

Final Assignment
(Thematic Analysis Starts Page 17)

Topic Modeling to Extract Meaningful Insights
from Free-text in Electronic Health Records of Long-COVID Patients

Cynthia Yip
INF2209H: Human-Centered Topic Models
Shion Guha
14 Dec 2022

Introduction

Contrary to the initial characterization of COVID-19 as a relatively short-term, acute disease, some Covid-19 survivors start to report prolonged physical, cognitive and psychological symptoms that linger beyond the typical post-viral recovery period [4]. Many long COVID patients find the lingering health problems debilitating. Some even face enormous occupational, economic, social, and emotional difficulties (Davis et al., 2021). Clinical evidence and research are evolving on long COVID, [2,7,13,14]. Many unknowns remain, however, around the diagnosis, prevalence, phenotypes, and management of long COVID.

With rapid growth in volume and diversity of electronic health record (HER) datasets, topic modelling has been increasingly used for text analysis of EHR [3,18]. Topic modelling allows us to extract underlying themes from text data. Incorporating topic modeling techniques in long COVID research can allow a more rapid process in understanding unknowns of long COVID, seeking effective interventions, and warranting future investigations. This study seeks to apply four topic modeling techniques, namely latent Dirichlet allocation (LDA), Non-negative Matrix Factorization (NMF), Biterm Topic Model (BTM) and semi-supervised topic models: Correlation Explanation (CorEx) on 655 EHR of patients from the Post-COVID clinic of Parkview Health. I focus the analysis only on the dataset's *Journey / Major events* attribute, a free text recoding the patients journey after their COVID-19 infection. I evaluate the performance of the models using the average number of overlapping top topics, perplexity, coherence, Jaccard similarity. Finally, the lists of topics for each of the best models are printed and labeled for detailed comparison within and between the models.

Background

At the initial stage of the pandemic, research on COVID-19 largely centered around understanding the virus and managing the acute phase of COVID-19. The term long COVID and COVID-19 long hauler first appeared among COVID-19 survivors sharing the experience of ongoing, returning, or new symptoms that persist for months or even longer after being infected initially on Twitter [4]. Unfortunately, some medical professionals initially dismissed this idea of COVID-19 survivors experiencing lingering symptoms as merely a psychiatric condition [4]. It is the insistence on the need for serious investigation coming directly out of patient testimonies that the matter starts to gain momentum in medical research.

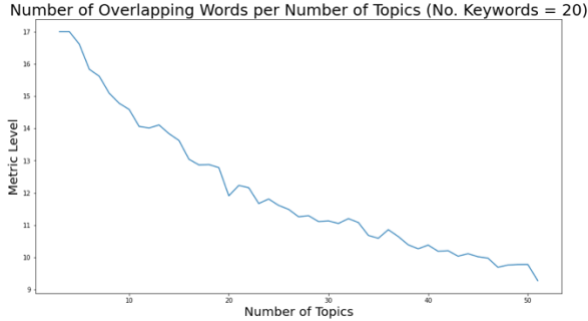
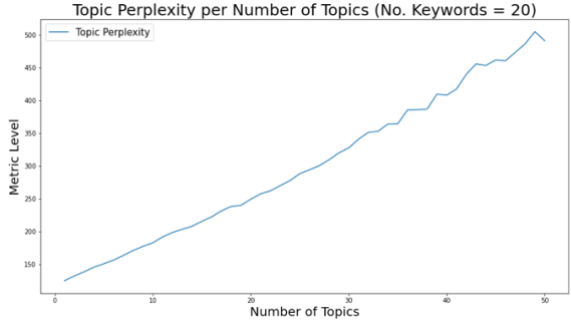
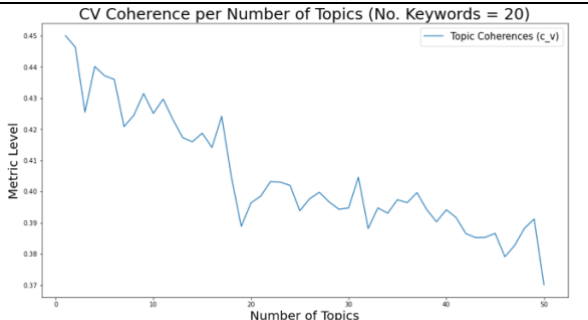
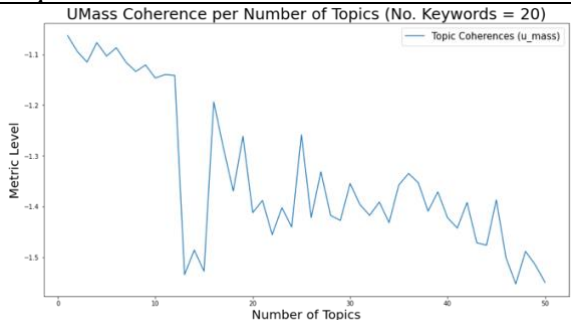
This natural history of long COVID as a patient-made illness makes patient input particularly important in studying long COVID [4]. Traditional medical research, driven mostly by health practitioners and researchers, can often result in discrepancies between the research that patients want and the research that actually happens [9,10]. Patient-centred angle data focusing on the experiences and needs of the patients themselves can greatly help in developing better treatment options. Therefore, I choose to analyze the free-text fields of long COVID patients' journey to uncover hidden characteristics and patterns of long COVID that are otherwise unavailable using other research methods.

the number of topics ranging from 1 to 50. Next, I evaluate each model's average number of overlapping top 20 keywords, perplexity score, coherence score, and Jaccard similarity to determine the optimal number of topics. Lastly, I print the list of the top ten keywords, add labels to each topic in the best model, and review the quality of the topics.

Some additional steps are performed for CoReX. Apart from running it as an unsupervised topic model, I also run the anchored CoReX, the semi-supervised topic model, by adding anchor words to topics. The list of anchor words is compiled from previous long COVID research results in characterizing long COVID symptoms [1,4,14], cluster analysis of the same dataset, and results from the LDA, NMF, and BTM topic models.

