

Symptom Extraction and Linking from Vaccine Adverse Event Reports

CS 584 - A

Student information:

Full name: Thanapoom Phatthanaphan

CWID: 20011296

Background and the Goal

Sequence labeling has been one of the most well-known topics in linguistics and computational linguistics history. Some examples of sequence labeling include part of speech (POS) tagging, named entity recognition (NER), and event detection (ED). In this project, I will apply sequence labeling techniques to a healthcare data, namely VAERS data, from Vaccine Adverse Events Reporting Systems (VAERS). The goal of the project is to automatically identify symptoms after vaccination (i.e., vaccine adverse events) from VAERS reports.

Goal

1. Use existing or develop new named entity recognition packages to identify symptom-related terms from narrative text reports.
2. Develop named entity linking methods to link the identified terms to standard terms in a dictionary.

Dataset

The Vaccine Adverse Event Reporting System (VAERS) is a national early warning system to detect possible safety problems in U.S.-licensed vaccines. VAERS is co-managed by the Centers for Disease Control and Prevention (CDC) and the U.S. Food and Drug Administration (FDA). VAERS accepts and analyzes reports of adverse events (possible side effects) after a person has received a vaccination. There are three tables in the VAERS dataset that will be used in this project, including VAERS Data, VAERS Symptoms, and VAERS Vaccine.

- **VAERS Data:** The dataset contains detailed information about individual VAER submitted to the VAERS system.
- **VAERS Symptoms:** The dataset contains a list of symptoms and their corresponding codes that are reported in adverse event narratives.
- **VAERS Vaccine:** The dataset provides information about the vaccines including their characteristics and administration.

The datasets spanning from 2010 to the present day will serve as the foundation for creating new named entity recognition packages to identify terms related to symptoms and developing named entity linking methods to connect these identified terms to standardized terms present in a dictionary.

Experiments

The evaluation of the machine learning model's performance is a critical aspect of this project. However, it's essential to note that there is no ground truth annotation available for the data, making evaluation challenging. To overcome this limitation, a combination of automatic and manual evaluations will be employed.

- **Automatic Evaluation:** The model's accuracy in correctly classifying symptoms will be assessed through automatic evaluation metrics.
- **Manual Evaluation:** To ensure the accuracy of the model's predictions, a sample of clinical notes (typically 20~50) will be selected for manual evaluation. Experts manually review the results, verifying the correctness of the model's predictions.

The three key metrics, including precision, recall, and F1-score, will serve as essential tools for assessing the model's performance in identifying and linking symptoms.

- **Precision:** Precision will measure the accuracy of identified symptoms, ensuring their trustworthiness and relevance
- **Recall:** Recall will determine the model's ability to capture all relevant symptoms, minimizing the risk of missing significant health indicators
- **F1-score:** F1-score will provide an overall evaluation of the model's effectiveness in both accuracy and completeness

```
In [1]: from google.colab import drive
drive.mount('/content/drive')
```

Mounted at /content/drive

1. Import libraries

```
In [2]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import glob
```

2. Data preprocessing

Importing and Combining the datasets

Import three datasets, including VAERS Data, VAERS Symptoms, and VAERS Vaccine. Since each year of data is separated into its individual file, thus I will combine them into one file that contains data from 2010 to the present.

```
In [3]: # Import datasets
# Combine separated datasets each year into one dataset
# VAERS Data
vaers_data_path = '/content/drive/My Drive/Colab Notebooks/Vaers_d
vaers_data_files = glob.glob(vaers_data_path)

vaers_data = pd.DataFrame()

for file in vaers_data_files:
    try:
        data = pd.read_csv(file)
    except UnicodeDecodeError:
        data = pd.read_csv(file, encoding='latin1')
    vaers_data = pd.concat([vaers_data, data], ignore_index=True)

# VAERS Symptoms
vaers_symp_path = '/content/drive/My Drive/Colab Notebooks/Vaers_s
vaers_symp_files = glob.glob(vaers_symp_path)

vaers_symp = pd.DataFrame()

for file in vaers_symp_files:
    try:
        data = pd.read_csv(file)
    except UnicodeDecodeError:
        data = pd.read_csv(file, encoding='latin1')
    vaers_symp = pd.concat([vaers_symp, data], ignore_index=True)

# VAERS Vaccine
vaers_vax_path = '/content/drive/My Drive/Colab Notebooks/Vaers_va
vaers_vax_files = glob.glob(vaers_vax_path)

vaers_vax = pd.DataFrame()

for file in vaers_vax_files:
    try:
        data = pd.read_csv(file)
    except UnicodeDecodeError:
        data = pd.read_csv(file, encoding='latin1')
    vaers_vax = pd.concat([vaers_vax, data], ignore_index=True)
```

```
<ipython-input-3-682894d3668c>:13: DtypeWarning: Columns (30) hav
e mixed types. Specify dtype option on import or set low_memory=F
alse.
    data = pd.read_csv(file, encoding='latin1')
<ipython-input-3-682894d3668c>:13: DtypeWarning: Columns (30) hav
e mixed types. Specify dtype option on import or set low_memory=F
alse.
    data = pd.read_csv(file, encoding='latin1')
<ipython-input-3-682894d3668c>:13: DtypeWarning: Columns (30) hav
e mixed types. Specify dtype option on import or set low_memory=F
alse.
    data = pd.read_csv(file, encoding='latin1')
<ipython-input-3-682894d3668c>:13: DtypeWarning: Columns (31) hav
e mixed types. Specify dtype option on import or set low_memory=F
alse.
    data = pd.read_csv(file, encoding='latin1')
<ipython-input-3-682894d3668c>:13: DtypeWarning: Columns (7,9,10,
12,15,16,23,31) have mixed types. Specify dtype option on import
or set low_memory=False.
    data = pd.read_csv(file, encoding='latin1')
<ipython-input-3-682894d3668c>:13: DtypeWarning: Columns (7,12,1
5,23) have mixed types. Specify dtype option on import or set low
_memory=False.
    data = pd.read_csv(file, encoding='latin1')
<ipython-input-3-682894d3668c>:13: DtypeWarning: Columns (7,12,2
3) have mixed types. Specify dtype option on import or set low_me
mory=False.
    data = pd.read_csv(file, encoding='latin1')
```

```
In [4]: # Check VAERS Data after combined
        vaers_data.head()
```

Out [4]:

	VAERS_ID	RECVDATE	STATE	AGE_YRS	CAGE_YR	CAGE_MO	SEX	RPT_DATE	SYM
									AT /
0	375646	01/01/2010	FL	28.0	28.0	NaN	F	01/01/2010	W,
1	375647	01/01/2010	IL	75.0	75.0	NaN	M	01/01/2010	lo
2	375648	01/01/2010	MN	30.0	30.0	NaN	F	01/01/2010	vorr 12
3	375650	01/01/2010	ID	35.0	35.0	NaN	M	01/01/2010	Dia ne
4	375651	01/01/2010	WA	65.0	65.0	NaN	F	01/01/2010	Bet

5 rows x 35 columns



```
In [5]: vaers_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1504868 entries, 0 to 1504867
Data columns (total 35 columns):
#   Column                Non-Null Count  Dtype
---  -
0   VAERS_ID              1504868 non-null  int64
1   RECVDATE              1504868 non-null  object
2   STATE                1246120 non-null  object
3   AGE_YRS              1277028 non-null  float64
4   CAGE_YR              1154313 non-null  float64
5   CAGE_MO              52941 non-null    float64
6   SEX                  1504868 non-null  object
7   RPT_DATE             260228 non-null  object
8   SYMPTOM_TEXT         1503931 non-null  object
9   DIED                 20508 non-null    object
10  DATEDIED             17770 non-null    object
11  L_THREAT             19893 non-null    object
12  ER_VISIT             68265 non-null    object
13  HOSPITAL             106905 non-null   object
14  HOSPDAYS             64474 non-null    float64
15  X_STAY               1706 non-null     object
16  DISABLE              25479 non-null    object
17  RECOVD              1347421 non-null  object
18  VAX_DATE             1346023 non-null  object
19  ONSET_DATE           1295583 non-null  object
20  NUMDAYS             1245190 non-null  float64
21  LAB_DATA             528172 non-null   object
22  V_ADMINBY           1504868 non-null  object
23  V_FUNDBY            266074 non-null   object
24  OTHER_MEDS          793142 non-null   object
25  CUR_ILL              710438 non-null   object
26  HISTORY              798639 non-null   object
27  PRIOR_VAX           64392 non-null    object
28  SPLTTYPE            514896 non-null   object
29  FORM_VERS           1504868 non-null  int64
30  TODAYS_DATE         1228217 non-null  object
31  BIRTH_DEFECT         712 non-null      object
32  OFC_VISIT           251900 non-null   object
33  ER_ED_VISIT         136880 non-null   object
34  ALLERGIES           552982 non-null   object
dtypes: float64(5), int64(2), object(28)
memory usage: 401.8+ MB
```

In [6]: *# Check VAERS Symptoms after combined*
 vaers_symp.head()

Out [6]:

	VAERS_ID	SYMPTOM1	SYMPTOMVERSION1	SYMPTOM2	SYMPTOMVERSION2	SYMPTOM3
0	375646	Chest X-ray normal	12.1	Chest discomfort	12.1	
1	375647	Erythema	12.1	Rash	12.1	n
2	375648	Nausea	12.1	Vomiting	12.1	
3	375650	Abdominal pain upper	12.1	Diarrhoea	12.1	l
4	375650	Nausea	12.1	NaN	NaN	

In [7]: vaers_symp.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1947800 entries, 0 to 1947799
Data columns (total 11 columns):
#   Column                Dtype
---  -
0   VAERS_ID               int64
1   SYMPTOM1               object
2   SYMPTOMVERSION1       float64
3   SYMPTOM2               object
4   SYMPTOMVERSION2       float64
5   SYMPTOM3               object
6   SYMPTOMVERSION3       float64
7   SYMPTOM4               object
8   SYMPTOMVERSION4       float64
9   SYMPTOM5               object
10  SYMPTOMVERSION5        float64
dtypes: float64(5), int64(1), object(5)
memory usage: 163.5+ MB
```

```
In [8]: # Check VAERS Vaccine after combined
vaers_vax.head()
```

```
Out [8]:
```

	VAERS_ID	VAX_TYPE	VAX_MANU	VAX_LOT	VAX_DOSE_SERIES	VAX_ROUTE
0	375646	FLU(H1N1)	SANOFI PASTEUR	NaN	1	IM
1	375647	FLU(H1N1)	SANOFI PASTEUR	UP078AA	1	IM
2	375648	FLUX(H1N1)	UNKNOWN MANUFACTURER	NaN	1	NaN
3	375650	FLUX(H1N1)	UNKNOWN MANUFACTURER	NaN	1	IM
4	375651	VARZOS	UNKNOWN MANUFACTURER	NaN	1	NaN

```
In [9]: vaers_vax.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1754353 entries, 0 to 1754352
Data columns (total 8 columns):
#   Column          Dtype
---  -
0   VAERS_ID        int64
1   VAX_TYPE        object
2   VAX_MANU        object
3   VAX_LOT         object
4   VAX_DOSE_SERIES object
5   VAX_ROUTE       object
6   VAX_SITE        object
7   VAX_NAME        object
dtypes: int64(1), object(7)
memory usage: 107.1+ MB
```

Selecting significant samples

Since the VAERS Data has more than 1 million records, I will select a small sample based on one vaccine type.

```
In [10]: # Get the number of data based on VAX_TYPE for proper selection of  
vaers_vax['VAX_TYPE'].value_counts().head()
```

```
Out[10]: COVID19      1009575  
VARZOS      108667  
FLU3        57602  
FLU4        43724  
PPV         42530  
Name: VAX_TYPE, dtype: int64
```

From the above number, I will select samples for developing a system to identify symptoms based on 'COVID19' vaccine type because the number of data is sufficient.

```
In [11]: # Based on the VAERS_ID, get 10,000 reports related to the COVID19  
# Firstly, get 10,000 reports related to the COVID19 vaccine from  
covid_vaers_id = []  
for i in range(len(vaers_vax)):  
    if vaers_vax['VAX_TYPE'][i] != 'COVID19':  
        continue  
    else:  
        if vaers_vax['VAERS_ID'][i] not in covid_vaers_id:  
            covid_vaers_id.append(vaers_vax['VAERS_ID'][i])  
            if len(covid_vaers_id) >= 10000:  
                break  
  
# Check that the list contains 10,000 VAER_ID  
len(covid_vaers_id)
```

```
Out[11]: 10000
```



```
In [12]: # Secondly, get 10,000 reports from the column SYMPTOM_TEXT in the
covid_reports = pd.DataFrame({'VAERS_ID': [], 'SYMPTOM_TEXT': []})
for vaers_id in covid_vaers_id:

    new_row = {'VAERS_ID': vaers_id,
               'SYMPTOM_TEXT': vaers_data[vaers_data['VAERS_ID'] =
covid_reports = pd.concat([covid_reports, pd.DataFrame(new_row

# Convert 'VAERS_ID' to integers
covid_reports['VAERS_ID'] = covid_reports['VAERS_ID'].astype(int)

# Check the context in the corpus
covid_reports
```

Out [12]:

	VAERS_ID	SYMPTOM_TEXT
0	2669769	body aches, fatigue Narrative: Took OTC Tyleno...
1	2527460	Headache, Myalgia, NauseaVomiting, chills Narr...
2	2673135	Headache, Fever, Body aches Narrative: Other ...
3	2672717	Headache & Myalgia Narrative: Other Relevant...
4	902418	Patient experienced mild numbness traveling fr...
...
9995	916173	REDNESS TO INJECTION SITE 12-30-20. PROGRESSED...
9996	916174	Patient described joint and muscle pain in the...
9997	916176	Numbness and tingling on left side of face, ey...
9998	916177	HIVES, tachypnea, vomiting - normal saline, ne...
9999	916178	Excessive swelling to left axillary lymph node...

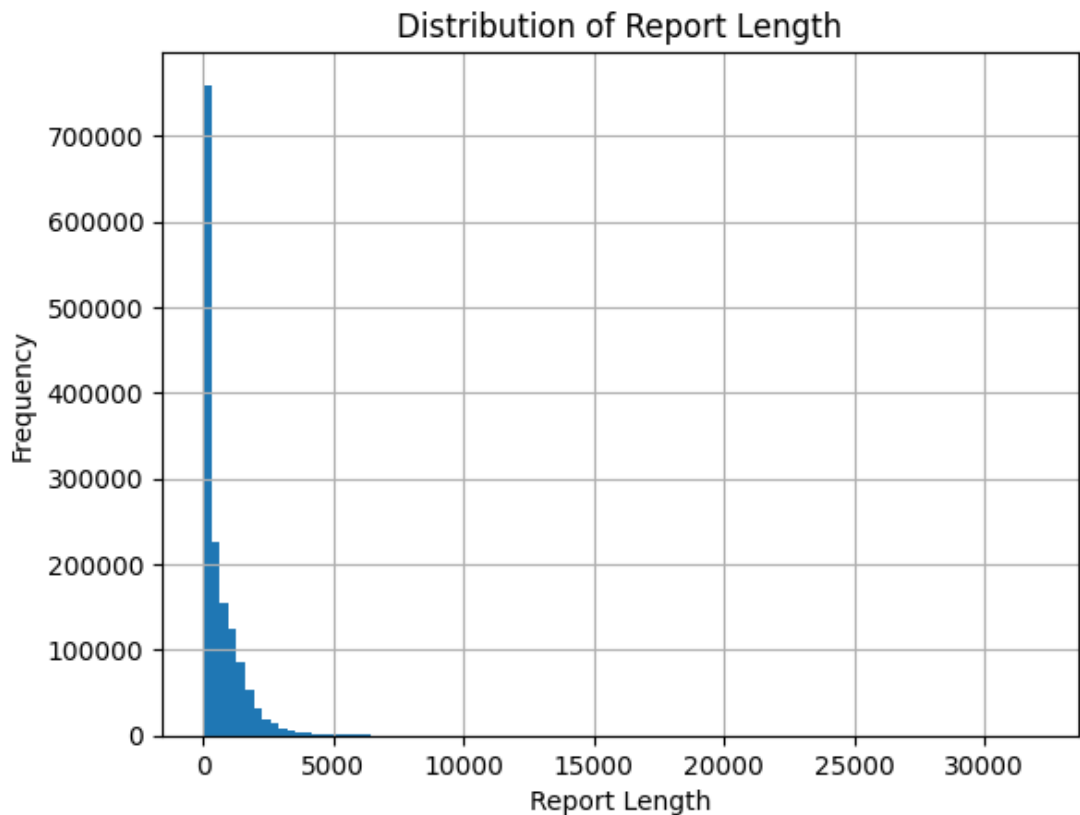
10000 rows × 2 columns

The distribution of the length of different reports

