	Ridge	Validation Column	267	46371.141	48518.328	0.3337222	0.3005086
<input checked="" type="checkbox"/>	Multinomial	Maximum Likelihood	Validation Column	267	46299.815	48447.003	0.3360759	0.2981779

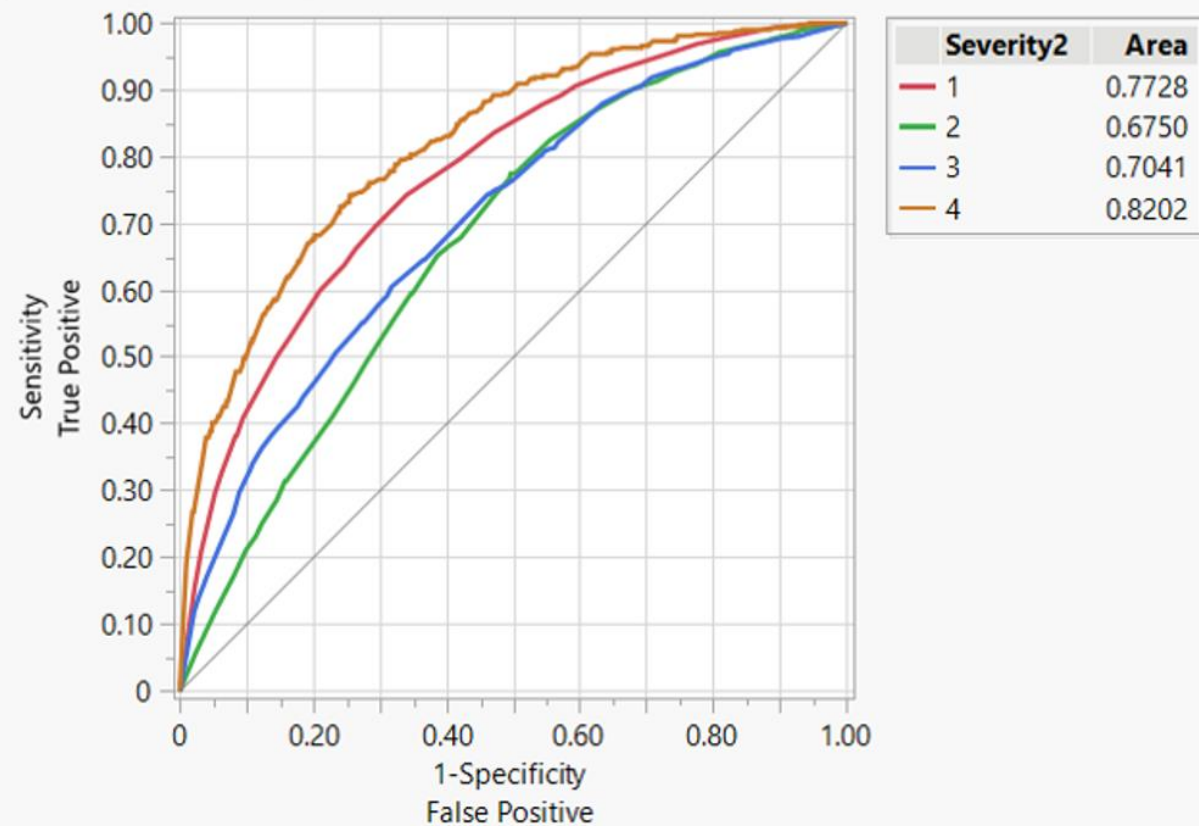
Maximum Likelihood model

Receiver Operating Characteristic on Validation Data



Lasso model

Receiver Operating Characteristic on Validation Data



Effect Tests

Source	Nparm	DF	Wald ChiSquare	Prob > ChiSquare ^	
NSYMP 2	12	11	1607.2519	<.0001*	Levels removed: 1
NUMDAYS 2	18	16	1714.6293	<.0001*	Levels removed: 2
intensive care Binary	3	3	950.82905	<.0001*	
V_ADMINBY	24	21	677.50105	<.0001*	Levels removed: 3
AGE 2	12	12	292.45944	<.0001*	
cerebrovascular accident Binary	3	3	110.64681	<.0001*	
pulmonary embolism Binary	3	3	90.417622	<.0001*	
cardiac arrest Binary	3	3	84.420657	<.0001*	
deafness unilateral Binary	3	3	76.936318	<.0001*	
myocardial infarction Binary	3	3	67.479073	<.0001*	
SEX	3	3	63.714765	<.0001*	
bell's palsy Binary	3	3	53.791218	<.0001*	
VAX_MANU	6	5	51.628245	<.0001*	Levels removed: 1
clavulan· Binary	3	2	37.66094	<.0001*	Levels removed: 1
lumbar puncture Binary	3	3	37.725313	<.0001*	
nerve Binary	3	3	27.183425	<.0001*	
chronic· kidney· diseas· Binary	3	3	25.917356	<.0001*	
Regions	21	20	56.461763	<.0001*	Levels removed: 1
pneumonia Binary	3	3	23.342908	<.0001*	
gait inability Binary	3	3	22.842835	<.0001*	
penicillin· Binary	3	3	20.265984	0.0001*	
cardiovascular diseas· Binary	3	3	17.67727	0.0005*	
lung· diseas· Binary	3	3	16.151242	0.0011*	
covid -19 pneumonia Binary	3	3	16.038722	0.0011*	
pennicillin· Binary	3	3	15.412066	0.0015*	
thrombocytopenia Binary	3	3	15.154965	0.0017*	
emphysema Binary	3	2	10.803767	0.0045*	Levels removed: 1
failure Binary	3	3	10.11761	0.0176*	
crestor Binary	3	3	9.3365888	0.0251*	
miralax Binary	3	3	8.4062086	0.0383*	