Results

Model 1: Latent Dirichlet Allocation (LDA)

<p>Figure 3 <i>Number of Overlapping top 20 Words among Topics per Number of Topics</i></p> 	<p>Figure 4 <i>Topic Perplexity per Number of Topics</i></p> 
<p>Figure 5 <i>CV Coherence Score per Number of Topics</i></p> 	<p>Figure 6 <i>UMass Coherence Score per Number of Topics</i></p> 
<p>Figure 7 <i>Average Topic Overlap (Jaccard Similarity), CV Coherences and Ideal Number of Topics</i></p>	<p>Figure 8 <i>Average Topic Overlap (Jaccard Similarity), UMass Coherences and Ideal Number of Topics</i></p>

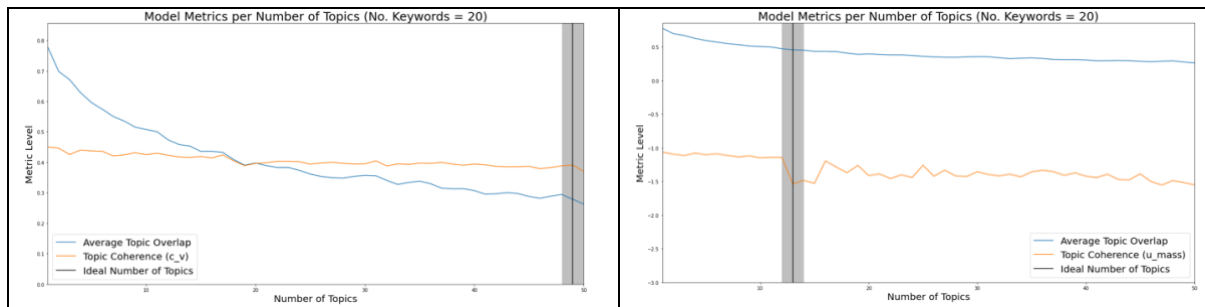


Figure 3 shows the average number of overlapping words when comparing the top 20 keywords between topics per number of topics. When the number of topics is small, the resulting topics in the LDA models are very similar, with a high of 17 overlapping keywords when the number of topics is 2. The number of overlapping top 20 words decreases as the number of topics increases. The perplexity scores of the models show the same phenomenon (figure 4), starting at around 120 and going up to around 500. Topic overlap decreases as the number of topics in the model increases.

Figure 5 and 6 shows the CV coherence score and UMass coherence Score. Both scores have a decreasing trend as the number of topics increases. They both show a steep decrease at around the number of topics equal to 15 to 20.

When plotting the average topic overlap calculated through Jaccard similarity, coherence score, and the optimal number of topics, using the CV coherence score versus the UMass coherence score give a very different optimal number of topics. Figure 7, using the CV coherence score, shows an ideal number of 49, while figure 8, using the UMass coherence score, shows an ideal number of 13.

Although some metrics (figure 7) suggest a much higher number of topics, I pick the 13 topics and 25 topics models for further analysis. These two models consider all of the above metrics with a balance between having a better metric score and interpretability. The topics in the 13-topics model are similar to each other. Although there are only an average of around 13 overlapping keywords between the topics (figure 3), printing out the keywords shows that the remaining non-overlapping keywords provide little useful information in distinguishing between the topics. While the model presents 13 topics, there are in fact, only a few noticeable different themes present:

- Fatigue and headache
- body pain (chest, headache) and respiratory-related symptoms
- cognitive deficits (taste/smell) and headache

The 25 topics model performs slightly better with more varieties among the topics. Some new topics observed are:

- Cardiac (chest, heart)
- Body pain (chest, muscle, joint)
- Cognitive (memory loss)
- Topics related to work and going to hospitals

LDA 13 Topics Model		
Topic	Labels	Top 10 Keywords
Topic 1	Brainfog, headache, tiredness, body aches	brain, fog, headache, fatigue, pain, day, go, smell, sleep, chest

Topic 2	Body aches (chest, headache), respiratory	pain, chest, headache, fatigue, cough, smell, symptom, start, go, day
Topic 3	Smell, taste, headache	day, symptom, smell, headache, taste, feel, bad, cough, pain, much
Topic 4	Smell, taste, headache	day, symptom, smell, headache, taste, feel, bad, cough, pain, much
Topic 5	Headache, smell, taste, cough, fever	headache, smell, taste, pain, day, start, cough, fatigue, loss, fever
Topic 6	Feel bad, home	day, bad, get, start, loss, feel, headache, home, cough, fatigue
Topic 7	Fatigue, headache, taste, smell	fatigue, headache, taste, start, cough, symptom, smell, go, feel, body
Topic 8	Headache, cough, fever	pain, start, headache, fatigue, smell, cough, taste, fever, get, go
Topic 9	Fatigue, headache, pain	fatigue, headache, pain, symptom, start, heart, feel, chest, brain, fog
Topic 10	Fatigue, headache, smell and taste	symptom, fatigue, pain, headache, smell, taste, cough, loss, brain, day
Topic 11	Pain, smell, taste	pain, go, day, smell, taste, fatigue, headache, fog, brain, feel
Topic 12	Taste, smell, headache	taste, headache, symptom, pain, smell, start, get, continue, fatigue, day
Topic 13	?	start, go, get, week, feel, back, come, symptom, pain, leg

LDA 25 Topics Model		
Topic	Labels	Top 10 Keywords
Topic 1	Brain fog, headache, fatigue	brain, fog, headache, fatigue, increase, heart, work, pain, chest, continue
Topic 2	Pain, chest discomfort, fatigue, headache, muscle pain	pain, chest, fatigue, symptom, headache, cough, muscle, day, work, go
Topic 3	Headache, cough, work	day, symptom, headache, bad, cough, much, work, time, feel, go
Topic 4	Pain, short of breath, chest discomfort, cough	pain, sob, fatigue, headache, time, cough, brain, day, feel, chest
Topic 5	Body aches, taste/smell, fever	headache, pain, taste, smell, start, cough, day, chest, loss, fever
Topic 6	Worse during certain time, short of breath, work	day, get, bad, sob, feel, symptom, go, fatigue, headache, work
Topic 7	Headache, fatigue, difficulty, breath	headache, fatigue, symptom, difficulty, breathe, start, walk, taste, feel, back
Topic 8	Pain, fatigue, cough, breath	pain, start, get, fatigue, cough, breath, headache, chest, fever, work
Topic 9	Pain, heart and chest, memory loss	pain, heart, chest, memory, headache, symptom, fatigue, start, severe, issue
Topic 10	Fatigue, brainfog, headache	fatigue, symptom, brain, fog, headache, pain, day, cough, loss, taste
Topic 11	Pain, brainfog, day, tremor, continuous symptoms, liver	pain, go, fog, brain, day, feel, tremor, continue, shortness, liver
Topic 12	Headache, taste, continuous symptoms	headache, taste, start, pain, get, continue, symptom, smell, muscle, see
Topic 13	?	week, start, diagnose, high, notice, go, get, blood, negative, come
Topic 14	Hospital, heart, fever, fatigue	hospital, heart, day, fever, go, week, fatigue, get, blood, work