```
In [13]: # The distribution of the length of different reports
vaers_data['REPORT_LENGTH'] = vaers_data['SYMPTOM_TEXT'].str.len()
report_length_stats = vaers_data['REPORT_LENGTH'].describe()
report_length_stats
```

```
Out [13]: count    1.503931e+06
mean      6.496922e+02
std       8.788270e+02
min       1.000000e+00
25%      1.080000e+02
50%      3.130000e+02
75%      9.310000e+02
max      3.199200e+04
Name: REPORT_LENGTH, dtype: float64
```

```
In [14]: # Plot the graph to see the distribution of the length of differen
vaers_data['REPORT_LENGTH'].hist(bins=100)
plt.xlabel('Report Length')
plt.ylabel('Frequency')
plt.title('Distribution of Report Length')
plt.show()
```

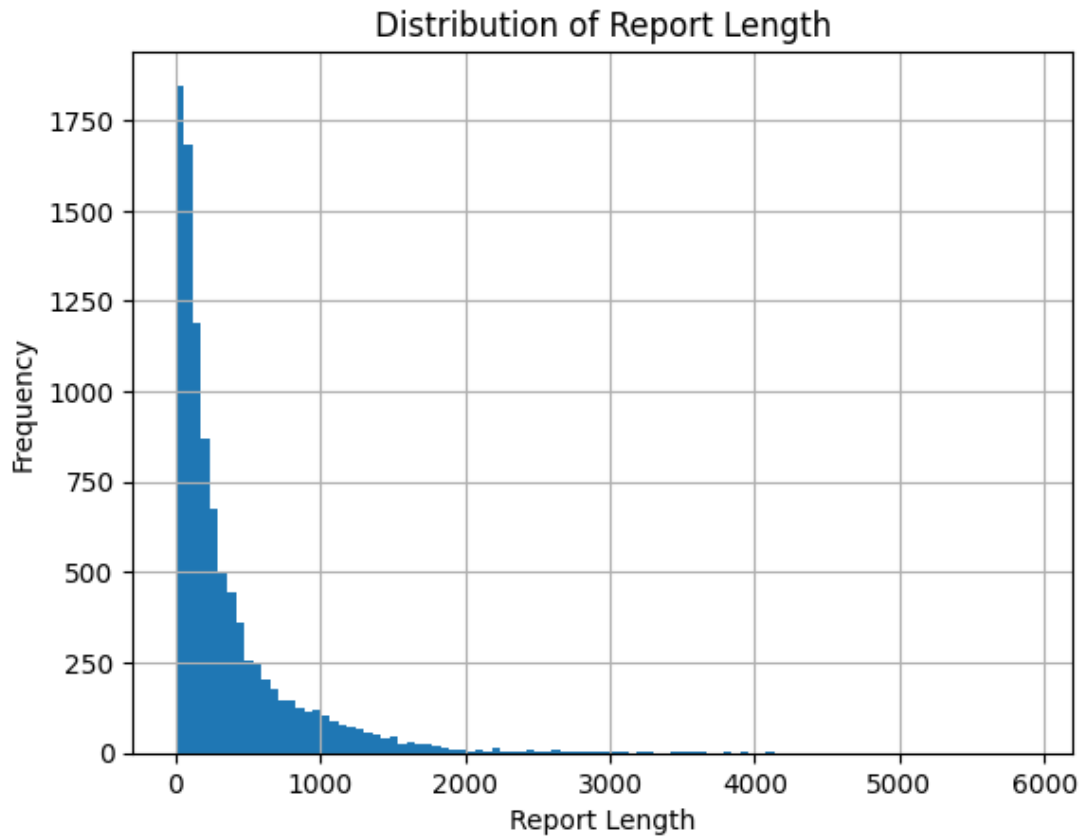


The distribution of the length of different reports (10,000 Covid-19 reports)

```
In [15]: # The distribution of the length of different reports (10,000 Covi
covid_reports['REPORT_LENGTH'] = covid_reports['SYMPTOM_TEXT'].str
covid_report_length_stats = covid_reports['REPORT_LENGTH'].describ
covid_report_length_stats
```

```
Out[15]: count    10000.000000
mean       375.680100
std        494.695037
min         2.000000
25%        81.000000
50%       198.000000
75%       466.000000
max       5904.000000
Name: REPORT_LENGTH, dtype: float64
```

```
In [16]: # Plot the graph to see the distribution of the length of differen
covid_reports['REPORT_LENGTH'].hist(bins=100)
plt.xlabel('Report Length')
plt.ylabel('Frequency')
plt.title('Distribution of Report Length')
plt.show()
```



Standard symptoms

Build the corresponding standard symptom list with the symptoms in VAERS Symptoms table, and find the most frequent 100 symptoms.

```
In [17]: # Build the standard symptom list
std_symp_dict = {}
for vaers_id in covid_reports['VAERS_ID']:
    for col_symp in ['SYMPTOM1', 'SYMPTOM2', 'SYMPTOM3', 'SYMPTOM4']:
        symptom = vaers_symp.loc[vaers_symp['VAERS_ID'] == vaers_id][col_symp]
        if pd.notna(symptom):
            std_symp_dict[symptom] = std_symp_dict.get(symptom, 0)