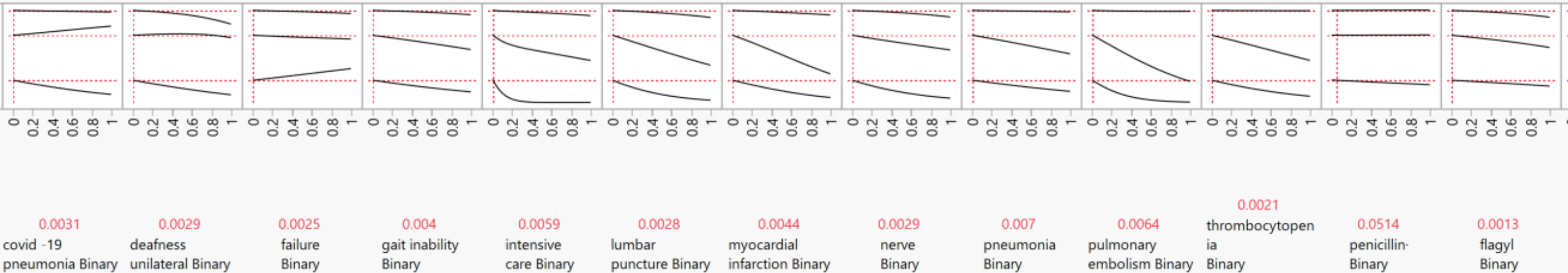
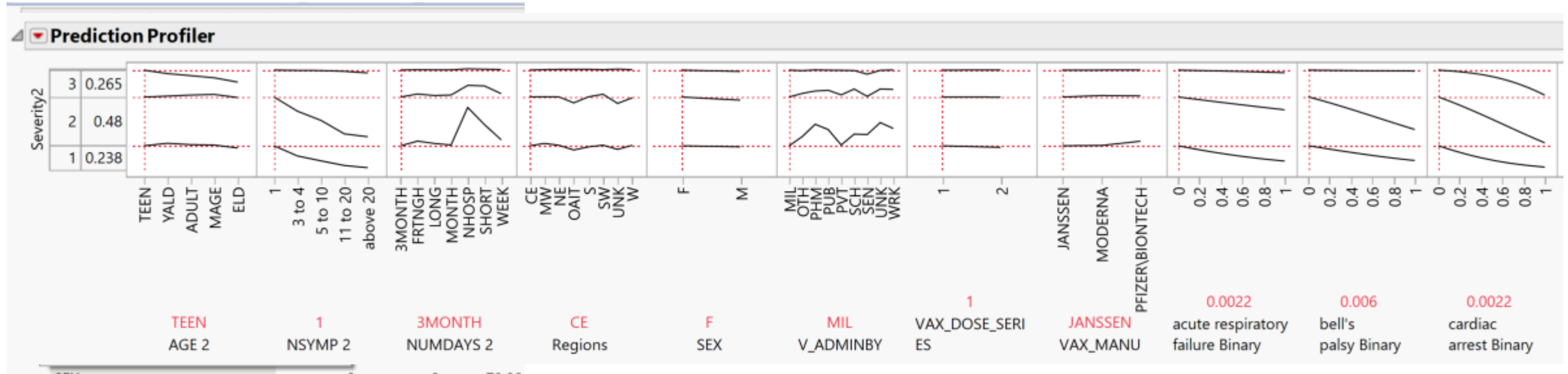
Effects Test

- hypothesis tests of the null hypothesis that the variable has no effect on the response.
- Type III sum of squares for ANOVA
- Suitable for model reduction

Results

- Long list of significant variables
- Most variables are symptoms related to cardiopulmonary illnesses.

Model Evaluation - Profiler



Conclusion

- Several decisions were made in the grouping and classification of variables. Although these decision were made to the best of our understanding of the variables and to ensure a suitable sample size, an expert familiar with vaccine studies or clinical trials could be consulted on these decisions.
- Based on model building of structured and unstructured data based, we have identified key factors that varies with the severity a reaction to the COVID_19 vaccination. However, we note that the effect of these key factors on the response variable “Severity” is very small.
- The binary DTM was found to be most effective in representing the weights of terms in each document and the Generalised linear model with Lasso penalised regression technique produced the optimal model.