Topic 15	Fatigue, smell/taste loss, pain, work	fatigue, start, smell, day, pain, taste, work, loss, feel, go
Topic 16	Smell/taste, cough, fatigue	smell, taste, cough, fatigue, pain, start, day, loss, symptom, headache
Topic 17	Headache, cough, chest	get, symptom, smell, start, headache, go, cough, back, chest, work
Topic 18	Smell/taste, fatigue, headache	smell, taste, feel, body, fatigue, headache, go, pain, symptom, back
Topic 19	Pain headache, cough, chest	pain, headache, fatigue, cough, chest, start, go, symptom, bad, smell
Topic 20	Pain, go for walk, fatigue, fever	day, pain, go, walk, start, fatigue, fever, work, much, loss
Topic 21	Brain fog, cough, joint pain, headache	fog, cough, brain, start, joint, bad, fatigue, headache, fever, pain
Topic 22	Headache, taste/smell, body aches, fatigue	headache, taste, smell, pain, fatigue, ache, day, go, body, feel
Topic 23	Body aches	pain, feel, day, start, right, symptom, back, smell, headache, fatigue
Topic 24	Headache, fatigue, pain, short of breath	headache, fatigue, pain, sob, day, continue, go, smell, body, start
Topic 25	Smell/taste, fatigue, headache, cough	smell, taste, fatigue, headache, cough, pain, day, symptom, muscle, start

Model 2: NMF

Figure 9

Number of Overlapping top 20 Words among Topics per Number of Topics

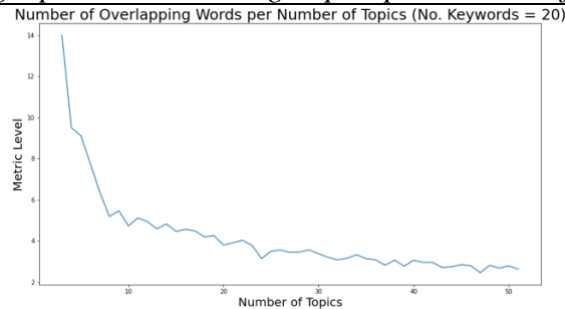


Figure 10

Topic CV Coherences Score per Number of Topics

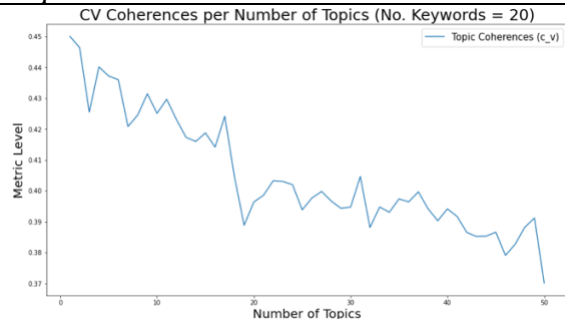


Figure 11

Topic UMass Coherences Score per Number of Topics

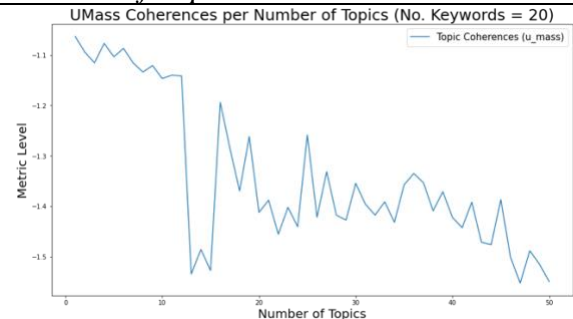


Figure 12

Average Topic Overlap (Jaccard Similarity), CV Coherences and Ideal Number of Topics

Figure 13

Average Topic Overlap (Jaccard Similarity), UMass Coherences and Ideal Number of Topics

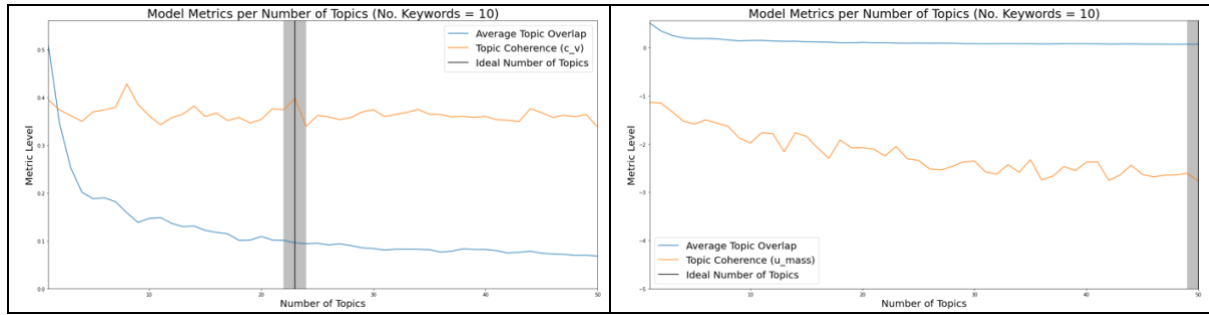


Figure 9 shows the average number of overlapping words when comparing the top 20 keywords between topics per number of topics. For NMF, the number of overlapping of the top 20 words goes from a high of 14 at number of topics equal to 2 and decreases as the number of topics increases. There is a steep decrease as the number of topics increases to 10. The line becomes relatively flat afterward.