# Get the list of standard symptoms
std_symp = list(std_symp_dict.keys())
std_symp
```

```
Out[17]: ['Fatigue',
          'Pain',
          'Chills',
          'Headache',
          'Myalgia',
          'Nausea',
          'Vomiting',
          'Pyrexia',
          'Hypoaesthesia',
          'Injection site hypoaesthesia',
          'Erythema',
          'Feeling hot',
          'Flushing',
          'Dizziness',
          'Electrocardiogram normal',
          'Hyperhidrosis',
          'Laboratory test normal',
          'Presyncope',
          'Dysgeusia',
          'Oral pruritus',
          'Paraesthesia',
          'Paraesthesia oral',
          'Parosmia',
          'Chest discomfort',
          'Defaecation urgency',
          'Diarrhoea',
          'Chest pain',
          'Feeling abnormal',
          'Intensive care',
          'Blood pressure increased',
          'Dyspnoea',
          'Visual impairment',
          'Ear discomfort',
          'Palpitations',
          'Throat tightness',
          'Inappropriate schedule of product administration',
          'Eye pruritus',
          'Asthenia',
          'Heart rate increased',
          'Hypertension',
          'Blood glucose normal',
          'Blood test',
          'Injection site pain',
          'Pain in extremity',
          'Malaise',
          'Lip swelling',
          'Swelling face',
          'Feeling of body temperature change',
          'Rash',
          'Oropharyngeal pain',
          'Cold sweat',
          'Condition aggravated',
          'Panic attack',
          'Abdominal pain upper',
          'Decreased appetite',
          'Injection site erythema',
          'Injection site induration',
          'Local reaction',
          'Immediate post-injection reaction',
          'Joint swelling',
          'Peripheral swelling',
```

'Pruritus',
'Somnolence',
'Mobility decreased',
'Product administered at inappropriate site',
'Injected limb mobility decreased',
'Injection site oedema',
'Musculoskeletal chest pain',
'Hypotension',
'Pallor',
'Disturbance in attention',
'Impaired work ability',
'Abnormal sensation in eye',
'Dry mouth',
'Euphoric mood',
'Arthralgia',
'Injection site swelling',
'Injection site pruritus',
'Injection site bruising',
'Cough',
'Vaccination site bruising',
'Vaccination site haemorrhage',
'Heart rate decreased',
'Tremor',
'Rash maculo-papular',
'Influenza like illness',
'Eye movement disorder',
'Memory impairment',
'Unresponsive to stimuli',
'Dysphagia',
'Hypersensitivity',
'Hypoaesthesia oral',
'Swollen tongue',
'Device connection issue',
'Product design issue',
'Product quality issue',
'Syringe issue',
'Hot flush',
'Circumstance or information capable of leading to medication error',
'Contraindicated product administered',
'Anxiety',
'Fall',
'Seizure like phenomena',
'Wheezing',
'Loss of consciousness',
'Orthostatic hypotension',
'Tearfulness',
'Head discomfort',
'Disorientation',
'Lacrimation increased',
'Throat irritation',
'Dry throat',
'Gait inability',
'Tachycardia',
'Urticaria',
'Oedema peripheral',
'Atelectasis',
'Chest X-ray abnormal',
'Computerised tomogram thorax normal',
'Tenderness',
'Anaesthesia oral',

'Dysarthria',
'Pharyngeal hypoaesthesia',
'Angioedema',
'Rash pustular',
'Pharyngeal paraesthesia',
'Incorrect dose administered',
'Body temperature increased',
'Anosmia',
'Influenza virus test',
'Chest X-ray normal',
'Computerised tomogram normal',
'Glassy eyes',
'Visual tracking test abnormal',
'Abdominal pain',
'Bowel movement irregularity',
'Exercise tolerance decreased',
'Confusional state',
'Slow response to stimuli',
'Haematuria',
'Injection site reaction',
'Migraine',
'Muscle fatigue',
'Anaphylactic reaction',
'Chest X-ray',
'Lymphadenopathy',
'Eye pain',
'Rosacea',
'Vertigo',
'Rash erythematous',
'Back pain',
'Limb discomfort',
'Neck pain',
'Syncope',
'SARS-CoV-2 test negative',
'Mental status changes',
'Ear pruritus',
'Eye swelling',
'Ocular discomfort',
'Flank pain',
'Feeling cold',
'Electrocardiogram abnormal',
'Heart rate irregular',
'Ventricular extrasystoles',
'Injection site rash',
'Activated partial thromboplastin time shortened',
'Alanine aminotransferase normal',
'Anion gap',
'Aspartate aminotransferase normal',
'Basophil count normal',
'Dysphonia',
'Pharyngeal swelling',
'Pain in jaw',
'Wheelchair user',
'Dysstasia',
'Muscle spasms',
'Injection site urticaria',
'Muscular weakness',
'Sensation of foreign body',
'Neurological examination normal',
'Faeces discoloured',
'Palatal oedema',

'Nasal congestion',
'Photophobia',
'Circumoral swelling',
'Oropharyngeal discomfort',
'Cranial nerve disorder',
'Nerve conduction studies abnormal',
'Fasting',
'Swelling',
'Rash pruritic',
'Rhinorrhoea',
'Product preparation issue',
'Asthma',
'Frequent bowel movements',
'Injection site joint pain',
'Taste disorder',
'Increased upper airway secretion',
'Electrocardiogram',
'Laboratory test',
'Angina pectoris',
'Burning sensation',
'Cardiac discomfort',
'Product preparation error',
'Spinal pain',
'Blood urea increased',
'C-reactive protein increased',
'Full blood count normal',
'International normalised ratio normal',
'Sleep disorder',
'Human chorionic gonadotropin',
'Sinus tachycardia',
'Tracheomalacia',
'Skin warm',
'Dyspepsia',
'Hunger',
'Blepharospasm',
'Peripheral coldness',
'Skin discolouration',
'Tinnitus',
'Ageusia',
'Exposure during pregnancy',
'Injection site nodule',
'Blood thyroid stimulating hormone decreased',
'Urine analysis normal',
'Computerised tomogram',
'Facial paralysis',
'Magnetic resonance imaging',
'Oral discomfort',
'Blood chloride normal',
'Blood creatinine normal',
'Blood potassium decreased',
'Blood sodium normal',
'Seizure',
'Insomnia',
'Acoustic stimulation tests abnormal',
'Deafness unilateral',
'Maternal exposure during breast feeding',
'Palmar erythema',
'Breast swelling',
'Ocular hyperaemia',
'Extra dose administered',
'Injection site haemorrhage',

'Injection site warmth',
'Dry skin',
'Skin exfoliation',
'Blood glucose increased',
'Unevaluable event',
'Body temperature decreased',
'Hypothermia',
'Dyskinesia',
'Tongue disorder',
'Facial discomfort',
'Muscle disorder',
'Balance disorder',
'Abdominal discomfort',
'Rash macular',
'Eosinophil percentage decreased',
'Tongue pruritus',
'Nervousness',
'Vision blurred',
'Full blood count',
'Metabolic function test',
'Urine analysis',
'Blood pressure abnormal',
'Eosinophil percentage',
'Heart rate abnormal',
'Discomfort',
'Dizziness postural',
'Joint stiffness',
'Antinuclear antibody positive',
'Feeling drunk',
'Eye disorder',
'Gait disturbance',
'Blood pressure decreased',
'Computerised tomogram head normal',
'Cardiac monitoring',
'Angiopathy',
'Feeding disorder',
'Throat clearing',
'Injection site discomfort',
'Thirst',
'Vaccination site pain',
'Retching',
'Dysphemia',
'Lethargy',
'Allergy test',
'Pain of skin',
'Ear disorder',
'Feeling jittery',
'Oxygen saturation decreased',
'Extrasystoles',
'Increased appetite',
'Muscle tightness',
'Sinus bradycardia',
'Transient ischaemic attack',
'Fear of injection',
'Nodule',
'Jaw disorder',
'Blood glucose decreased',
'Respiratory tract congestion',
'Periorbital swelling',
'Skin tightness',
'Diplopia',

'Vitreous floaters',
'Hyperglycaemia',
'Aphonia',
'Obstructive airways disorder',
'Odynophagia',
'Autoimmune disorder',
'Posture abnormal',
'Neuralgia',
'Rash vesicular',
'Musculoskeletal discomfort',
'SARS-CoV-2 test positive',
'Gastric disorder',
'Blood pressure fluctuation',
'Injection site vesicles',
'Blindness transient',
'Loss of personal independence in daily activities',
'Musculoskeletal stiffness',
'Eyelids pruritus',
'Skin burning sensation',
'Blood gases normal',
'Calcium ionised normal',
'Metabolic function test normal',
'Blood albumin normal',
'Blood calcium decreased',
'Blood magnesium normal',
'Eye discharge',
'Intercepted product preparation error',
'Product dispensing error',
'Angiogram',
'Blood pressure systolic',
'Eyelid function disorder',
'Axillary mass',
'Electrocardiogram QT prolonged',
'Troponin normal',
'Blood creatinine increased',
'Computerised tomogram abdomen',
'Computerised tomogram abnormal',
'Rash papular',
'Dyspnoea exertional',
'Vaginal haemorrhage',
'Facial pain',
'Hemiparaesthesia',
'Pulse abnormal',
'Lip pruritus',
'Injection site streaking',
'Pelvic pain',
'Device malfunction',
'Product administration error',
'Underdose',
'Hyporesponsive to stimuli',
'Bone pain',
'Facial asymmetry',
'Blood glucose',
'Differential white blood cell count',
'Papule',
'Cyanosis',
'Influenza A virus test negative',
'Dehydration',
'Gout',
'Blood alkaline phosphatase normal',
'Blood creatine phosphokinase decreased',

'Haematoma muscle',
'Sensory loss',
'Injection site paraesthesia',
'Tunnel vision',
'Formication',
'Eye irritation',
'Night sweats',
'Vaccination site erythema',
'Vaccination site swelling',
'Axillary pain',
'Induration',
'Lymphoedema',
'Petechiae',
'Cardiac monitoring normal',
'Differential white blood cell count normal',
'Skin ulcer',
'Lymph node pain',
'SARS-CoV-2 test',
'SARS-CoV-2 antibody test',
'Photopsia',
'Panic reaction',
'Gastrointestinal disorder',
'Mydriasis',
'Pupillary reflex impaired',
'Rash morbilliform',
'Lung disorder',
'Ear swelling',
'Abdominal pain lower',
'Blood triglycerides normal',
'Constipation',
'Oedema',
'Autoscopy',
'Incorrect route of product administration',
'Abdominal distension',
'Flatulence',
'Blister',
'Ear pain',
'Pelvic girdle pain',
'Hypoacusis',
'Deafness neurosensory',
'Painful respiration',
'Pleurisy',
'Pleuritic pain',
'Metrorrhagia',
'Contusion',
'Lipase',
'Pancreatitis acute',
'Oxygen saturation normal',
'Breast mass',
'Breast oedema',
'Influenza B virus test',
'Respiratory syncytial virus test negative',
'Cardiac flutter',
'HIV antibody negative',
'Sinus congestion',
'Comminuted fracture',
'Epistaxis',
'Facial bones fracture',
'Haemorrhage',
'Supraventricular tachycardia',
'Anaphylactoid reaction',

'Oxygen therapy',
'Stridor',
'Influenza virus test negative',
'Hypersomnia',
'Productive cough',
'Lymphadenitis',
'Motion sickness',
'COVID-19',
'Herpes zoster',
'Penile rash',
'Scrotal dermatitis',
'Galactostasis',
'Herpes simplex',
'Herpes simplex reactivation',
'Atrial fibrillation',
'Scalloped tongue',
'Tongue discomfort',
'Periorbital oedema',
'Ear congestion',
'Sneezing',
'Computerised tomogram head',
'Electrocardiogram ambulatory',
'Blood culture',
'Culture urine',
'Cardioversion',
'Ventricular tachycardia',
'Blood test normal',
'Dermatitis allergic',
'Neuropathy peripheral',
'Lactation disorder',
'Menstruation delayed',
'Pregnancy test negative',
'Joint range of motion decreased',
'Joint noise',
'Breath sounds',
'Paraesthesia ear',
'Sensory disturbance',
'Blood calcium increased',
'Crying',
'Tenosynovitis',
'Glossitis',
'Coccydynia',
'Hyperventilation',
'Arrhythmia',
'Cerebellar ataxia',
'Arthritis',
'Weight bearing difficulty',
'Ultrasound scan',
'Eye paraesthesia',
'Conjunctival haemorrhage',
'Angiogram cerebral abnormal',
'Central nervous system lesion',
'Demyelination',
'Fine motor skill dysfunction',
'Alanine aminotransferase increased',
'Blood sodium decreased',
'Eyelid rash',
'Product delivery mechanism issue',
'Injection site mass',
'Fear',
'Drainage',

'Hypogeusia',
'Asthenopia',
'Magnetic resonance imaging brain',
'Eye haemorrhage',
'Retinal exudates',
'Increased viscosity of upper respiratory secretion',
'Vaccination site induration',
'Hospice care',
'Upper respiratory tract irritation',
'Chapped lips',
'Appendicectomy',
'Appendicitis',
'Computerised tomogram abdomen abnormal',
'Laparoscopic surgery',
'Periorbital pain',
'Swelling of eyelid',
'Platelet count decreased',
'Thrombocytopenia',
'Lip erythema',
'Face oedema',
'Influenza A virus test',
'Activated partial thromboplastin time',
'Exposure to SARS-CoV-2',
'Blood cholesterol increased',
'Blood triglycerides increased',
'Computerised tomogram head abnormal',
'Migraine with aura',
'Hyperaesthesia',
'Temperature intolerance',
'Sluggishness',
'Cardiac telemetry normal',
'Lip blister',
'Magnetic resonance imaging normal',
'Musculoskeletal disorder',
'Respiratory rate increased',
'Bradycardia',
'Blindness',
'Nasopharyngitis',
'Photosensitivity reaction',
'Gastrooesophageal reflux disease',
'Piloerection',
'Influenza',
'Paranasal sinus discomfort',
'Sinus headache',
'Poor quality sleep',
'Blood magnesium',
'Blood thyroid stimulating hormone',
'Periorbital discomfort',
'Myocardial infarction',
'Hypoglycaemia',
'Treatment noncompliance',
'Trigeminal neuralgia',
'Fibrin D dimer normal',
'Adverse reaction',
'C-reactive protein',
'Herpes virus test',
'Red blood cell sedimentation rate',
'Angiogram normal',
'Vaccination site irritation',
'Hypertonia',
'Mouth injury',

'Oral mucosal blistering',
'Echocardiogram',
'Allergy to synthetic fabric',
'Breath sounds abnormal',
'Angiogram cerebral normal',
'Arteriogram carotid normal',
'Drug screen positive',
'Mass',
'Fibrin D dimer',
'No adverse event',
'Haemoglobin decreased',
'Oversensing',
'Injection site inflammation',
'Blood potassium normal',
'Blood urea decreased',
'SARS-CoV-2 antibody test negative',
'Deafness',
'Oral herpes',
'Cellulitis',
'Histamine intolerance',
'Upper-airway cough syndrome',
'Dysuria',
'Irritability',
'Facial paresis',
'Basophil percentage decreased',
'Aphasia',
'Hypopnoea',
'Acute coronary syndrome',
'Altered state of consciousness',
'Blood electrolytes normal',
'Streptococcus test negative',
'Acute myocardial infarction',
'Blood phosphorus normal',
'Brain natriuretic peptide increased',
'Middle insomnia',
'Blood thyroid stimulating hormone increased',
'Brain natriuretic peptide normal',
'Enlarged uvula',
'Exposure via skin contact',
'Injection site extravasation',
'Nasal mucosal discolouration',
'Oedema mouth',
'Livedo reticularis',
'Hyposmia',
'Micturition urgency',
'Vaccination complication',
'Anaemia',
'Hyperacusis',
'Migraine without aura',
'Adverse drug reaction',
'Liver function test normal',
'Platelet count normal',
'Electrocardiogram QRS complex normal',
'Electrocardiogram ST segment normal',
'Premature delivery',
'Premature labour',
'Premature rupture of membranes',
'Conjunctivitis',
'Intraocular pressure test',
'Ophthalmological examination',
'Uveitis',

'Blood pressure immeasurable',
'Respiratory tract irritation',
'Diagnostic procedure',
'Gingival bleeding',
'Glossodynia',
'Mouth swelling',
'Bronchitis',
'Skin swelling',
'Blood pressure systolic increased',
'Blood urine present',
'Pharyngeal erythema',
'Rectal haemorrhage',
'Histamine level increased',
'Middle ear effusion',
'Skin laceration',
'Wound closure',
'Base excess decreased',
'Blood pH increased',
'Blood pressure measurement',
'Acute generalised exanthematous pustulosis',
'Injection site coldness',
'Polymerase chain reaction',
'Blood bilirubin increased',
'Blood osmolarity increased',
'Cardiovascular evaluation',
'Electrocardiogram change',
'Prostatomegaly',
'Testicular swelling',
'Speech disorder',
'Muscle twitching',
'Limb immobilisation',
'Bilevel positive airway pressure',
'Status asthmaticus',
'Refusal of treatment by patient',
'Grip strength decreased',
'Basophil count decreased',
'Basophil percentage',
'Blood bicarbonate normal',
'Endotracheal intubation',
'Bedridden',
'Dry eye',
'Muscle swelling',
'C-reactive protein normal',
'Clumsiness',
'Eosinophil count increased',
'Magnetic resonance imaging brain normal',
'Neurological symptom',
'Bronchospasm',
'Dissociation',
'Hidradenitis',
'Tooth fracture',
'Auditory disorder',
'Respiratory tract oedema',
'Haemodynamic test normal',
'Eye inflammation',
'Initial insomnia',
'Menorrhagia',
'Respiratory distress',
'Bradykinesia',
'Skin irritation',
'Chromaturia',

'Tryptase',
'Testicular pain',
'Facial spasm',
'Trichoglossia',
'Gastrointestinal sounds abnormal',
'Lung hyperinflation',
'Blood pressure orthostatic increased',
'Blood thyroid stimulating hormone normal',
'Cardiac telemetry',
'Blood pressure systolic abnormal',
'Nasal pruritus',
'Decreased activity',
'Tongue eruption',
'Anal incontinence',
'Acute kidney injury',
'Atrial tachycardia',
'Glycosylated haemoglobin decreased',
'Activated partial thromboplastin time normal',
'Illness',
'Electroencephalogram',
'Generalised tonic-clonic seizure',
'Haematology test normal',
'Troponin I',
'Injection site cellulitis',
'Impaired driving ability',
'Gastrointestinal haemorrhage',
'SARS-CoV-2 antibody test positive',
'White coat hypertension',
'Exposure to extreme temperature',
'Erythema of eyelid',
'Tachypnoea',
'Drug eruption',
'Antibody test',
'Inflammation',
'Adverse event',
'Gaze palsy',
'Hallucination',
'Lip oedema',
'Alcohol use',
'Blood pressure systolic decreased',
'Blood albumin increased',
'Full blood count abnormal',
'Liver function test',
'Blood parathyroid hormone increased',
'Occupational exposure to SARS-CoV-2',
'Bed rest',
'Blood pressure normal',
'Troponin increased',
'Angiogram cerebral',
'Arteriogram carotid',
'Culture urine negative',
'Emotional distress',
'Product administered to patient of inappropriate age',
'Hypertensive urgency',
'Supraventricular extrasystoles',
'Neurological examination',
'Mouth ulceration',
'Scab',
'Gingival pain',
'Gingival swelling',
'Lip pain',

'Sedation',
'Injection site lymphadenopathy',
'Ultrasound scan abnormal',
'Anticonvulsant drug level',
'Blood urea normal',
'Acne',
'Haemoglobin normal',
'Red blood cell count normal',
'Imaging procedure',
'Metamorphopsia',
'Mechanical ventilation',
'Blood creatine phosphokinase normal',
'Immunisation',
'Incomplete course of vaccination',
'Product dose omission in error',
'Nasal discomfort',
'Phonophobia',
'Hypoaesthesia eye',
'Nerve compression',
'Incoherent',
'Alpha 1 globulin normal',
'Anion gap normal',
'Radiculopathy',
'Viral titre',
'Delusion',
'Tongue biting',
'Dermatitis exfoliative generalised',
'Joint dislocation',
'Upper limb fracture',
'X-ray abnormal',
'Computerised tomogram thorax',
'Aura',
'Brain natriuretic peptide',
'Nipple pain',
'Nipple swelling',
'Injection site discolouration',
'Breast pain',
'Lymphocyte count increased',
'Blood pressure orthostatic abnormal',
'Cheilitis',
'Dental paraesthesia',
'Tension headache',
'Echocardiogram abnormal',
'Energy increased',
'Neurological examination abnormal',
'Bronchial hyperreactivity',
'Saliva altered',
'COVID-19 pneumonia',
'Clonic convulsion',
'Stress',
'Injection site hypersensitivity',
'Sensitive skin',
'Blood pressure diastolic increased',
'Pharyngeal oedema',
'Restlessness',
'Skin lesion',
'Oral mucosal erythema',
'Pneumonia',
'Rales',
'Diarrhoea haemorrhagic',
'Product label issue',

'Alopecia',
'Thrombosis',
'Breast milk discolouration',
'Immunoglobulin therapy',
'Blood creatine increased',
'Blood potassium increased',
'Chronic kidney disease',
'Injection site haematoma',
'Binocular eye movement disorder',
'Coordination abnormal',
'Fibrin D dimer increased',
'Haemoptysis',
'Skin hyperpigmentation',
'Viral rash',
'Lip haemorrhage',
'Angina unstable',
'Skin disorder',
'Blood lactic acid decreased',
'Similar reaction on previous exposure to drug',
'Gastrointestinal pain',
'Herpes virus infection',
'Foaming at mouth',
'Pulse absent',
'Respiratory arrest',
'Blood culture negative',
'Computerised tomogram abdomen normal',
'Scan with contrast normal',
'Hallucination, visual',
'Catheterisation cardiac abnormal',
'Coronary artery occlusion',
'Conjunctival hyperaemia',
'Muscle discomfort',
'Genital ulceration',
'Injection site irritation',
'Tongue ulceration',
'Tension',
'Hypertensive emergency',
'Dysacusis',
'Raynaud's phenomenon',
'Ulcer',
'Delirium',
'Skin reaction',
'Speech sound disorder',
'Computerised tomogram pelvis',
'Contraindication to vaccination',
'Bursitis',
'Sciatica',
'Blood lactic acid normal',
'Therapeutic response unexpected',
'Vaginal ulceration',
'Eyelid ptosis',
'Mental impairment',
'Blood insulin increased',
'Electromyogram',
'X-ray',
'Nodular rash',
'Incisional drainage',
'Tongue dry',
'Head injury',
'Blood calcium',
'Blood creatine phosphokinase',

'Blood creatinine',
'Blood electrolytes',
'Lymph node palpable',
'Tonsillar hypertrophy',
'Abnormal behaviour',
'Echocardiogram normal',
'Faeces pale',
'Rhinalgia',
'Rhinitis',
'Groin pain',
'C-reactive protein decreased',
'Bruxism',
'Hypertensive crisis',
'Illusion',
'Nystagmus',
'Psychogenic seizure',
'Dermatitis',
'Eyelid sensory disorder',
'Muscle strain',
'Agitation',
'Oral disorder',
'Oral pain',
'Choking',
'Hemiparesis',
'Mononucleosis heterophile test negative',
'Pustule',
'Hypotonia',
'Vein discolouration',
'Coxsackie virus test positive',
'Eating disorder',
'Herpangina',
'Abnormal dreams',
'Depressed mood',
'Corrective lens user',
'Eczema',
'Suprapubic pain',
'Respiratory viral panel',
'Pulmonary congestion',
'X-ray normal',
'Vocal cord disorder',
'Acoustic stimulation tests',
'Ear infection',
'Myosclerosis',
'Aphthous ulcer',
'Hyperaesthesia teeth',
'Ophthalmological examination abnormal',
'Vitreous detachment',
'Intestinal obstruction',
'Troponin',
'Menstruation normal',
'Death',
'Hypophagia',
'Hypoxia',
'Food poisoning',
'Computerised tomogram pancreas abnormal',
'Lipase increased',
'White blood cell count increased',
'Claustrophobia',
'Nightmare',
'Antinuclear antibody',
'Abdominal rigidity',

'Tonsillitis',
'Pulmonary pain',
'Trismus',
'Pancreatitis',
'Weight increased',
'Haematochezia',
'Mucous stools',
'Basal ganglia haemorrhage',
'Iridocyclitis',
'Diabetes mellitus',
'Genital swelling',
'Cardiac disorder',
'Neutrophil percentage increased',
'Limb mass',
'Synovitis',
'Blood cholesterol normal',
'Monoplegia',
'Hangover',
'Renal pain',
'Candida infection',
'Atopy',
'Toothache',
'Tonsillar erythema',
'Sialoadenitis',
'Submaxillary gland enlargement',
'Needle issue',
'Inflammatory marker test',
'Physical examination normal',
'Motor dysfunction',
'Dysmenorrhoea',
'Aspartate aminotransferase increased',
'Halo vision',
'Cerebrovascular accident',
'Muscle contractions involuntary',
'Catatonia',
'HIV test',
'Stomatitis',
'Hysteroscopy',
'Asymptomatic COVID-19',
'Herpes zoster cutaneous disseminated',
'Dyschezia',
'Gallbladder disorder',
'Blood cholesterol',
'Body temperature',
'Exposure via breast milk',
'Capillary nail refill test abnormal',
'Extensive swelling of vaccinated limb',
'Musculoskeletal pain',
'Eyelid disorder',
'Product storage error',
'Eyelid irritation',
'Oral contusion',
'Administration site pain',
'Antipsychotic drug level',
'Dyslexia',
'Eyelid contusion',
'Accidental exposure to product',
'Injury associated with device',
'Tongue coated',
'Sinusitis',
'Endothelial dysfunction',

```
'Coma scale normal',  
'Thinking abnormal',  
'Bradyphrenia',  
'Blood alkaline phosphatase increased',  
'Blood calcium normal',  
'Irregular breathing',  
'Snoring',  
'Staring',  
'Oesophageal spasm',  
'Paranasal sinus hypersecretion',  
'Dermatitis psoriasiform',  
'Breast inflammation',  
'Deja vu',  
'Facial nerve disorder',  
'Uterine spasm',  
'Culture stool negative',  
'Apnoea',  
'Wound complication',  
'Butterfly rash',  
'Immunodeficiency',  
'Pulmonary embolism',  
'Injection site muscle weakness',  
'Palatal swelling',  
'Miliaria',  
'Arthropod bite',  
...]
```

```
In [18]: # Sort the dictionary by values (frequencies) in descending order
sorted_symp_freq = sorted(std_symp_dict.items(), key=lambda item:

# Get the top 100 symptoms with the highest frequencies
top_symp = [symp for symp, freq in sorted_symp_freq[:100]]

# Get the list of the top 100 symptoms
top_symp
```

```
Out[18]: ['Headache',  
          'Fatigue',  
          'Chills',  
          'Dizziness',  
          'Pain',  
          'Nausea',  
          'Injection site pain',  
          'Pyrexia',  
          'Pain in extremity',  
          'Arthralgia',  
          'Myalgia',  
          'Pruritus',  
          'Paraesthesia',  
          'Dyspnoea',  
          'Flushing',  
          'Rash',  
          'Hypoaesthesia',  
          'Cough',  
          'Diarrhoea',  
          'Injection site erythema',  
          'Urticaria',  
          'Erythema',  
          'Chest discomfort',  
          'Feeling hot',  
          'Asthenia',  
          'Hyperhidrosis',  
          'Heart rate increased',  
          'Palpitations',  
          'Malaise',  
          'Paraesthesia oral',  
          'Feeling abnormal',  
          'Injection site swelling',  
          'Blood pressure increased',  
          'Vomiting',  
          'Tachycardia',  
          'Lymphadenopathy',  
          'Throat irritation',  
          'Hypoaesthesia oral',  
          'Anxiety',  
          'Body temperature increased',  
          'Dysgeusia',  
          'Oropharyngeal pain',  
          'Chest pain',  
          'Back pain',  
          'Throat tightness',  
          'Injection site pruritus',  
          'Immediate post-injection reaction',  
          'Condition aggravated',  
          'Injection site warmth',  
          'Unevaluable event',  
          'Dysphagia',  
          'Lip swelling',  
          'Hot flush',  
          'Hypertension',  
          'Electrocardiogram',  
          'SARS-CoV-2 test negative',  
          'Abdominal pain',  
          'Tremor',  
          'Decreased appetite',  
          'Neck pain',  
          'Peripheral swelling',
```

```
'Cold sweat',  
'Impaired work ability',  
'Electrocardiogram normal',  
'Feeling cold',  
'Burning sensation',  
'Pharyngeal swelling',  
'Rash erythematous',  
'Migraine',  
'Swollen tongue',  
'Swelling face',  
'Skin warm',  
'Rhinorrhoea',  
'Abdominal discomfort',  
'Abdominal pain upper',  
'Nasal congestion',  
'Swelling',  
'SARS-CoV-2 test positive',  
'Injection site induration',  
'Limb discomfort',  
'Dry mouth',  
'Rash pruritic',  
'Dysphonia',  
'Ageusia',  
'Full blood count',  
'Syncope',  
'Injected limb mobility decreased',  
'Pallor',  
'Lymph node pain',  
'Mobility decreased',  
'Influenza like illness',  
'Hypersensitivity',  
'Blood glucose normal',  
'Discomfort',  
'Musculoskeletal stiffness',  
'Anosmia',  
'Chest X-ray',  
'Injection site rash',  
'Anaphylactic reaction',  
'Muscle spasms']
```

The distribution of the number of different symptoms

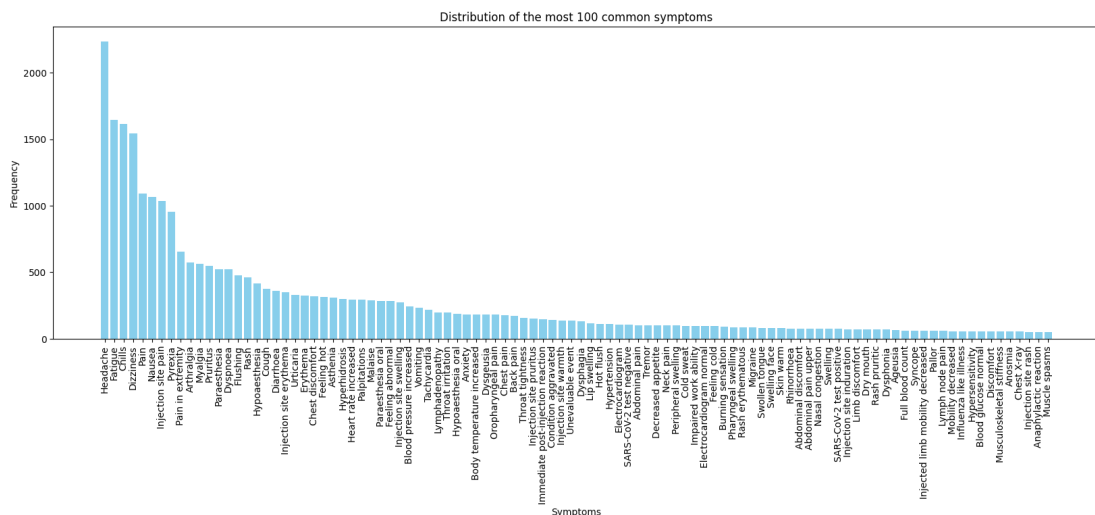

```
In [19]: # Extract the symptoms and their corresponding frequencies
symptoms = list(std_symp_dict.keys())
frequencies = list(std_symp_dict.values())

# Sort symptoms and frequencies by frequency in descending order
sorted_indices = sorted(range(len(frequencies)), key=lambda k: frequencies[k], reverse=True)
symptoms = [symptoms[i] for i in sorted_indices]
frequencies = [frequencies[i] for i in sorted_indices]

# Take only the top 100 symptoms
top_symptoms = symptoms[:100]
top_frequencies = frequencies[:100]

# Create a bar plot to visualize the distribution
plt.figure(figsize=(20, 6)) # Set the figure size
plt.bar(top_symptoms, top_frequencies, color='skyblue')
plt.xlabel('Symptoms')
plt.ylabel('Frequency')
plt.title('Distribution of the most 100 common symptoms')
plt.xticks(rotation=90)

# Show the plot
plt.show()
```



3. Extracting Symptom-related Entities

```
In [20]: !pip install stanza

# Import necessary library
import stanza

# download and initialize a mimic pipeline with an i2b2 NER model
stanza.download('en', package='mimic', processors={'ner': 'i2b2'})
nlp = stanza.Pipeline('en', package='mimic', processors={'ner': 'i
```

Collecting stanza

Downloading stanza-1.7.0-py3-none-any.whl (933 kB)

933.2/933.2 kB 11.6

MB/s eta 0:00:00

Collecting emoji (from stanza)

Downloading emoji-2.9.0-py2.py3-none-any.whl (397 kB)

397.5/397.5 kB 33.3

MB/s eta 0:00:00

Requirement already satisfied: numpy in /usr/local/lib/python3.10/dist-packages (from stanza) (1.23.5)

Requirement already satisfied: protobuf>=3.15.0 in /usr/local/lib/python3.10/dist-packages (from stanza) (3.20.3)

Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-packages (from stanza) (2.31.0)

Requirement already satisfied: networkx in /usr/local/lib/python3.10/dist-packages (from stanza) (3.2.1)

Requirement already satisfied: toml in /usr/local/lib/python3.10/dist-packages (from stanza) (0.10.2)

Requirement already satisfied: torch>=1.3.0 in /usr/local/lib/python3.10/dist-packages (from stanza) (2.1.0+cu118)

Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages (from stanza) (4.66.1)

Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-packages (from torch>=1.3.0->stanza) (3.13.1)

Requirement already satisfied: typing-extensions in /usr/local/lib/python3.10/dist-packages (from torch>=1.3.0->stanza) (4.5.0)

Requirement already satisfied: sympy in /usr/local/lib/python3.10/dist-packages (from torch>=1.3.0->stanza) (1.12)

Requirement already satisfied: jinja2 in /usr/local/lib/python3.10/dist-packages (from torch>=1.3.0->stanza) (3.1.2)

Requirement already satisfied: fsspec in /usr/local/lib/python3.10/dist-packages (from torch>=1.3.0->stanza) (2023.6.0)

Requirement already satisfied: triton==2.1.0 in /usr/local/lib/python3.10/dist-packages (from torch>=1.3.0->stanza) (2.1.0)

Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests->stanza) (3.3.2)

Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests->stanza) (3.6)

Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests->stanza) (2.0.7)

Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests->stanza) (2023.11.17)

Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.10/dist-packages (from jinja2->torch>=1.3.0->stanza) (2.1.3)

Requirement already satisfied: mpmath>=0.19 in /usr/local/lib/python3.10/dist-packages (from sympy->torch>=1.3.0->stanza) (1.3.0)

Installing collected packages: emoji, stanza

Successfully installed emoji-2.9.0 stanza-1.7.0

Downloading https://raw.githubusercontent.com/stanfordnlp/stanza-resources/main/resources_1.7.0.json: (https://raw.githubusercontent.com/stanfordnlp/stanza-resources/main/resources_1.7.0.json)

0%| ...

INFO:stanza:Downloading these customized packages for language: en (English)...

Processor	Package
tokenize	mimic
pos	mimic_charlm
lemma	mimic_nocharlm
depparse	mimic_charlm
ner	i2b2
pretrain	mimic
backward_charlm	mimic
forward_charlm	mimic

Downloading <https://huggingface.co/stanfordnlp/stanza-en/resolve/v1.7.0/models/tokenize/mimic.pt>: (<https://huggingface.co/stanfordnlp/stanza-en/resolve/v1.7.0/models/tokenize/mimic.pt>:) 0%|

...

Downloading https://huggingface.co/stanfordnlp/stanza-en/resolve/v1.7.0/models/pos/mimic_charlm.pt: (https://huggingface.co/stanfordnlp/stanza-en/resolve/v1.7.0/models/pos/mimic_charlm.pt:) 0%|

...

Downloading https://huggingface.co/stanfordnlp/stanza-en/resolve/v1.7.0/models/lemma/mimic_nocharlm.pt: (https://huggingface.co/stanfordnlp/stanza-en/resolve/v1.7.0/models/lemma/mimic_nocharlm.pt:) 0%| ...

Downloading https://huggingface.co/stanfordnlp/stanza-en/resolve/v1.7.0/models/depparse/mimic_charlm.pt: (https://huggingface.co/stanfordnlp/stanza-en/resolve/v1.7.0/models/depparse/mimic_charlm.pt:) 0%|...

Downloading <https://huggingface.co/stanfordnlp/stanza-en/resolve/v1.7.0/models/ner/i2b2.pt>: (<https://huggingface.co/stanfordnlp/stanza-en/resolve/v1.7.0/models/ner/i2b2.pt>:) 0%| | 0...

Downloading <https://huggingface.co/stanfordnlp/stanza-en/resolve/v1.7.0/models/pretrain/mimic.pt>: (<https://huggingface.co/stanfordnlp/stanza-en/resolve/v1.7.0/models/pretrain/mimic.pt>:) 0%|

...

Downloading https://huggingface.co/stanfordnlp/stanza-en/resolve/v1.7.0/models/backward_charlm/mimic.pt: (https://huggingface.co/stanfordnlp/stanza-en/resolve/v1.7.0/models/backward_charlm/mimic.pt:) 0%|...

Downloading https://huggingface.co/stanfordnlp/stanza-en/resolve/v1.7.0/models/forward_charlm/mimic.pt: (https://huggingface.co/stanfordnlp/stanza-en/resolve/v1.7.0/models/forward_charlm/mimic.pt:) 0%| ...

INFO:stanza:Finished downloading models and saved to /root/stanza_resources.

INFO:stanza:Checking for updates to resources.json in case models have been updated. Note: this behavior can be turned off with download_method=None or download_method=DownloadMethod.REUSE_RESOURCES

Downloading https://raw.githubusercontent.com/stanfordnlp/stanza-resources/main/resources_1.7.0.json: (https://raw.githubusercontent.com/stanfordnlp/stanza-resources/main/resources_1.7.0.json)

0%| ...

INFO:stanza:Loading these models for language: en (English):

=====		
Processor	Package	

tokenize	mimic	
pos	mimic_charlm	
lemma	mimic_nocharlm	
depparse	mimic_charlm	
ner	i2b2	
=====		

INFO:stanza:Using device: cuda

INFO:stanza:Loading: tokenize

INFO:stanza:Loading: pos

INFO:stanza:Loading: lemma

INFO:stanza:Loading: depparse

INFO:stanza:Loading: ner

INFO:stanza:Done loading processors!

```

In [21]: # Extract Symptom-related Entities in each symptom text
# Print out 50 reports to see the example of output (Not print all
extracted_symptom_list = []
count_symptom = {}
for i, sentence in enumerate(covid_reports['SYMPTOM_TEXT'][:50]):
    # annotate clinical text
    doc = nlp(sentence)

    # Iterate to extracted symptom
    for ent in doc.entities:
        if ent.type == 'PROBLEM':
            symptom = ent.text.lower()
            extracted_symptom_list.append(symptom)
            count_symptom[symptom] = count_symptom.get(symptom, 0)

    # Print out 50 reports
    print("Vaers ID:", covid_reports['VAERS_ID'][i])
    print("Input:", sentence)
    print("Output:")
    for ent in doc.entities:
        print(f'\t{ent.text}\t{ent.type}')
    print()

```

Vaers ID: 2669769

Input: body aches, fatigue Narrative: Took OTC Tylenol Other
Relevant History:

Output:

body aches	PROBLEM
fatigue	PROBLEM
OTC Tylenol	TREATMENT

Vaers ID: 2527460

Input: Headache, Myalgia, NauseaVomiting, chills Narrative:

Output:

Headache	PROBLEM
Myalgia	PROBLEM
NauseaVomiting	PROBLEM
chills	PROBLEM

Vaers ID: 2673135

Input: Headache, Fever, Body aches Narrative: Other Relevant
History:

Output:

In [22]: `extracted_symptom_list`

```
Out[22]: ['body aches',
          'fatigue',
          'headache',
          'myalgia',
          'nauseavomiting',
          'chills',
          'headache',
          'fever',
          'body aches',
          'headache',
          'myalgia',
          'mild numbness',
          'injection site',
          'headache',
          'warm',
          'progressive light-headedness',
          'near-syncope',
          'diaphoresis',
          '20 minutes symptoms',
          'bad taste in mouth',
          'tingling in body , legs, back , across stomach',
          'symptoms',
          'shaking of hands',
          'tingling',
          'tongue tingling',
          'smelling chemical smell',
          'mouth itching',
          'symptoms',
          'lightheadedness',
          'flushing',
          'symptoms',
          'vagal and "spacey',
          'chest heaviness',
          'shortness of breath',
          'rigors',
          'urge to defecate',
          'chest pain',
          'really bad heartburn',
          'headache',
          'body ache',
          'lightheadedness',
          'shortness of breath',
          'headache',
          'some nausea',
          'some redness to her neck and upper chest',
          'recent illness',
          'elevated bps',
          'dizziness',
          'being hot',
          'nauseated',
          'difficulty breathing',
          'chest pain',
          'nauseated',
          'dizzy',
          'lightheaded',
          'feet elevated',
          'any other complaints',
          'complaints',
          'facial flushing',
          'pounding in chest',
          'burning and hot ears',
```


'tingly in right arm and chest and hands',
'symptoms',
'this reaction',
'tightness',
'her tongue',
'tingling',
'anthrax',
'tingling of upper lip and cheeks',
'warmth in face',
'symptom',
'symptoms',
'shaky',
'shaking',
'short of breath',
'wheeze',
'tightness in the chest or throat',
'pulse weak',
'sweating on torso',
'red',
'the first sob',
'chills',
'nausea',
'vomiting',
'potential fever',
'symptoms',
'pain in deltoid muscle',
'pain in the arm',
'headache',
'nausea',
'energy',
'injection site pain',
'a fever',
'symptoms',
'right sided facial/lip swelling',
'angioedema',
'chills',
'hot',
'cold',
'nausea',
'headache',
'extreme fatigue',
'low grade temp',
'associate vomited',
'vomiting',
'some nausea',
'diffuse rash at anterior and right lateral neck',
'warmth',
'woozy',
'dizzy',
'dizzy',
'chills',
'body ache',
'headache',
'headache',
'sore throat',
'headache',
'lightheadedness',
'lightheaded',
'clammy',
'dizzy',
'some chest tightness',

'bilateral hand tingling',
'deep breath',
'rash',
'swelling',
'itching',
'throat tightness',
'a panic attack"',
'chest tightness',
'redness',
'the injection spot',
'fever',
'chills',
'stomach ache',
'body ache',
'short of breath',
'headaches',
'appetite',
'localized reaction in left deltoid',
'redness',
'firm to touch',
'redness',
'chills',
'body aches',
'feverish',
'a little soreness',
'other symptoms',
'severely nausead',
'lightheaded',
'really loopy in the head',
'the dizziness',
'the nausea',
'being "lightheaded',
'slightly dizzy',
'patient symptoms',
'left arm swelling of forearm',
'old l wrist tattoo',
'itching',
'several red dots',
'reaction',
'body wide itching',
'nausea',
'symptoms',
'a prior anaphylaxis reaction',
'sweaty palms',
'lightheadedness',
'throat swelling',
'difficulty breathing',
'the symptoms',
'pain',
'pain',
'the pain',
'nauseous',
'body aches',
'chills',
'really tired',
'a dull headache',
'crummy',
'fever',
'fatigued',
'a headache',
'the pain',