Figure 10 and 11 shows the CV coherence score and UMass coherence score. Similar to the LDA models, scores decrease as the number of topics increases. UMass coherence score fluctuates more than the CV coherence score. They both show a steep decrease at around the number of topics equal to 15 to 20.

When plotting the average topic overlap calculated through Jaccard similarity, coherence score, and the optimal number of topics, using the CV coherence score versus the UMass coherence score again gives a very different optimal number of topics. Figure 9, using the CV coherence score, shows an ideal number of 23, while figure 6, using the UMass coherence score, shows an ideal number of 50.

In balancing all the above metrics, I choose the 23 topics model. There is only an average of around three overlapping keywords between the topics in this model. It, therefore, presents more separated topics than the LDA models. For example, in topics with pain as a top occurring keyword, the word appears with respiratory symptoms (topic 1); taste and smell loss (topic 3); fatigue, headache, and other extreme symptoms (topic 7); and weakness and fatigue (topic 20). There are also new topics, such as topic 9, which relates fatigue with sleep, and topic 15, which discusses worsening symptoms.

NMF 23 Topics Model		
Topic	Labels	Top 10 Keywords
Topic 1	Respiratory, pain (back, neck)	cough, pain, chest, low, smell, back, neck, fever, breath, start
Topic 2	Headache, pain	headache, start, feel, pain, much, go, chest, day, normal, week
Topic 3	Pain, cough, taste/smell loss	pain, cough, headache, fatigue, taste, ache, smell, body, chest, sob
Topic 4	Work, home	work, day, get, feel, start, much, home, migraine, issue, go
Topic 5	Feeling sick, nebulizer	go, day, use, sick, heart, time, get, inhaler, nebulizer, even
Topic 6	Taste/smell loss	smell, taste, loss, symptom, fatigue, get, day, bad, time, issue
Topic 7	Pain, fatigue, headache, extreme	pain, fatigue, muscle, visit, begin, headache, resolve, extreme, ear, joint
Topic 8	Work	right, back, go, work, chest, head, day, cough, headache, side
Topic 9	Fatigue, sleep	fatigue, first, body, sleep, day, extreme, smell, high, get,

		notice
Topic 10	Taste/smell, brainfog	taste, smell, feel, brain, fog, symptom, get, start, chest, normal
Topic 11	Walk	much, go, day, get, good, walk, also, bad, come, week
Topic 12	Increased / continuous symptoms	increase, day, pain, first, continue, bp, time, symptom, start, sleep
Topic 13	Sleep, respiratory, heart	sleep, chest, heart, brain, breath, fog, throat, shortness, trouble, fatigue
Topic 14	Brainfog, fatigue, muscle, anxiety	brain, fog, fatigue, low, muscle, headache, severe, symptom, anxiety, joint
Topic 15	Worsen symptoms, chronic	rec, see, wue, week, worsen, asthma, sid, sharaan, chronic, loss
Topic 16	Tiredness, episode	right, hand, head, arm, day, tired, leg, left, episode, foot
Topic 17	Chest discomfort, walk	chest, feel, walk, go, inhaler, start, body, severe, appt, use
Topic 18	Work, normal time	work, feel, symptom, back, time, see, normal, well, come, go
Topic 19	SOB, fatigue, increased symptom	sob, fatigue, heart, headache, increase, leg, rate, symptom, diagnose, able
Topic 20	Pain, weakness, fatigue	pain, day, weakness, extreme, muscle, symptom, continue, bad, week, fatigue
Topic 21	Asthma, chronic	rec, see, asthma, headache, wue, chronic, sharaan, sid, first, order
Topic 22	Brainfog, headache, joint pain	brain, fog, headache, joint, bad, wake, head, bed, get, continue
Topic 23	Shaking, walk, sit	see, shake, symptom, go, walk, mile, sit, side, mostly, ct

Model 3: BTM

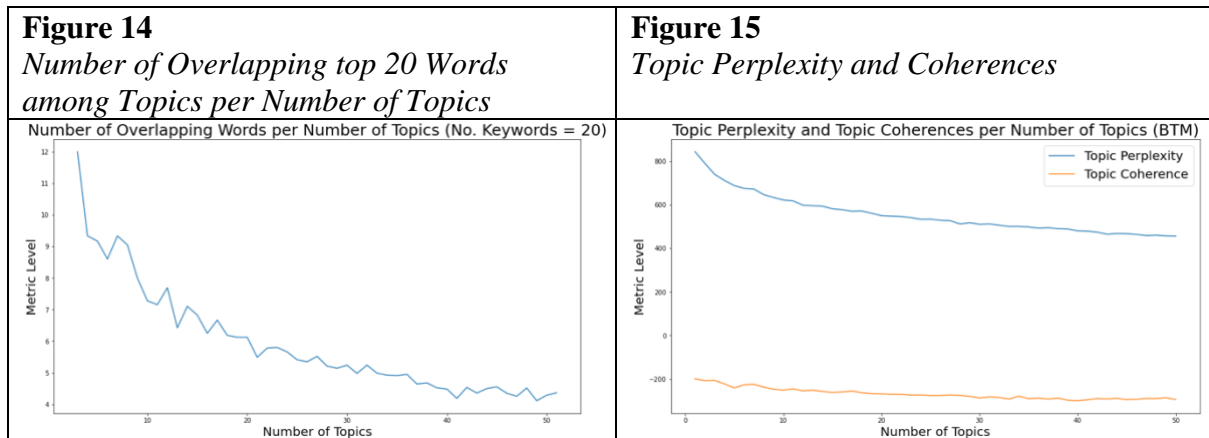


Figure 14 shows the average number of overlapping words when comparing the top 20 words between topics per number of topics. For BTM, the number of overlapping of the top 20 words goes from a high of 11 at around the number of topics = 3 and decreases as the number of topics increases. It flattens at around the number of topics = 30 and further flattens at around the number of topics = 40

Figure 15 shows the topic perplexity and coherence Score. For BTM, the topic perplexity is much higher than the other models, while the topic coherence score is much lower than the other models.

In balancing all the above metrics, I choose the 25 topics model. There is only an average of around five overlapping keywords between the topics in this model. There are a

few topics with the overlapping top one keyword pain and brain fog (e.g., topics 3, 14, 22, 23, and 25; topics 5 and 6), but that is expected as these are the most prevalent symptoms of long COVID and the rest of the keywords are quite different. Other than long COVID symptoms, this BTM model can find other hidden structures, such as patients discussing their symptoms' starting, work, and physical activities like walking. Topic 21 about antibodies and CT, topic 22 about blood clots, oxygen, and therapy, and topic 24 about asthma and cough are not present in the previous models.

BTM 25 Topics Model		
Topic	Labels	Top 10 Keywords
Topic 1	Starting of symptoms, work, pain	start, symptom, much, pain, work, go, visit, feel, time, day
Topic 2	Walk time, work, feel bad	go, day, walk, get, work, much, bad, feel, hour, take
Topic 3	Chest pain, brainfog, headache, nausea	pain, chest, brain, headache, fog, start, issue, nausea, fatigue, normal
Topic 4	Body aches (head, back, leg, foot)	right, pain, head, back, leg, left, leave, side, feel, foot
Topic 5	Brainfog, fatigue, muscle, anxiety, affected sleep	brain, fatigue, fog, issue, muscle, start, headache, anxiety, sleep, symptom
Topic 6	Brainfog, fatigue, headache, pain	brain, fog, bad, fatigue, headache, smell, symptom, pain, get, much
Topic 7	Taste/smell, start/begin of symptoms	smell, taste, food, much, day, start, continue, come, begin, symptom
Topic 8	Work, smell/taste, fatigue	back, start, work, feel, smell, fatigue, day, taste, come, symptom
Topic 9	Chest, heart, blood, liver	see, symptom, chest, normal, heart, visit, order, continue, blood, liver
Topic 10	Work, cough, chest	work, go, feel, get, day, back, cough, chest, pain, good
Topic 11	Fever, headache, cough	symptom, day, fever, headache, cough, pain, back, work, go, ache
Topic 12	Taste/smell, cough, fatigue	smell, cough, taste, fatigue, loss, headache, pain, fever, sob, ache
Topic 13	Heart, work related	day, get, feel, go, time, heart, much, week, work, start
Topic 14	Body aches, fever	pain, headache, chest, fatigue, fever, cough, ache, joint, body, fog
Topic 15	Headache, taste/smell loss, fatigue	headache, smell, taste, loss, fatigue, brain, ache, bad, joint, get
Topic 16	Taste/smell, pain, sleep	smell, taste, see, also, day, refer, right, pain, headache, sleep
Topic 17	Taste, tiredness, diarrhea	start, feel, day, symptom, taste, fatigue, diarrhea, tired, headache, smell
Topic 18	Hospital, home, pneumonia	hospital, parkview, home, go, admit, start, cough, give, walk, pneumonia
Topic 19	Work, home, breathing	get, day, work, week, go, home, breathing, good, take, feel
Topic 20	Headache, fatigue, body aches (cough, body, brain, muscle)	headache, fatigue, pain, cough, body, ache, brain, fever, muscle, fog
Topic 21	Antibody, CT, sinus, sleep	antibody, ct, see, negative, sinus, appt, show, start, sleep, go
Topic 22	Blood clot, oxygen, therapy, home	pain, blood, hospital, home, clot, leg, oxygen, go, normal, therapy

Topic 23	Pain, headache	pain, headache, taste, sob, visit, chest, cough, oxygen, prmc, home
Topic 24	Asthma, cough, steroid, inhaler, SOB	see, asthma, cough, rec, day, steroid, inhaler, sob, allergy, severe
Topic 25	Cardiac, respiratory	pain, chest, heart, breath, go, start, increase, headache, high, sob

Method 4: CoRex

Figure 16

CV Topic Coherence Score

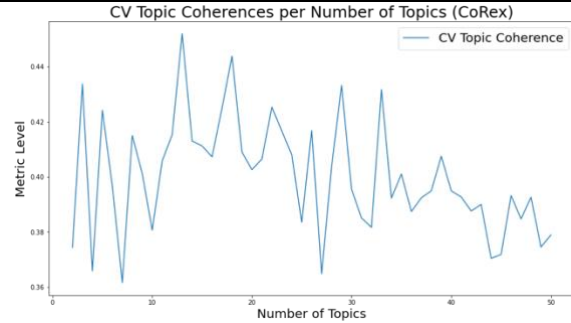


Figure 17

UMass Topic Coherence Score

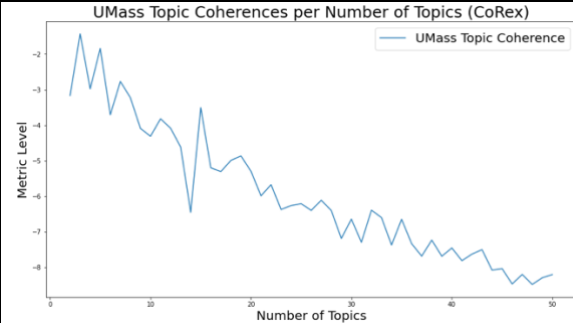


Figure 18

Total Correlation per Number of Topics

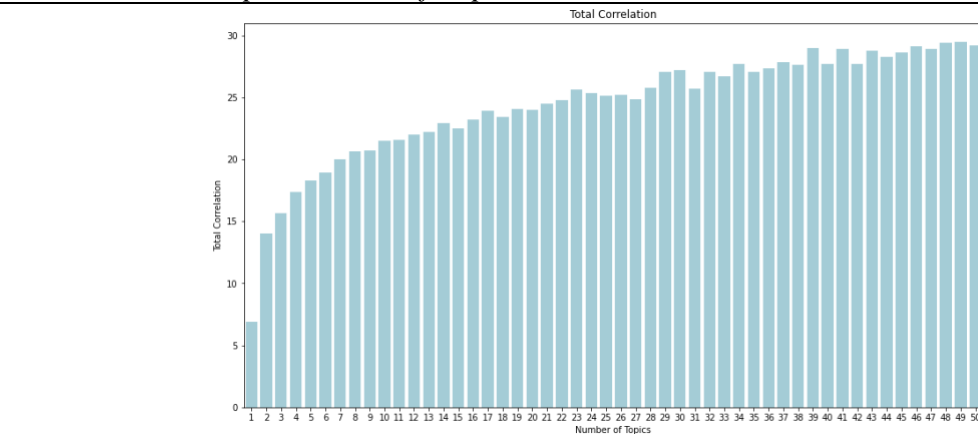


Figure 16 shows the CV topic coherence score, and figure 17 shows the UMass topic coherence score. The CV topic coherence score fluctuates as the number of topics increases, with the score relatively lower at around the number of topics equals 24. The UMass topic coherence score decreases as the number of topic increase. The line starts to flatten when the number of topics is around 20.