```
'the headache',  
'normal arm pain',  
'itchy flushed skin',  
'a rash',  
'the itchy skin',  
'a sore arm',  
'swelling on arm',  
'a metallic taste in her mouth',  
'nauseous',  
'myalgias',  
'nauseas',  
'headache',  
'stomach pain',  
'pain',  
'edema',  
'decreased range of motion',  
'left arm rib cage pain',  
'muscle pain right and left thighs',  
'worsening upon ambulation muscle pain',  
'tingling right upper extremity',  
'hypotensive',  
'pale',  
'diaphoretic',  
'syncope']
```

In [23]: count_symptom

```
Out[23]: {'body aches': 4,
          'fatigue': 1,
          'headache': 12,
          'myalgia': 2,
          'nauseavomiting': 1,
          'chills': 7,
          'fever': 3,
          'mild numbness': 1,
          'injection site': 1,
          'warm': 1,
          'progressive light-headedness': 1,
          'near-syncope': 1,
          'diaphoresis': 1,
          '20 minutes symptoms': 1,
          'bad taste in mouth': 1,
          'tingling in body , legs, back , across stomach': 1,
          'symptoms': 8,
          'shaking of hands': 1,
          'tingling': 2,
          'tongue tingling': 1,
          'smelling chemical smell': 1,
          'mouth itching': 1,
          'lightheadedness': 4,
          'flushing': 1,
          'vagal and "spacey': 1,
          'chest heaviness': 1,
          'shortness of breath': 2,
          'rigors': 1,
          'urge to defecate': 1,
          'chest pain': 2,
          'really bad heartburn': 1,
          'body ache': 3,
          'some nausea': 2,
          'some redness to her neck and upper chest': 1,
          'recent illness': 1,
          'elevated bps': 1,
          'dizziness': 1,
          'being hot': 1,
          'nauseated': 2,
          'difficulty breathing': 2,
          'dizzy': 4,
          'lightheaded': 3,
          'feet elevated': 1,
          'any other complaints': 1,
          'complaints': 1,
          'facial flushing': 1,
          'pounding in chest': 1,
          'burning and hot ears': 1,
          'tingly in right arm and chest and hands': 1,
          'this reaction': 1,
          'tightness': 1,
          'her tongue': 1,
          'anthrax': 1,
          'tingling of upper lip and cheeks': 1,
          'warmth in face': 1,
          'symptom': 1,
          'shaky': 1,
          'shaking': 1,
          'short of breath': 2,
          'wheeze': 1,
          'tightness in the chest or throat': 1,
```

```
'pulse weak': 1,  
'sweating on torso': 1,  
'red': 1,  
'the first sob': 1,  
'nausea': 4,  
'vomiting': 2,  
'potential fever': 1,  
'pain in deltoid muscle': 1,  
'pain in the arm': 1,  
'energy': 1,  
'injection site pain': 1,  
'a fever': 1,  
'right sided facial/lip swelling': 1,  
'angioedema': 1,  
'hot': 1,  
'cold': 1,  
'extreme fatigue': 1,  
'low grade temp': 1,  
'associate vomited': 1,  
'diffuse rash at anterior and right lateral neck': 1,  
'warmth': 1,  
'woozy': 1,  
'sore throat': 1,  
'clammy': 1,  
'some chest tightness': 1,  
'bilateral hand tingling': 1,  
'deep breath': 1,  
'rash': 1,  
'swelling': 1,  
'itching': 2,  
'throat tightness': 1,  
'a panic attack': 1,  
'chest tightness': 1,  
'redness': 3,  
'the injection spot': 1,  
'stomach ache': 1,  
'headaches': 1,  
'appetite': 1,  
'localized reaction in left deltoid': 1,  
'firm to touch': 1,  
'feverish': 1,  
'a little soreness': 1,  
'other symptoms': 1,  
'severely nauseated': 1,  
'really loopy in the head': 1,  
'the dizziness': 1,  
'the nausea': 1,  
'being "lightheaded': 1,  
'slightly dizzy': 1,  
'patient symptoms': 1,  
'left arm swelling of forearm': 1,  
'old l wrist tattoo': 1,  
'several red dots': 1,  
'reaction': 1,  
'body wide itching': 1,  
'a prior anaphylaxis reaction': 1,  
'sweaty palms': 1,  
'throat swelling': 1,  
'the symptoms': 1,  
'pain': 3,  
'the pain': 2,
```

```
'nauseous': 2,  
'really tired': 1,  
'a dull headache': 1,  
'crummy': 1,  
'fatigued': 1,  
'a headache': 1,  
'the headache': 1,  
'normal arm pain': 1,  
'itchy flushed skin': 1,  
'a rash': 1,  
'the itchy skin': 1,  
'a sore arm': 1,  
'swelling on arm': 1,  
'a metallic taste in her mouth': 1,  
'myalgias': 1,  
'nauseas': 1,  
'stomach pain': 1,  
'edema': 1,  
'decreased range of motion': 1,  
'left arm rib cage pain': 1,  
'muscle pain right and left thighs': 1,  
'worsening upon ambulation muscle pain': 1,  
'tingling right upper extremity': 1,  
'hypotensive': 1,  
'pale': 1,  
'diaphoretic': 1,  
'syncope': 1}
```

4. Link Entities to Standard Symptoms

I will link 50 reports to the standard symptom to see the performance of the model, using different methods.

4.1 Exact matching

```
In [24]: # Define the dictionary for mapping
exact_linked_std_symp_dict = {}
exact_linked_top_symp_dict = {}

# Iterate to find the exact match symptom
for symp in extracted_symptom_list:

    # Define as None at first
    exact_linked_std_symp_dict[symp.lower()] = 'None'
    exact_linked_top_symp_dict[symp.lower()] = 'None'

    # Iterate to find the exact match symptom from the list of sta
    for std_symptom in std_symp:
        if symp.lower() == std_symptom.lower():
            exact_linked_std_symp_dict[symp.lower()] = std_symptom
            break

    # Iterate to find the exact match symptom from the list of top
    for top_symptom in top_symp:
        if symp.lower() == top_symptom.lower():
            exact_linked_top_symp_dict[symp.lower()] = top_symptom
            break
```


In [25]: `exact_linked_std_symp_dict`

```
Out[25]: {'body aches': 'None',
'fatigue': 'fatigue',
'headache': 'headache',
'myalgia': 'myalgia',
'nauseavomiting': 'None',
'chills': 'chills',
'fever': 'None',
'mild numbness': 'None',
'injection site': 'None',
'warm': 'None',
'progressive light-headedness': 'None',
'near-syncope': 'None',
'diaphoresis': 'None',
'20 minutes symptoms': 'None',
'bad taste in mouth': 'None',
'tingling in body , legs, back , across stomach': 'None',
'symptoms': 'None',
'shaking of hands': 'None',
'tingling': 'None',
'tongue tingling': 'None',
'smelling chemical smell': 'None',
'mouth itching': 'None',
'lightheadedness': 'None',
'flushing': 'flushing',
'vagal and "spacey": 'None',
'chest heaviness': 'None',
'shortness of breath': 'None',
'rigors': 'None',
'urge to defecate': 'None',
'chest pain': 'chest pain',
'really bad heartburn': 'None',
'body ache': 'None',
'some nausea': 'None',
'some redness to her neck and upper chest': 'None',
'recent illness': 'None',
'elevated bps': 'None',
'dizziness': 'dizziness',
'being hot': 'None',
'nauseated': 'None',
'difficulty breathing': 'None',
'dizzy': 'None',
'lightheaded': 'None',
'feet elevated': 'None',
'any other complaints': 'None',
'complaints': 'None',
'facial flushing': 'None',
'pounding in chest': 'None',
'burning and hot ears': 'None',
'tingly in right arm and chest and hands': 'None',
'this reaction': 'None',
'tightness': 'None',
'her tongue': 'None',
'anthrax': 'None',
'tingling of upper lip and cheeks': 'None',
'warmth in face': 'None',
'symptom': 'None',
'shaky': 'None',
'shaking': 'None',
'short of breath': 'None',
'wheeze': 'None',
'tightness in the chest or throat': 'None',
```

```
'pulse weak': 'None',
'sweating on torso': 'None',
'red': 'None',
'the first sob': 'None',
'nausea': 'nausea',
'vomiting': 'vomiting',
'potential fever': 'None',
'pain in deltoid muscle': 'None',
'pain in the arm': 'None',
'energy': 'None',
'injection site pain': 'injection site pain',
'a fever': 'None',
'right sided facial/lip swelling': 'None',
'angioedema': 'angioedema',
'hot': 'None',
'cold': 'None',
'extreme fatigue': 'None',
'low grade temp': 'None',
'associate vomited': 'None',
'diffuse rash at anterior and right lateral neck': 'None',
'warmth': 'None',
'woozy': 'None',
'sore throat': 'None',
'clammy': 'None',
'some chest tightness': 'None',
'bilateral hand tingling': 'None',
'deep breath': 'None',
'rash': 'rash',
'swelling': 'swelling',
'itching': 'None',
'throat tightness': 'throat tightness',
'a panic attack": 'None',
'chest tightness': 'None',
'redness': 'None',
'the injection spot': 'None',
'stomach ache': 'None',
'headaches': 'None',
'appetite': 'None',
'localized reaction in left deltoid': 'None',
'firm to touch': 'None',
'feverish': 'None',
'a little soreness': 'None',
'other symptoms': 'None',
'severely nausead': 'None',
'really loopy in the head': 'None',
'the dizziness': 'None',
'the nausea': 'None',
'being "lightheaded': 'None',
'slightly dizzy': 'None',
'patient symptoms': 'None',
'left arm swelling of forearm': 'None',
'old l wrist tattoo': 'None',
'several red dots': 'None',
'reaction': 'None',
'body wide itching': 'None',
'a prior anaphylaxis reaction': 'None',
'sweaty palms': 'None',
'throat swelling': 'None',
'the symptoms': 'None',
'pain': 'pain',
'the pain': 'None',
```

```
'nauseous': 'None',  
'really tired': 'None',  
'a dull headache': 'None',  
'crummy': 'None',  
'fatigued': 'None',  
'a headache': 'None',  
'the headache': 'None',  
'normal arm pain': 'None',  
'itchy flushed skin': 'None',  
'a rash': 'None',  
'the itchy skin': 'None',  
'a sore arm': 'None',  
'swelling on arm': 'None',  
'a metallic taste in her mouth': 'None',  
'myalgias': 'None',  
'nauseas': 'None',  
'stomach pain': 'None',  
'edema': 'None',  
'decreased range of motion': 'None',  
'left arm rib cage pain': 'None',  
'muscle pain right and left thighs': 'None',  
'worsening upon ambulation muscle pain': 'None',  
'tingling right upper extremity': 'None',  
'hypotensive': 'None',  
'pale': 'None',  
'diaphoretic': 'None',  
'syncope': 'syncope'}
```

In [26]: `exact_linked_top_symp_dict`

```
Out[26]: {'body aches': 'None',
'fatigue': 'fatigue',
'headache': 'headache',
'myalgia': 'myalgia',
'nauseavomiting': 'None',
'chills': 'chills',
'fever': 'None',
'mild numbness': 'None',
'injection site': 'None',
'warm': 'None',
'progressive light-headedness': 'None',
'near-syncope': 'None',
'diaphoresis': 'None',
'20 minutes symptoms': 'None',
'bad taste in mouth': 'None',
'tingling in body , legs, back , across stomach': 'None',
'symptoms': 'None',
'shaking of hands': 'None',
'tingling': 'None',
'tongue tingling': 'None',
'smelling chemical smell': 'None',
'mouth itching': 'None',
'lightheadedness': 'None',
'flushing': 'flushing',
'vagal and "spacey": 'None',
'chest heaviness': 'None',
'shortness of breath': 'None',
'rigors': 'None',
'urge to defecate': 'None',
'chest pain': 'chest pain',
'really bad heartburn': 'None',
'body ache': 'None',
'some nausea': 'None',
'some redness to her neck and upper chest': 'None',
'recent illness': 'None',
'elevated bps': 'None',
'dizziness': 'dizziness',
'being hot': 'None',
'nauseated': 'None',
'difficulty breathing': 'None',
'dizzy': 'None',
'lightheaded': 'None',
'feet elevated': 'None',
'any other complaints': 'None',
'complaints': 'None',
'facial flushing': 'None',
'pounding in chest': 'None',
'burning and hot ears': 'None',
'tingly in right arm and chest and hands': 'None',
'this reaction': 'None',
'tightness': 'None',
'her tongue': 'None',
'anthrax': 'None',
'tingling of upper lip and cheeks': 'None',
'warmth in face': 'None',
'symptom': 'None',
'shaky': 'None',
'shaking': 'None',
'short of breath': 'None',
'wheeze': 'None',
'tightness in the chest or throat': 'None',
```

```
'pulse weak': 'None',
'sweating on torso': 'None',
'red': 'None',
'the first sob': 'None',
'nausea': 'nausea',
'vomiting': 'vomiting',
'potential fever': 'None',
'pain in deltoid muscle': 'None',
'pain in the arm': 'None',
'energy': 'None',
'injection site pain': 'injection site pain',
'a fever': 'None',
'right sided facial/lip swelling': 'None',
'angioedema': 'None',
'hot': 'None',
'cold': 'None',
'extreme fatigue': 'None',
'low grade temp': 'None',
'associate vomited': 'None',
'diffuse rash at anterior and right lateral neck': 'None',
'warmth': 'None',
'woozy': 'None',
'sore throat': 'None',
'clammy': 'None',
'some chest tightness': 'None',
'bilateral hand tingling': 'None',
'deep breath': 'None',
'rash': 'rash',
'swelling': 'swelling',
'itching': 'None',
'throat tightness': 'throat tightness',
'a panic attack": 'None',
'chest tightness': 'None',
'redness': 'None',
'the injection spot': 'None',
'stomach ache': 'None',
'headaches': 'None',
'appetite': 'None',
'localized reaction in left deltoid': 'None',
'firm to touch': 'None',
'feverish': 'None',
'a little soreness': 'None',
'other symptoms': 'None',
'severely nausead': 'None',
'really loopy in the head': 'None',
'the dizziness': 'None',
'the nausea': 'None',
'being "lightheaded': 'None',
'slightly dizzy': 'None',
'patient symptoms': 'None',
'left arm swelling of forearm': 'None',
'old l wrist tattoo': 'None',
'several red dots': 'None',
'reaction': 'None',
'body wide itching': 'None',
'a prior anaphylaxis reaction': 'None',
'sweaty palms': 'None',
'throat swelling': 'None',
'the symptoms': 'None',
'pain': 'pain',
'the pain': 'None',
```