Figure 18 shows the total correlation per number of topic, which is a better metric in evaluating CoRex models. A higher total correlation indicates that the model explains more about the dataset. In general, total correlation increases from around 6 to a high of 29 as the number of topics increases. There is a steep increase as we go from one topic to four topics. In balancing all the above metrics, I choose the 23 topics model. It is the point where previous figures of the coherence score start to flatten, and the total correlation has a relatively big increase from the 23 topics model before decreasing again after the 23 topics model. The topics generated using the CoRex topic model technique are generally very different from the other methods. The topics within each model are also different, with no overlap regarding the top occurring words. The CoRex model displays some topics not seen

in other topic models, for example, topic one about uterus-related diseases and blood-related symptoms, topic six about constipation and GI, and topic 21 about activities at home and family members. The topics about the top occurring long COVID are only assigned to a few topics (topics 15, 16, and 17) instead of distributed among all topics like the other topic models.

CoRex 23 Topics Model		
Topic	Labels	Top 10 Keywords
Topic 1	Uterus and blood related symptoms	tank, quite, hysterectomy, result, indicate, hematologist, carry, visual, mood, teh
Topic 2	Job related topics, lung related discomforts	appt, discuss, job, worsen, recommend, pulmonologist, stuffy, response, wayne, physically
Topic 3	Improvement of symptoms	note, improvement, post, progress, virtual, tend, recognize, ultrasound, itching, kid
Topic 4	Time and day	day, hour, take, time, bed, also, night, week, year, never
Topic 5	CT scan, blood	see, scan, show, order, negative, ct, prescribe, blood, give, ekg
Topic 6	Constipation and GI	constipation, include, early, mostly, case, induce, panel, draw, tender, gi
Topic 7	Work, sore throat	feel, go, work, good, bad, get, back, throat, sore, well
Topic 8	Liver	liver, june, stress, cleveland, reduce, fatty, suffer, entire, assume, injury
Topic 9	Home, hospital, clinic	home, admit, hospital, send, discharge, parkview, walk, release, clinic, oxygen
Topic 10	Lung, cancer	need, lung, advise, anti, cancer, restriction, contact, injection, strong, part
Topic 11	Discomfort in specific body parts (leg, hand, neck, foot, arm, etc.)	leg, hand, neck, left, side, foot, arm, face, ear, knee
Topic 12	Life, time	much, life, stand, sit, hold, look, morning, evening, force, bathroom
Topic 13	Normal and increased Symptoms	normal, symptom, food, several, increase, improve, period, second, experience, random
Topic 14	?	weird, big, north, block, colon, frequent, florida, reading, history, chicken
Topic 15	Body ache, fever, fatigue, cough, headache	ache, body, fever, fatigue, cough, headache, sob, chill, sinus, low
Topic 16	Brainfog, joint pain, anxiety, weakness	fog, brain, joint, pain, anxiety, muscle, weakness, dizziness, memory, complete
Topic 17	Taste/smell loss	smell, taste, loss, lose, sense, exhaust, chart, meat, sensitivity, regain
Topic 18	Sleep	come, lot, sleep, eat, fall, able, couple, sometimes, water, almost
Topic 19	Staying late	keep, late, doc, turn, stuff, stay, ask, monday, law, mildly
Topic 20	Diarrhea, vomiting	issue, diarrhea, vomiting, energy, start, find, nerve, decrease, exhausted, flash
Topic 21	Activities at home, family members	monitor, last, significant, thought, husband, pfizer, chemical, minute, cook, pound
Topic 22	Chest, breath	chest, ray, tightness, continue, breath, severe, antibiotic, head, wear, remain
Topic 23	Heart, tiredness	rate, heart, movement, next, check, house, extremely, trouble, tired, talk

Anchored CoRex Topic Model

Next, I apply a semi-supervised variant of CoRex by incorporating domain knowledge through anchor words. The following ten lists of anchor words are passed to train the CoRex model.

1. pain, headache
2. anxiety, mental
3. tired, fatigue
4. breath, shortness
5. diabetes
6. pressure, blood
7. asthma
8. brainfog, brain, memory
9. taste, smell, loss
10. dizziness, vertigo, balance

Figure 19

Total Correlation per Anchor Strength

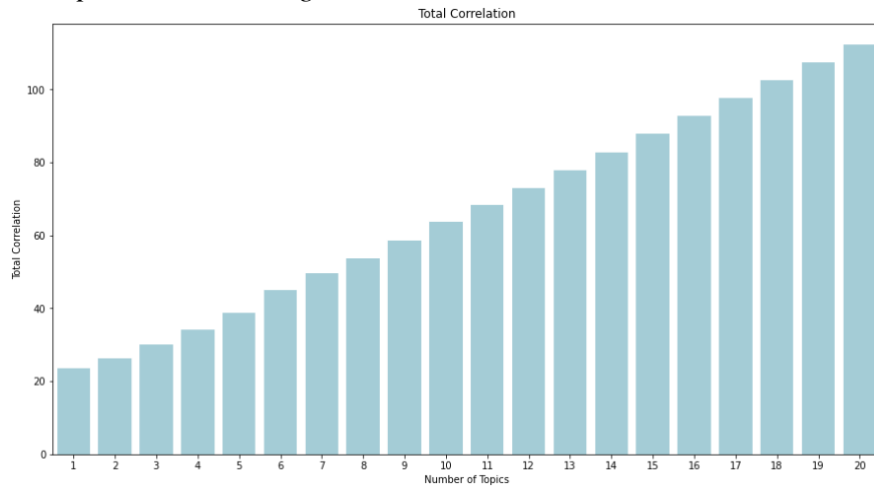


Figure 19 shows the total correlation of the anchored CoRex 10 topic model per anchor strength. The total correlation of the model increases as the anchor strength increase. To get a closer look at the results, below are the top 10 keywords of each topic in the model with anchor strengths 2, 11, and 20. Compared to the unsupervised CoRex models, the semi-supervised models have a much higher total correlation score. The topics presented are also more interpretable. Adding the anchor words allows us to uncover words that are related to our anchor words. An example in the anchor strength = 2 models is topic 1 with the anchor words ‘pain’ and ‘headache’ that gives us other related words ‘chest’, ‘muscle’, ‘throat,’ and ‘joint’. This gives us more idea about what specific parts long COVID patients find discomfort. Topic 6, which aims to find topics related to blood pressure, displays related keywords ‘lung’, blood clot’, ‘sinus’, and ‘liver’. Another interesting example is topic 7 with the anchor word ‘asthma’. While the anchor strength = 2 model is hard to interpret, the other two models give us more helpful information by relating it to ‘allergy’, and we see the word ‘husband’ appearing in all three models.