```
'nauseous': 'None',
'really tired': 'None',
'a dull headache': 'None',
'crummy': 'None',
'fatigued': 'None',
'a headache': 'None',
'the headache': 'None',
'normal arm pain': 'None',
'itchy flushed skin': 'None',
'a rash': 'None',
'the itchy skin': 'None',
'a sore arm': 'None',
'swelling on arm': 'None',
'a metallic taste in her mouth': 'None',
'myalgias': 'None',
'nauseas': 'None',
'stomach pain': 'None',
'edema': 'None',
'decreased range of motion': 'None',
'left arm rib cage pain': 'None',
'muscle pain right and left thighs': 'None',
'worsening upon ambulation muscle pain': 'None',
'tingling right upper extremity': 'None',
'hypotensive': 'None',
'pale': 'None',
'diaphoretic': 'None',
'syncope': 'syncope'}
```

4.2 Fuzzy matching

In [27]: `# Install fuzzywuzzy`
`!pip install fuzzywuzzy`

```
Collecting fuzzywuzzy
  Downloading fuzzywuzzy-0.18.0-py2.py3-none-any.whl (18 kB)
Installing collected packages: fuzzywuzzy
Successfully installed fuzzywuzzy-0.18.0
```

In [28]: `# Import the libraries for fuzzywuzzy`
`from fuzzywuzzy import fuzz`

```
/usr/local/lib/python3.10/dist-packages/fuzzywuzzy/fuzz.py:11: UserWarning: Using slow pure-python SequenceMatcher. Install python-Levenshtein to remove this warning
  warnings.warn('Using slow pure-python SequenceMatcher. Install python-Levenshtein to remove this warning')
```



```
In [29]: # Define the dictionary for mapping
fuzzy_linked_std_symp_dict = {}
fuzzy_linked_top_symp_dict = {}

# Iterate to find the most similar symptom
for symp in extracted_symptom_list:

    # Define as None at first
    fuzzy_linked_std_symp_dict[symp.lower()] = 'None'
    fuzzy_linked_top_symp_dict[symp.lower()] = 'None'

    # Iterate to find the most similar symptom from the list of std
    max_fuzz_ratio = 0.0
    for std_symptom in std_symp:
        fuzz_ratio = fuzz.partial_ratio(symp.lower(), std_symptom)
        if fuzz_ratio > max_fuzz_ratio:
            max_fuzz_ratio = fuzz_ratio
            most_sim_std_symp = std_symptom.lower()
    fuzzy_linked_std_symp_dict[symp.lower()] = most_sim_std_symp

    # Iterate to find the most similar symptom from the list of top
    max_fuzz_ratio = 0.0
    for top_symptom in top_symp:
        fuzz_ratio = fuzz.partial_ratio(symp.lower(), top_symptom)
        if fuzz_ratio > max_fuzz_ratio:
            max_fuzz_ratio = fuzz_ratio
            most_sim_top_symp = top_symptom.lower()
    fuzzy_linked_top_symp_dict[symp.lower()] = most_sim_top_symp
```

```
In [30]: fuzzy_linked_std_symp_dict
```

```

Out[30]: {'body aches': 'acne',
'fatigue': 'fatigue',
'headache': 'headache',
'myalgia': 'myalgia',
'nauseavomiting': 'nausea',
'chills': 'chills',
'fever': 'thyroxine free',
'mild numbness': 'blindness',
'injection site': 'injection site hypoaesthesia',
'warm': 'skin warm',
'progressive light-headedness': 'deafness',
'near-syncope': 'syncope',
'diaphoresis': 'hemiparesis',
'20 minutes symptoms': 'stress',
'bad taste in mouth': 'gout',
'tingling in body , legs, back , across stomach': 'back pain',
'symptoms': 'neurological symptom',
'shaking of hands': 'pain',
'tingling': 'blood creatinine',
'tongue tingling': 'tongue biting',
'smelling chemical smell': 'swelling',
'mouth itching': 'choking',
'lightheadedness': 'deafness',
'flushing': 'flushing',
'vagal and "spacey': 'acne',
'chest heaviness': 'illness',
'shortness of breath': 'death',
'rigors': 'vertigo',
'urge to defecate': 'fear',
'chest pain': 'pain',
'really bad heartburn': 'fall',
'body ache': 'acne',
'some nausea': 'nausea',
'some redness to her neck and upper chest': 'neck pain',
'recent illness': 'illness',
'elevated bps': 'biopsy',
'dizziness': 'dizziness',
'being hot': 'feeling hot',
'nauseated': 'nausea',
'difficulty breathing': 'retching',
'dizzy': 'dizziness',
'lightheaded': 'headache',
'feet elevated': 'death',
'any other complaints': 'pain',
'complaints': 'pain',
'facial flushing': 'flushing',
'pounding in chest': 'moaning',
'burning and hot ears': 'fear',
'tingly in right arm and chest and hands': 'chest pain',
'this reaction': 'anaphylactic reaction',
'tightness': 'throat tightness',
'her tongue': 'swollen tongue',
'anthrax': 'arthralgia',
'tingling of upper lip and cheeks': 'lip pain',
'warmth in face': 'acne',
'symptom': 'neurological symptom',
'shaky': 'oropharyngeal pain',
'shaking': 'flushing',
'short of breath': 'death',
'wheeze': 'wheezing',
'tightness in the chest or throat': 'dry throat',

```

```
'pulse weak': 'injection site muscle weakness',  
'sweating on torso': 'pain',  
'red': 'product administered at inappropriate site',  
'the first sob': 'thirst',  
'nausea': 'nausea',  
'vomiting': 'vomiting',  
'potential fever': 'hypotonia',  
'pain in deltoid muscle': 'pain',  
'pain in the arm': 'pain',  
'energy': 'energy increased',  
'injection site pain': 'pain',  
'a fever': 'dyspnoea exertional',  
'right sided facial/lip swelling': 'lip swelling',  
'angioedema': 'angioedema',  
'hot': 'feeling hot',  
'cold': 'cold sweat',  
'extreme fatigue': 'fatigue',  
'low grade temp': 'exposure to extreme temperature',  
'associate vomited': 'injury associated with device',  
'diffuse rash at anterior and right lateral neck': 'rash',  
'warmth': 'injection site warmth',  
'woozy': 'haematology test normal',  
'sore throat': 'dry throat',  
'clammy': 'injection site inflammation',  
'some chest tightness': 'skin tightness',  
'bilateral hand tingling': 'blister',  
'deep breath': 'death',  
'rash': 'rash',  
'swelling': 'lip swelling',  
'itching': 'muscle twitching',  
'throat tightness': 'throat tightness',  
'a panic attack': 'panic attack',  
'chest tightness': 'muscle tightness',  
'redness': 'tenderness',  
'the injection spot': 'ear infection',  
'stomach ache': 'acne',  
'headaches': 'headache',  
'appetite': 'decreased appetite',  
'localized reaction in left deltoid': 'localised infection',  
'firm to touch': 'accidental exposure to product',  
'feverish': 'rash',  
'a little soreness': 'stress',  
'other symptoms': 'allergic respiratory symptom',  
'severely nauseated': 'nausea',  
'really loopy in the head': 'fall',  
'the dizziness': 'dizziness',  
'the nausea': 'nausea',  
'being "lightheaded": 'headache',  
'slightly dizzy': 'dizziness',  
'patient symptoms': 'pain',  
'left arm swelling of forearm': 'swelling',  
'old l wrist tattoo': 'atopy',  
'several red dots': 'viral titre',  
'reaction': 'local reaction',  
'body wide itching': 'choking',  
'a prior anaphylaxis reaction': 'anaphylactic reaction',  
'sweaty palms': 'night sweats',  
'throat swelling': 'swelling',  
'the symptoms': 'allergic respiratory symptom',  
'pain': 'pain',  
'the pain': 'pain',
```

```
'nauseous': 'nausea',  
'really tired': 'fall',  
'a dull headache': 'headache',  
'crummy': 'circumstance or information capable of leading to med  
ication error',  
'fatigued': 'fatigue',  
'a headache': 'headache',  
'the headache': 'headache',  
'normal arm pain': 'pain',  
'itchy flushed skin': 'dry skin',  
'a rash': 'rash',  
'the itchy skin': 'dry skin',  
'a sore arm': 'fear',  
'swelling on arm': 'swelling',  
'a metallic taste in her mouth': 'dry mouth',  
'myalgias': 'myalgia',  
'nauseas': 'nausea',  
'stomach pain': 'pain',  
'edema': 'injection site oedema',  
'decreased range of motion': 'drainage',  
'left arm rib cage pain': 'pain',  
'muscle pain right and left thighs': 'pain',  
'worsening upon ambulation muscle pain': 'pain',  
'tingling right upper extremity': 'tremor',  
'hypotensive': 'hypotension',  
'pale': 'faeces pale',  
'diaphoretic': 'diarrhoea',  
'syncope': 'presyncope'}
```

```
In [31]: fuzzy_linked_top_symp_dict
```

```

Out[31]: {'body aches': 'headache',
'fatigue': 'fatigue',
'headache': 'headache',
'myalgia': 'myalgia',
'nauseavomiting': 'nausea',
'chills': 'chills',
'fever': 'injection site erythema',
'mild numbness': 'influenza like illness',
'injection site': 'injection site pain',
'warm': 'injection site warmth',
'progressive light-headedness': 'headache',
'near-syncope': 'syncope',
'diaphoresis': 'diarrhoea',
'20 minutes symptoms': 'dizziness',
'bad taste in mouth': 'dry mouth',
'tingling in body , legs, back , across stomach': 'back pain',
'symptoms': 'rash erythematous',
'shaking of hands': 'pain',
'tingling': 'vomiting',
'tongue tingling': 'fatigue',
'smelling chemical smell': 'swelling',
'mouth itching': 'vomiting',
'lightheadedness': 'throat tightness',
'flushing': 'flushing',
'vagal and "spacey': 'headache',
'chest heaviness': 'pain',
'shortness of breath': 'asthenia',
'rigors': 'rhinorrhoea',
'urge to defecate': 'fatigue',
'chest pain': 'pain',
'really bad heartburn': 'rash',
'body ache': 'body temperature increased',
'some nausea': 'nausea',
'some redness to her neck and upper chest': 'neck pain',
'recent illness': 'chills',
'elevated bps': 'immediate post-injection reaction',
'dizziness': 'dizziness',
'being hot': 'feeling hot',
'nauseated': 'nausea',
'difficulty breathing': 'fatigue',
'dizzy': 'dizziness',
'lightheaded': 'headache',
'feet elevated': 'sars-cov-2 test negative',
'any other complaints': 'pain',
'complaints': 'pain',
'facial flushing': 'flushing',
'pounding in chest': 'chest pain',
'burning and hot ears': 'feeling hot',
'tingly in right arm and chest and hands': 'chest pain',
'this reaction': 'anaphylactic reaction',
'tightness': 'throat tightness',
'her tongue': 'swollen tongue',
'anthrax': 'arthralgia',
'tingling of upper lip and cheeks': 'headache',
'warmth in face': 'injection site warmth',
'symptom': 'dyspnoea',
'shaky': 'oropharyngeal pain',
'shaking': 'flushing',
'short of breath': 'hot flush',
'wheeze': 'headache',
'tightness in the chest or throat': 'chest x-ray',

```

```
'pulse weak': 'musculoskeletal stiffness',  
'sweating on torso': 'pain',  
'red': 'impaired work ability',  
'the first sob': 'chest discomfort',  
'nausea': 'nausea',  
'vomiting': 'vomiting',  
'potential fever': 'asthenia',  
'pain in deltoid muscle': 'pain',  
'pain in the arm': 'pain',  
'energy': 'injection site erythema',  
'injection site pain': 'pain',  
'a fever': 'body temperature increased',  
'right sided facial/lip swelling': 'lip swelling',  
'angioedema': 'erythema',  
'hot': 'feeling hot',  
'cold': 'cold sweat',  
'extreme fatigue': 'fatigue',  
'low grade temp': 'tremor',  
'associate vomited': 'vomiting',  
'diffuse rash at anterior and right lateral neck': 'rash',  
'warmth': 'injection site warmth',  
'woozy': 'chest discomfort',  
'sore throat': 'blood pressure increased',  
'clammy': 'electrocardiogram',  
'some chest tightness': 'throat tightness',  
'bilateral hand tingling': 'arthralgia',  
'deep breath': 'lymphadenopathy',  
'rash': 'rash',  
'swelling': 'injection site swelling',  
'itching': 'vomiting',  
'throat tightness': 'throat tightness',  
'a panic attack': 'pain',  
'chest tightness': 'throat tightness',  
'redness': 'dizziness',  
'the injection spot': 'immediate post-injection reaction',  
'stomach ache': 'headache',  
'headaches': 'headache',  
'appetite': 'decreased appetite',  
'localized reaction in left deltoid': 'pain in extremity',  
'firm to touch': 'cough',  
'feverish': 'rash',  
'a little soreness': 'influenza like illness',  
'other symptoms': 'rash erythematous',  
'severely nauseated': 'nausea',  
'really loopy in the head': 'headache',  
'the dizziness': 'dizziness',  
'the nausea': 'nausea',  
'being "lightheaded": 'headache',  
'slightly dizzy': 'dizziness',  
'patient symptoms': 'pain',  
'left arm swelling of forearm': 'swelling',  
'old l wrist tattoo': 'rash',  
'several red dots': 'dry mouth',  
'reaction': 'immediate post-injection reaction',  
'body wide itching': 'vomiting',  
'a prior anaphylaxis reaction': 'anaphylactic reaction',  
'sweaty palms': 'muscle spasms',  
'throat swelling': 'swelling',  
'the symptoms': 'rash erythematous',  
'pain': 'pain',  
'the pain': 'pain',
```



```
'nauseous': 'nausea',  
'really tired': 'musculoskeletal stiffness',  
'a dull headache': 'headache',  
'crummy': 'pain in extremity',  
'fatigued': 'fatigue',  
'a headache': 'headache',  
'the headache': 'headache',  
'normal arm pain': 'pain',  
'itchy flushed skin': 'hot flush',  
'a rash': 'rash',  
'the itchy skin': 'skin warm',  
'a sore arm': 'injection site warmth',  
'swelling on arm': 'swelling',  
'a metallic taste in her mouth': 'dry mouth',  
'myalgias': 'myalgia',  
'nauseas': 'nausea',  
'stomach pain': 'pain',  
'edema': 'headache',  
'decreased range of motion': 'decreased appetite',  
'left arm rib cage pain': 'pain',  
'muscle pain right and left thighs': 'pain',  
'worsening upon ambulation muscle pain': 'pain',  
'tingling right upper extremity': 'tremor',  
'hypotensive': 'hypertension',  
'pale': 'hypoesthesia',  
'diaphoretic': 'diarrhoea',  
'syncope': 'syncope'}
```



```

In [32]: from gensim.models import KeyedVectors

# Load GloVe embeddings
def load_glove_embeddings(file_path):

    embeddings_index = {}
    with open(file_path, 'r', encoding='utf-8') as f:
        for line in f:
            values = line.split()
            word = values[0]
            coefs = np.asarray(values[1:], dtype='float32')
            embeddings_index[word] = coefs
    return embeddings_index

# Get embeddings
def get_embedding(word, embeddings_index, default_vector=None):

    try:
        return embeddings_index[word]
    except KeyError:
        return default_vector

# Specify the path to the GloVe file
glove_file_path = '/content/drive/My Drive/Colab Notebooks/convert

# Load GloVe embeddings
glove_embeddings = KeyedVectors.load_word2vec_format(glove_file_pa

# Define the dictionary for mapping
sim_linked_std_symp_dict = {}
sim_linked_top_symp_dict = {}

# Define a default vector (vector of zeros)
default_vector = np.zeros_like(glove_embeddings['headache'])

for symp in extracted_symptom_list:

    symp_vec = get_embedding(symp.lower(), glove_embeddings, defau

# Find the most similar symptoms from the list of standard sym
max_sim_score = 0.0
most_symp = 'None'
for std_symp in std_symp:

    std_sympom_vec = get_embedding(std_sympom.lower(), glove
    sim_score = np.dot(symp_vec, std_sympom_vec) / (np.linalg
    if sim_score > max_sim_score:
        max_sim_score = sim_score
        most_symp = std_sympom

sim_linked_std_symp_dict[symp] = most_symp

# Find the most similar symptoms from the list of the most com
max_sim_score = 0.0
most_symp = 'None'
for top_sympom in top_symp:

    top_sympom_vec = get_embedding(top_sympom.lower(), glove
    sim_score = np.dot(symp_vec, top_sympom_vec) / (np.linalg
    if sim_score > max_sim_score:
        max_sim_score = sim_score

```

```
most_symp = top_syptom
```

```
sim_linked_top_symp_dict[symp] = most_symp
```

```
<ipython-input-32-023315687be3>:46: RuntimeWarning: invalid value encountered in float_scalars
```

```
    sim_score = np.dot(symp_vec, std_syptom_vec) / (np.linalg.norm(symp_vec) * np.linalg.norm(std_syptom_vec))
```

```
<ipython-input-32-023315687be3>:59: RuntimeWarning: invalid value encountered in float_scalars
```

```
    sim_score = np.dot(symp_vec, top_syptom_vec) / (np.linalg.norm(symp_vec) * np.linalg.norm(top_syptom_vec))
```

```
In [33]: sim_linked_std_symp_dict
```

```
Out[33]: {'body aches': 'None',
'fatigue': 'Fatigue',
'headache': 'Headache',
'myalgia': 'Myalgia',
'nauseavomiting': 'None',
'chills': 'Chills',
'fever': 'Cough',
'mild numbness': 'None',
'injection site': 'None',
'warm': 'Flushing',
'progressive light-headedness': 'None',
'near-syncope': 'Glossodynia',
'diaphoresis': 'Paraesthesia',
'20 minutes symptoms': 'None',
'bad taste in mouth': 'None',
'tingling in body , legs, back , across stomach': 'None',
'symptoms': 'Pain',
'shaking of hands': 'None',
'tingling': 'Paraesthesia',
'tongue tingling': 'None',
'smelling chemical smell': 'None',
'mouth itching': 'None',
'lightheadedness': 'Palpitations',
'flushing': 'Flushing',
'vagal and "spacey': 'None',
'chest heaviness': 'None',
'shortness of breath': 'None',
'rigors': 'Chills',
'urge to defecate': 'None',
'chest pain': 'None',
'really bad heartburn': 'None',
'body ache': 'None',
'some nausea': 'None',
'some redness to her neck and upper chest': 'None',
'recent illness': 'None',
'elevated bps': 'None',
'dizziness': 'Dizziness',
'being hot': 'None',
'nauseated': 'Bedridden',
'difficulty breathing': 'None',
'dizzy': 'Dizziness',
'lightheaded': 'Eructation',
'feet elevated': 'None',
'any other complaints': 'None',
'complaints': 'Pain',
'facial flushing': 'None',
'pounding in chest': 'None',
'burning and hot ears': 'None',
'tingly in right arm and chest and hands': 'None',
'this reaction': 'None',
'tightness': 'Tenderness',
'her tongue': 'None',
'anthrax': 'Gastroenteritis',
'tingling of upper lip and cheeks': 'None',
'warmth in face': 'None',
'symptom': 'Pain',
'shaky': 'Audiogram',
'shaking': 'Agitation',
'short of breath': 'None',
'wheeze': 'Wheezing',
'tightness in the chest or throat': 'None',
```

```
'pulse weak': 'None',
'sweating on torso': 'None',
'red': 'Mass',
'the first sob': 'None',
'nausea': 'Nausea',
'vomiting': 'Vomiting',
'potential fever': 'None',
'pain in deltoid muscle': 'None',
'pain in the arm': 'None',
'energy': 'Mass',
'injection site pain': 'None',
'a fever': 'None',
'right sided facial/lip swelling': 'None',
'angioedema': 'Angioedema',
'hot': 'Snoring',
'cold': 'Flushing',
'extreme fatigue': 'None',
'low grade temp': 'None',
'associate vomited': 'None',
'diffuse rash at anterior and right lateral neck': 'None',
'warmth': 'Hunger',
'woozy': 'Miliaria',
'sore throat': 'None',
'clammy': 'Hypoaesthesia',
'some chest tightness': 'None',
'bilateral hand tingling': 'None',
'deep breath': 'None',
'rash': 'Rash',
'swelling': 'Swelling',
'itching': 'Pruritus',
'throat tightness': 'None',
'a panic attack": 'None',
'chest tightness': 'None',
'redness': 'Erythema',
'the injection spot': 'None',
'stomach ache': 'None',
'headaches': 'Headache',
'appetite': 'Hunger',
'localized reaction in left deltoid': 'None',
'firm to touch': 'None',
'feverish': 'Choking',
'a little soreness': 'None',
'other symptoms': 'None',
'severely nauseated': 'None',
'really loopy in the head': 'None',
'the dizziness': 'None',
'the nausea': 'None',
'being "lightheaded': 'None',
'slightly dizzy': 'None',
'patient symptoms': 'None',
'left arm swelling of forearm': 'None',
'old l wrist tattoo': 'None',
'several red dots': 'None',
'reaction': 'Dissociation',
'body wide itching': 'None',
'a prior anaphylaxis reaction': 'None',
'sweaty palms': 'None',
'throat swelling': 'None',
'the symptoms': 'None',
'pain': 'Pain',
'the pain': 'None',
```

```
'nauseous': 'Crying',  
'really tired': 'None',  
'a dull headache': 'None',  
'crummy': 'None',  
'fatigued': 'Bedridden',  
'a headache': 'None',  
'the headache': 'None',  
'normal arm pain': 'None',  
'itchy flushed skin': 'None',  
'a rash': 'None',  
'the itchy skin': 'None',  
'a sore arm': 'None',  
'swelling on arm': 'None',  
'a metallic taste in her mouth': 'None',  
'myalgias': 'Myalgia',  
'nauseas': 'Eructation',  
'stomach pain': 'None',  
'edema': 'Oedema',  
'decreased range of motion': 'None',  
'left arm rib cage pain': 'None',  
'muscle pain right and left thighs': 'None',  
'worsening upon ambulation muscle pain': 'None',  
'tingling right upper extremity': 'None',  
'hypotensive': 'Hypotension',  
'pale': 'Papule',  
'diaphoretic': 'Formication',  
'syncope': 'Syncope'}
```



```
In [34]: sim_linked_top_symp_dict
```

```
Out[34]: {'body aches': 'None',
'fatigue': 'Fatigue',
'headache': 'Headache',
'myalgia': 'Myalgia',
'nauseavomiting': 'None',
'chills': 'Chills',
'fever': 'Cough',
'mild numbness': 'None',
'injection site': 'None',
'warm': 'Flushing',
'progressive light-headedness': 'None',
'near-syncope': 'Ageusia',
'diaphoresis': 'Paraesthesia',
'20 minutes symptoms': 'None',
'bad taste in mouth': 'None',
'tingling in body , legs, back , across stomach': 'None',
'symptoms': 'Pain',
'shaking of hands': 'None',
'tingling': 'Paraesthesia',
'tongue tingling': 'None',
'smelling chemical smell': 'None',
'mouth itching': 'None',
'lightheadedness': 'Palpitations',
'flushing': 'Flushing',
'vagal and "spacey': 'None',
'chest heaviness': 'None',
'shortness of breath': 'None',
'rigors': 'Chills',
'urge to defecate': 'None',
'chest pain': 'None',
'really bad heartburn': 'None',
'body ache': 'None',
'some nausea': 'None',
'some redness to her neck and upper chest': 'None',
'recent illness': 'None',
'elevated bps': 'None',
'dizziness': 'Dizziness',
'being hot': 'None',
'nauseated': 'Paraesthesia',
'difficulty breathing': 'None',
'dizzy': 'Dizziness',
'lightheaded': 'Ageusia',
'feet elevated': 'None',
'any other complaints': 'None',
'complaints': 'Pain',
'facial flushing': 'None',
'pounding in chest': 'None',
'burning and hot ears': 'None',
'tingly in right arm and chest and hands': 'None',
'this reaction': 'None',
'tightness': 'Malaise',
'her tongue': 'None',
'anthrax': 'Diarrhoea',
'tingling of upper lip and cheeks': 'None',
'warmth in face': 'None',
'symptom': 'Pain',
'shaky': 'Tremor',
'shaking': 'Flushing',
'short of breath': 'None',
'wheeze': 'Cough',
'tightness in the chest or throat': 'None',
```

```
'pulse weak': 'None',
'sweating on torso': 'None',
'red': 'Swelling',
'the first sob': 'None',
'nausea': 'Nausea',
'vomiting': 'Vomiting',
'potential fever': 'None',
'pain in deltoid muscle': 'None',
'pain in the arm': 'None',
'energy': 'Swelling',
'injection site pain': 'None',
'a fever': 'None',
'right sided facial/lip swelling': 'None',
'angioedema': 'Urticaria',
'hot': 'Pain',
'cold': 'Flushing',
'extreme fatigue': 'None',
'low grade temp': 'None',
'associate vomited': 'None',
'diffuse rash at anterior and right lateral neck': 'None',
'warmth': 'Discomfort',
'woozy': 'Flushing',
'sore throat': 'None',
'clammy': 'Hypoaesthesia',
'some chest tightness': 'None',
'bilateral hand tingling': 'None',
'deep breath': 'None',
'rash': 'Rash',
'swelling': 'Swelling',
'itching': 'Pruritus',
'throat tightness': 'None',
'a panic attack": 'None',
'chest tightness': 'None',
'redness': 'Erythema',
'the injection spot': 'None',
'stomach ache': 'None',
'headaches': 'Headache',
'appetite': 'Fatigue',
'localized reaction in left deltoid': 'None',
'firm to touch': 'None',
'feverish': 'Malaise',
'a little soreness': 'None',
'other symptoms': 'None',
'severely nauseated': 'None',
'really loopy in the head': 'None',
'the dizziness': 'None',
'the nausea': 'None',
'being "lightheaded': 'None',
'slightly dizzy': 'None',
'patient symptoms': 'None',
'left arm swelling of forearm': 'None',
'old l wrist tattoo': 'None',
'several red dots': 'None',
'reaction': 'Swelling',
'body wide itching': 'None',
'a prior anaphylaxis reaction': 'None',
'sweaty palms': 'None',
'throat swelling': 'None',
'the symptoms': 'None',
'pain': 'Pain',
'the pain': 'None',
```

```
'nauseous': 'Flushing',  
'really tired': 'None',  
'a dull headache': 'None',  
'crummy': 'None',  
'fatigued': 'Fatigue',  
'a headache': 'None',  
'the headache': 'None',  
'normal arm pain': 'None',  
'itchy flushed skin': 'None',  
'a rash': 'None',  
'the itchy skin': 'None',  
'a sore arm': 'None',  
'swelling on arm': 'None',  
'a metallic taste in her mouth': 'None',  
'myalgias': 'Myalgia',  
'nauseas': 'Paraesthesia',  
'stomach pain': 'None',  
'edema': 'Swelling',  
'decreased range of motion': 'None',  
'left arm rib cage pain': 'None',  
'muscle pain right and left thighs': 'None',  
'worsening upon ambulation muscle pain': 'None',  
'tingling right upper extremity': 'None',  
'hypotensive': 'Syncope',  
'pale': 'Erythema',  
'diaphoretic': 'Rash',  
'syncope': 'Syncope'}
```

5. Evaluation

5.1 Automatic evaluation

Calculate the precision, recall, and f1 score for evaluation

```

In [35]: def evaluation_metrics(extracted_symp_dict, std_symp_list, count_s

    default_vector = np.zeros_like(glove_embeddings['headache'])

    # Iterate through the linked symptom dictionary to calculate t
    for extracted_symp, std_symptom in extracted_symp_dict.items():
        extracted_symp_vec = get_embedding(extracted_symp.lower(),
        std_symptom_vec = get_embedding(std_symptom.lower(), glove
        sim_score = np.dot(extracted_symp_vec, std_symptom_vec) /
        if sim_score >= 0.8:
            true_positives += count_symptom[extracted_symp]
        else:
            false_positives += count_symptom[extracted_symp]

    # Calculate false negatives
    false_negatives = len(set(std_symp_list) - set(extracted_symp_
    # Calculate precision, recall, and F1 score
    precision = true_positives / (true_positives + false_positives
    recall = true_positives / (true_positives + false_negatives) i
    f1_score = 2 * (precision * recall) / (precision + recall) if

    return precision, recall, f1_score

# Building the lists of standard symptom from 50 reports that will
std_symp_50 = []
for vaers_id in covid_reports['VAERS_ID'][:50]:
    for col_symp in ['SYMPTOM1', 'SYMPTOM2', 'SYMPTOM3', 'SYMPTOM4
        symptom = vaers_symp.loc[vaers_symp['VAERS_ID'] == vaers_i
        if symptom not in std_symp_50 and pd.notna(symptom):
            std_symp_50.append(symptom.lower())

```

```
In [36]: # Evaluate the exact matching method with standard symptom
exact_std_precision, exact_std_recall, exact_std_f1_score = evaluation_metrics.evaluate_exact_matching(symptoms, diseases)
print("The result of Rule-based matching (Exact method), comparing exact_std_precision, exact_std_recall, exact_std_f1_score")
print(f"Precision: {exact_std_precision}")
print(f"Recall: {exact_std_recall}")
print(f"F1 Score: {exact_std_f1_score}")
print()

# Evaluate the exact matching method with the most common symptoms
exact_top_precision, exact_top_recall, exact_top_f1_score = evaluation_metrics.evaluate_exact_matching_top(symptoms, diseases)
print("The result of Rule-based matching (Exact method), comparing exact_top_precision, exact_top_recall, exact_top_f1_score")
print(f"Precision: {exact_top_precision}")
print(f"Recall: {exact_top_recall}")
print(f"F1 Score: {exact_top_f1_score}")
print()

# Evaluate the fuzzy matching method with standard symptom
fuzzy_std_precision, fuzzy_std_recall, fuzzy_std_f1_score = evaluation_metrics.evaluate_fuzzy_matching(symptoms, diseases)
print("The result of Rule-based matching (Fuzzy method), comparing fuzzy_std_precision, fuzzy_std_recall, fuzzy_std_f1_score")
print(f"Precision: {fuzzy_std_precision}")
print(f"Recall: {fuzzy_std_recall}")
print(f"F1 Score: {fuzzy_std_f1_score}")
print()

# Evaluate the fuzzy matching method with the most common symptoms
fuzzy_top_precision, fuzzy_top_recall, fuzzy_top_f1_score = evaluation_metrics.evaluate_fuzzy_matching_top(symptoms, diseases)
print("The result of Rule-based matching (Fuzzy method), comparing fuzzy_top_precision, fuzzy_top_recall, fuzzy_top_f1_score")
print(f"Precision: {fuzzy_top_precision}")
print(f"Recall: {fuzzy_top_recall}")
print(f"F1 Score: {fuzzy_top_f1_score}")
print()

# Evaluate the similarity matching method with standard symptom
sim_std_precision, sim_std_recall, sim_std_f1_score = evaluation_metrics.evaluate_similarity_matching(symptoms, diseases)
print("The result of Similarity-based matching, comparing to the 1st symptom, sim_std_precision, sim_std_recall, sim_std_f1_score")
print(f"Precision: {sim_std_precision}")
print(f"Recall: {sim_std_recall}")
print(f"F1 Score: {sim_std_f1_score}")
print()