Anchored CoRex 10 Topics Model (Anchor Strength = 2)	
Topic 1	pain, headache, chest, muscle, throat, cough, sore, joint, fever, nausea
Topic 2	anxiety, year, able, come, never, think, use, long, tell, struggle
Topic 3	fatigue, day, go, much, hour, get, take, feel, week, try
Topic 4	breath, shortness, short, catch, minute, pulmonary, plateau, store, soup, consistently
Topic 5	diabetes, refer, prescribe, cause, recommend, heartbeat, pet, calf, realize, parkview
Topic 6	blood, pressure, show, lung, high, right, side, clot, home, negative
Topic 7	asthma, normal, great, function, husband, stop, law, injection, lay, ent
Topic 8	brain, fog, begin, drive, forgetfulness, tooth, dysfunction, process, internal, blur
Topic 9	smell, taste, loss, lose, sense, chemical, sweet, meat, sensitivity, chart
Topic 10	dizziness, symptom, balance, steroid, office, order, increase, period, lab, random

Anchored CoRex 10 Topics Model (Anchor Strength = 11)	
Topic 1	pain, headache, joint, muscle, chest, throat, abdominal, musclejoint, foot, jaw
Topic 2	anxiety, mental, depression, increase, evening, difficulty, last, recognize, palpitation, intense
Topic 3	fatigue, tired, body, cough, extreme, fever, ache, improve, sleep, sob
Topic 4	breath, shortness, short, catch, minute, pulmonary, teh, breathlessness, liter, hte
Topic 5	diabetes, much, see, day, take, say, feel, go, hour, bed
Topic 6	blood, pressure, clot, high, lung, low, sinus, dehydrate, liver, sugar
Topic 7	asthma, normal, first, make, ct, improvement, allergy, need, full, husband
Topic 8	brain, memory, fog, brainfog, word, forgetfulness, begin, complete, conversation, tooth
Topic 9	smell, taste, loss, lose, sense, chemical, chart, sweet, metallic, regain
Topic 10	dizziness, balance, vertigo, hr, nausea, continue, stand, tend, tingling, turn

Anchored CoRex 10 Topics Model (Anchor Strength = 20)	
Topic 1	pain, headache, joint, muscle, chest, throat, low, abdominal, sore, musclejoint
Topic 2	anxiety, mental, depression, palpitation, medical, worker, panic, intense, concern, induce
Topic 3	fatigue, tired, body, cough, extreme, fever, ache, improve, sob, appetite
Topic 4	breath, shortness, short, catch, pulmonary, minute, flow, teh, afterwards, breathlessness
Topic 5	diabetes, much, see, day, take, get, feel, hour, go, say
Topic 6	blood, pressure, clot, lung, high, dehydrate, liver, sugar, like, xray
Topic 7	asthma, normal, function, first, allergy, rest, husband, improvement, need, check
Topic 8	brain, memory, fog, brainfog, word, forgetfulness, drive, begin, complete, tooth
Topic 9	smell, taste, loss, lose, sense, chemical, sweet, diarrhea, chart, regain
Topic 10	dizziness, balance, vertigo, nausea, clear, start, tingling, friend, suffer, focus

Discussion

Comparison between LDA, NMF, BTM, and CoRex Models

Figure 20

Comparison of Average Number of Overlapping Words per Number of Topics (LDA, NMF, BTM)

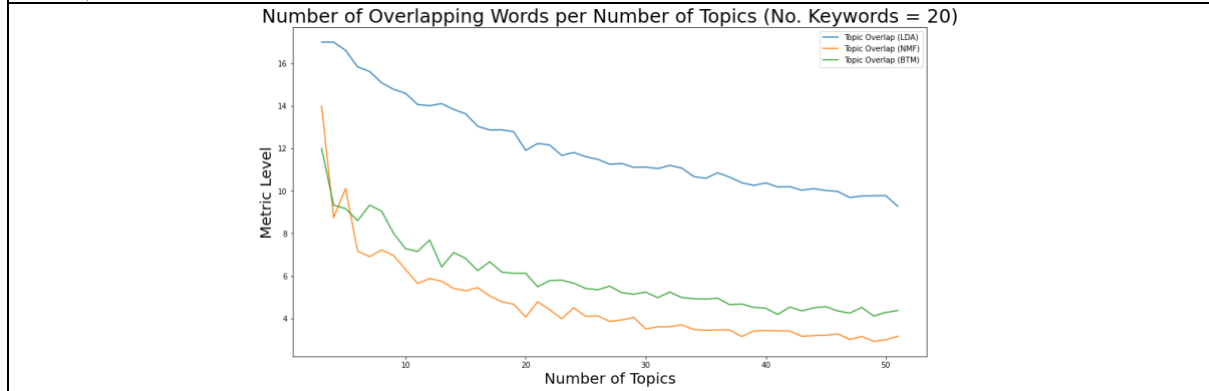


Figure 21

Comparison of Topic Coherence (LDA, NMF, BTM and CoRex)