# Evaluate the similarity matching method with the most common symptoms
sim_top_precision, sim_top_recall, sim_top_f1_score = evaluation_metrics.evaluate_similarity_matching_top(symptoms, diseases)
print("The result of Similarity-based matching, comparing to the 1st symptom, sim_top_precision, sim_top_recall, sim_top_f1_score")
print(f"Precision: {sim_top_precision}")
print(f"Recall: {sim_top_recall}")
print(f"F1 Score: {sim_top_f1_score}")
print()
```

The result of Rule-based matching (Exact method), comparing to the list of the standard symptoms:

Precision: 0.178743961352657

Recall: 0.39361702127659576

F1 Score: 0.24584717607973422

The result of Rule-based matching (Exact method), comparing to the list of the most common symptoms:

Precision: 0.17391304347826086

Recall: 0.3870967741935484

F1 Score: 0.24

The result of Rule-based matching (Fuzzy method), comparing to the list of the standard symptoms:

Precision: 0.178743961352657

Recall: 0.39361702127659576

F1 Score: 0.24584717607973422

The result of Rule-based matching (Fuzzy method), comparing to the list of the most common symptoms:

Precision: 0.17391304347826086

Recall: 0.3870967741935484

F1 Score: 0.24

The result of Similarity-based matching, comparing to the list of the standard symptoms:

Precision: 0.23671497584541062

Recall: 0.46226415094339623

F1 Score: 0.31309904153354634

The result of Similarity-based matching, comparing to the list of the most common symptoms:

Precision: 0.21739130434782608

Recall: 0.4411764705882353

F1 Score: 0.29126213592233013

<ipython-input-35-928e66737da6>:9: RuntimeWarning: invalid value encountered in float_scalars

```
sim_score = np.dot(extracted_symp_vec, std_symp_vec) / (np.linalg.norm(extracted_symp_vec) * np.linalg.norm(std_symp_vec))
```

5.2 Manual evaluation

Manually check 20 reports to see the performance of the model to extract the symptoms

In [37]: *# Creating the table containing covid symptoms with only those col*
 symptom_cols = vaers_symp[['VAERS_ID', 'SYMPTOM1', 'SYMPTOM2', 'SY
 symptom_cols

Out [37]:

	VAERS_ID	SYMPTOM1	SYMPTOM2	SYMPTOM3	SYMPTOM4	SYMPTOM5
0	375646	Chest X-ray normal	Chest discomfort	Cough	Dyspnoea	Wheezing
1	375647	Erythema	Rash	Rash macular	NaN	NaN
2	375648	Nausea	Vomiting	NaN	NaN	NaN
3	375650	Abdominal pain upper	Diarrhoea	Fatigue	Headache	Myalgia
4	375650	Nausea	NaN	NaN	NaN	NaN
...
1947795	2688371	Hip fracture	Hypokinesia	Hypophagia	Mobility decreased	Oesophagitis
1947796	2688371	Pain	Pyrexia	NaN	NaN	NaN
1947797	2688372	Chills	Decreased appetite	Dizziness	Fatigue	Headache
1947798	2688372	Nausea	NaN	NaN	NaN	NaN
1947799	2688373	Sleep disorder	Tremor	NaN	NaN	NaN

1947800 rows × 6 columns


```
In [38]: # Getting the standard symptoms from 20 reports
std_symptom_20 = []
report_number = 1
for i, sentence in enumerate(covid_reports['SYMPTOM_TEXT'][:20]):
    # annotate clinical text
    doc = nlp(sentence)

    # Print out 20 reports for manually evaluating
    print("Report No.:", report_number)
    print("Vaers ID:", covid_reports['VAERS_ID'][i])
    print("Input:", sentence)
    print("Symptoms:")
    for ent in doc.entities:
        if ent.type == 'PROBLEM':
            print(f'\t{ent.text}')
    print("\nStandard symptoms:")
    for col in ['SYMPTOM1', 'SYMPTOM2', 'SYMPTOM3', 'SYMPTOM4', 'S
        if pd.notna(symptom_cols[symptom_cols['VAERS_ID'] == covid
            std_symptom = symptom_cols[symptom_cols['VAERS_ID'] ==
            print('\t', std_symptom)
            std_symptom_20.append(std_symptom)

    print()
    report_number += 1
```

Report No.: 1

Vaers ID: 2669769

Input: body aches, fatigue Narrative: Took OTC Tylenol Other Relevant History:

Symptoms:

- body aches
- fatigue

Standard symptoms:

- Fatigue
- Pain

Report No.: 2

Vaers ID: 2527460

Input: Headache, Myalgia, Nausea Vomiting, chills Narrative:

Symptoms:

- Headache
- Myalgia
- Nausea Vomiting
- chills

Standard symptoms:

- Chills
- Headache
- Myalgia
- Nausea
- Vomiting

Report No.: 3

Vaers ID: 2673135

Input: Headache, Fever, Body aches Narrative: Other Relevant History:

Symptoms:

- Headache
- Fever
- Body aches

Standard symptoms:

- Headache
- Pain
- Pyrexia

Report No.: 4

Vaers ID: 2672717

Input: Headache & Myalgia Narrative: Other Relevant History:

Symptoms:

- Headache
- Myalgia

Standard symptoms:

- Headache
- Myalgia

Report No.: 5

Vaers ID: 902418

Input: Patient experienced mild numbness traveling from injection site up and down arm that subsided over 20 minutes.

Symptoms:

- mild numbness
- injection site

Standard symptoms:
Hypoaesthesia
Injection site hypoaesthesia

Report No.: 6
Vaers ID: 902440
Input: C/O Headache
Symptoms:
Headache

Standard symptoms:
Headache

Report No.: 7
Vaers ID: 902446
Input: felt warm, hot and face and ears were red and flushed.
Symptoms:
warm

Standard symptoms:
Erythema
Feeling hot
Flushing

Report No.: 8
Vaers ID: 902464
Input: within 15 minutes progressive light-headedness leading to near-syncope and diaphoresis. After 20 minutes symptoms subsided.
Symptoms:
progressive light-headedness
near-syncope
diaphoresis
20 minutes symptoms

Standard symptoms:
Dizziness
Electrocardiogram normal
Hyperhidrosis
Laboratory test normal
Presyncope

Report No.: 9
Vaers ID: 902465
Input: Pt felt wave come over body @ 1218 starting in head and going down. Bad taste in mouth, tingling in body , legs, back , across stomach, BP 150/100 P 120 @ 1219, EMS activated. BP 120/80, P 80 Pt alert and oriented, Pt declined transport and Benadryl. Symptoms come and go, pt feels better but then bad taste in mouth starts, shaking of hands, tingling starts again in stomach and back. @ 1300 pt requests Benadryl, 25 mg administered. Pt notified family by phone of circumstances and family in transit. @1324 BP 120/80, P 84, tongue tingling and pt reports smelling chemical smell. @1345 Pt complained of mouth itching, EMS activated and will transport to Medical Center. Pt oriented and transported at @1345
Symptoms:
Bad taste in mouth
tingling in body , legs, back , across stomach
Symptoms
shaking of hands
tingling

tongue tingling
smelling chemical smell
mouth itching

Standard symptoms:
Dysgeusia
Oral pruritus
Paraesthesia
Paraesthesia oral
Parosmia

Report No.: 10

Vaers ID: 902468

Input: Within 1 minute, patient complained of symptoms of lightheadedness, flushing, asked for water. Symptoms persisted, reported vagal and "spacey", vitals were 117/91, HR 67, O2 sat 99% on room air. Reported chest heaviness, shortness of breath and within 5 minutes developed rigors and urge to defecate. 911 called, repeat vital 150/89 HR 113 O2sat 97%, continues to want to defecate. 1 loose BM, transferred to ED

Symptoms:
symptoms
lightheadedness
flushing
Symptoms
vagal and "spacey"
chest heaviness
shortness of breath
rigors
urge to defecate

Standard symptoms:
Chest discomfort
Chills
Defaecation urgency
Diarrhoea
Dizziness

Report No.: 11

Vaers ID: 902479

Input: rPfizer-BionNTech COVID-19 Vaccine EUA 5-7 minutes after the vaccine Associate stated she did not feel right, mentioned chest pain. "My chest feels funny. It feels like when you have really bad heartburn coming on". "I feel flushed like when you get contrast for a CT". Pulse 90 BP 160/90 checked later 130/90

Symptoms:
chest pain
really bad heartburn

Standard symptoms:
Chest pain
Feeling abnormal
Flushing
Intensive care

Report No.: 12

Vaers ID: 902490

Input: Headache, body ache

Symptoms:
Headache
body ache

Standard symptoms:

Headache
Pain

Report No.: 13

Vaers ID: 902491

Input: Within a few minutes of receiving the COVID 19 vaccination, patient developed lightheadedness, shortness of breath, headache, and some nausea. She did get some redness to her neck and upper chest. No recent illness. Had elevated BPs ranged from 158/103 to 207/126. HR ranged from 82-106. O2 sats always > 96%. Temp 37.1 C. Received Tylenol 1000 mg PO, Dexamethasone 10 mg IV, diphenhydramine 50 mg IV, famotidine 20 mg IV, ketorolac 30 mg IV, ondansetron 4 mg IV, and 1 L NS. Patient prescribed EpiPen and prednisone and discharged.

Symptoms:

lightheadedness
shortness of breath
headache
some nausea
some redness to her neck and upper chest
recent illness
elevated BPs

Standard symptoms:

Blood pressure increased
Dizziness
Dyspnoea
Erythema
Headache

Report No.: 14

Vaers ID: 902492

Input: About 25 minutes after receiving vaccine complained of dizziness and being hot and nauseated. No difficulty breathing. No chest pain. B/P was 130/90 and was monitored. It went down to 124/80 after he started feeling better. He was wearing sweater over shirt and it was warm in building. Took sweater off. Cool wet cloth applied to back of neck. States he had only had a donut and cup of hot chocolate before receiving vaccine. Sprite and peanut butter crackers given. Became nauseated after eating peanut butter crackers. Blood pressure monitored. He laid on exam table for about 15 minutes. He felt better. Stood up and walked to conference room for another 15 minutes. Stated he felt much better and was ready to leave. Coworker drove him back. Received email from him letting us know he had made it back and they had stopped and eaten pizza on the way. Received text from coworker that he was dizzy and seeing spots and that his blood pressure had been 120/80 and then spiked to 160/100. Coworkers taking him to ER at Hospital for evaluation.

Symptoms:

dizziness
being hot
nauseated
difficulty breathing
chest pain
nauseated
dizzy

Standard symptoms:

Blood pressure increased
Dizziness
Feeling hot
Nausea
Visual impairment

Report No.: 15

Vaers ID: 902493

Input: At 12:55 pm 10 minutes following vaccine being given state s feeling lightheaded and flush. Was sitting in the chair. Encouraged him to lay down on the floor which he did on his own. Feet elevated. BP 174/70 pulse 82. Denies any other complaints. Laid on floor for 15 minutes then sat in chair. Denies complaints. 1:15 pm was allowed to leave. BP 120/80 and states feeling fine.

Symptoms:

lightheaded
Feet elevated
any other complaints
complaints

Standard symptoms:

Dizziness
Flushing

Report No.: 16

Vaers ID: 902505

Input: Patient felt facial flushing, pounding in chest, burning and hot ears and blood pressure went up. Tingly in right arm and chest and hands. Symptoms resolved, after a few minutes but then returned. Patient sat with nurse during this reaction.

Symptoms:

facial flushing
pounding in chest
burning and hot ears
Tingly in right arm and chest and hands
Symptoms
this reaction

Standard symptoms:

Blood pressure increased
Ear discomfort
Flushing
Palpitations
Paraesthesia

Report No.: 17

Vaers ID: 902508

Input: She claims she experienced tightness in the right side of throat and her tongue started tingling. Took her the Emergency Department, She decided to go and buy Benadryl

Symptoms:

tightness
her tongue
tingling

Standard symptoms:

Paraesthesia oral
Throat tightness

Report No.: 18

Vaers ID: 902514

Input: System was not populating immunization record, member denied having immunizations within last 14 days. Vaccine given, record populated and patient had anthrax on 12/10/20

Symptoms:

anthrax

Standard symptoms:

Inappropriate schedule of product administration

Report No.: 19

Vaers ID: 902518

Input: Tingling of upper lip and cheeks, warmth in face, and itchy eyes Treatment: diphenhydramine 50 mg PO x1 Outcome: symptom onset within 15 minutes of vaccine administration. Symptoms resolved within 20-30 minutes of diphenhydramine administration.

Symptoms:

Tingling of upper lip and cheeks

warmth in face

symptom

Symptoms

Standard symptoms:

Eye pruritus

Feeling hot

Paraesthesia

Paraesthesia oral

Report No.: 20

Vaers ID: 902524

Input: I am a immunization nurse at this location. I gave 2 of the first 4 Covid vaccinations given at our location. Then I received dose # 5. It was easy. I did a couple of things and then returned to my desk. As I sat down, my arm started feeling very heavy. I was unable to send a text. I told staff that I was feeling funny and that I was going to the other room to lay down. Staff followed me and took my Pulse 100 and BP 164/ 82 (high for me!) . I felt shaky, but my hands were not shaking. Put a wet cloth on my head and laid there a few minutes, telling staff stories and laughing at my BP. When I sat up, my BP was 126/74 and pulse was 80. I stood up for a minute or two, then my legs got heavy and I sat down for a few more minutes. I went to the bathroom and came back to my desk, but was weak and tired. I ate and drank some fluids. Because it was snowing and I live 25 miles away, I accepted a ride home from a co-worker. I walked across the parking lot without problems and talked all the way home. At home, I was tired, but had a sandwich and talked on the phone. I would still describe myself as tired, but functioning.

Symptoms:

shaky

shaking

Standard symptoms:

Asthenia

Fatigue

Feeling abnormal

Heart rate increased

Hypertension

In [39]: `std_symptom_20`


```
Out[39]: ['Fatigue',
          'Pain',
          'Chills',
          'Headache',
          'Myalgia',
          'Nausea',
          'Vomiting',
          'Headache',
          'Pain',
          'Pyrexia',
          'Headache',
          'Myalgia',
          'Hypoaesthesia',
          'Injection site hypoaesthesia',
          'Headache',
          'Erythema',
          'Feeling hot',
          'Flushing',
          'Dizziness',
          'Electrocardiogram normal',
          'Hyperhidrosis',
          'Laboratory test normal',
          'Presyncope',
          'Dysgeusia',
          'Oral pruritus',
          'Paraesthesia',
          'Paraesthesia oral',
          'Parosmia',
          'Chest discomfort',
          'Chills',
          'Defaecation urgency',
          'Diarrhoea',
          'Dizziness',
          'Chest pain',
          'Feeling abnormal',
          'Flushing',
          'Intensive care',
          'Headache',
          'Pain',
          'Blood pressure increased',
          'Dizziness',
          'Dyspnoea',
          'Erythema',
          'Headache',
          'Blood pressure increased',
          'Dizziness',
          'Feeling hot',
          'Nausea',
          'Visual impairment',
          'Dizziness',
          'Flushing',
          'Blood pressure increased',
          'Ear discomfort',
          'Flushing',
          'Palpitations',
          'Paraesthesia',
          'Paraesthesia oral',
          'Throat tightness',
          'Inappropriate schedule of product administration',
          'Eye pruritus',
          'Feeling hot',
```

```
'Paraesthesia',  
'Paraesthesia oral',  
'Asthenia',  
'Fatigue',  
'Feeling abnormal',  
'Heart rate increased',  
'Hypertension']
```

```
In [40]: len(std_symptom_20)
```

```
Out[40]: 68
```

```
In [41]: # Finding those misses symptoms and found symptoms  
missed_symptom = ['Erythema', 'Flushing', 'Blood pressure increase'  
found_symptom = 0  
for symp in std_symptom_20:  
    if symp not in missed_symptom:  
        found_symptom += 1  
  
# Calculate the percentage of discovered symptoms from 20 reports  
total_std_symptom_20 = len(std_symptom_20)  
percentage_found_symptoms = (found_symptom / total_std_symptom_20)  
print(f"The percentage of discovered symptoms from 20 report, comp  
print("Missed symptoms:")  
for i, symp in enumerate(missed_symptom):  
    print(f"{i + 1}. {symp}")
```

The percentage of discovered symptoms from 20 report, compared to the standard symptoms: 79.41176470588235 %

Missed symptoms:

1. Erythema
2. Flushing
3. Blood pressure increased
4. Visual impairment
5. Eye pruritus
6. Asthenia
7. Heart rate increased
8. Hypertension