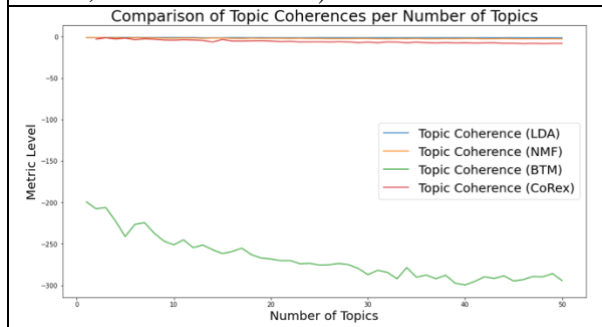
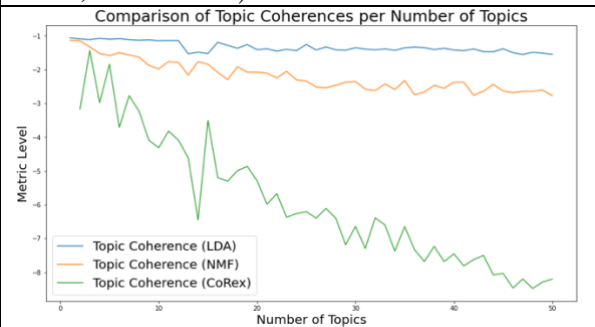


Figure 22

Comparison of Topic Coherence (LDA, NMF, and CoRex)



Average Number of Overlapping Words

Figure 20 shows the comparison of the average number of overlapping words when comparing the top 20 words between topics per number of topics of LDA, NMF, and BTM models. There is no overlap between topics in the CoRex model. The number of overlapping words between topics is consistently higher in the LDA models than in the NMF and BTM models for all number of topics. Even as we increase the number of topics, LDA models still have around half of their top words duplicated. This results in models having topics that are quite similar to each other. Topics generated from the NMF models have the least overlap among the three models, with a high of 14, and most of the models fall within 3-6 common words between the topics. BTM models have a high of 12, and most of the models fall within 5 to 7 common words between the topics.

Coherence

Figure 18 and figure 19 show the topic coherence of LDA, NMF, BTM, and CoRex models. As shown in figure 18, BTM has a much lower coherence score than the other models. LDA and NMF models have a relatively consistent coherence score across the models ranging from around -2 to -1. The coherence score of the CoRex models fluctuates the most. The list

of top keywords within the topics also shows that the unsupervised CoRex model is the most difficult to interpret.

Each Model is Best At

LDA models tend to be affected by the most frequently occurring keywords. In long COVID studies, the most common symptoms of brain fog, headache, fatigue, and smell or taste loss are found in many of the topics in the LDA models. Therefore, this method is good for initial exploration and analysis of a field to learn about the most commonly occurring themes.

However, because of the overemphasis on the frequently occurring topics, many of the topics generated with the LDA topic model technique are very similar. They do not provide any other meaningful topics. This may also be because LDA models tend to be not very effective in shorter texts, which are present in this dataset.

NMF models are less affected by the most frequently occurring words. Even though there are some overlappings in the top one keyword, we can still distinguish between the topics. The models are effective in segmenting the dataset. And we start to observe more hidden topics other than symptoms, such as medications the patients are taking, places they went, and activities they performed in the NMF models.

BTM works well with short texts. The topics derived from this method are the most interpretable. We can observe different groups of long COVID symptoms. As discussed in the previous section, an example is topics with pain as a top-occurring keyword. The model clearly defined four different categories of symptoms related to pain, namely respiratory (topic 1), taste and smell loss (topic 3), fatigue and headache (topic 7), and fatigue and weakness (topic 20). These results are very useful in segmenting long COVID patients and understanding other commonalities among these groups of patients.

CoRex generates very different topics within the model and when compared to the LDA, NMF, and BTM models. However, because they are very different from each other, it is also quite difficult to interpret some of the topics. However, the semi-supervised CoRex models are promising. Adding the anchored words greatly enhanced the models' performance and effectively discovered hidden topics on topics that I am interested in. An example is topic 2, with 'anxiety' and 'mental' as the anchor words. The resulting models give other words for related symptoms like 'depression', 'palpitation', 'panic'. They also pick on time-related themes such as 'evening'. Therefore as we gain more domain knowledge with our study, this method is very useful in helping us to uncover specific themes in the dataset.

Part II: Thematic Analysis

The free-form text responses of CLH patients highlighting their experiences since their COVID-19 infection are diverse and complex. Patients report not only COVID-19 or CLH symptoms but also CLH-related personal, family, social, and life impairments. Interpreting such documents also needs to consider the context of the language used. Following developing the four topic models, this study performs a thematic analysis on the best topic model, the Biterm Topic Model (BTM) with 25 topics. Through a close reading of each topic's keywords and top documents, four dominant themes emerge 1) severity of COVID-19: CLH patients in this study having mild, moderate, and severe illnesses during their COVID-19 infection show different CLH symptoms and post-COVID-19 recovery care needs; 2) CLH symptoms: CLH patients report a diverse range of symptoms 3) health-related quality of life (HRQL): persistent symptoms greatly affect CLH patients' physical, psychological, and social functioning; 4) social determinants of health (SDH): CLH patients continue to struggle in receiving appropriate care, returning to baseline health and securing employment.

Method

Unsupervised topic modeling techniques explored in part I excel in identifying hidden topics within the free text in electronic health records (EHR) of CLH patients. However, interpreting the detected topics using only the top occurring keywords of each topic can be confusing. Keywords are simply words that frequently occur together in documents. There is no guarantee that the keywords are related to one another conceptually. As a result, some words do not align with the general theme we think the topic is suggesting. Some pick up on noise. To better interpret the keywords, we need to understand how the words are used in the dataset and their intended meaning in the context.

This study uses thematic analysis, a method widely-used for organizing, describing and interpreting qualitative data [17]. Through iterations of reading and re-reading the data, we identify patterns across the dataset and derive themes [17]. Building on the best model (BTM with 25 topics) from the previous section, I generate the topic-term matrix to get the top 20 words with the highest probability per topic; and the document-topic matrix to get the top 10 documents for each topic. As I read through the keywords and top documents, I generate codes and themes within each topic and across the entire model.

Top 20 Keywords for BTM 25 topics model:

BTM 25 Topics Model	
Topic	Top 20 Keywords
Topic 1	cough, headache, pain, fatigue, ache, fever, body, chest, smell, sob, start, taste, throat, muscle, sore, symptom, sinus, congestion, loss, feel
Topic 2	go, home, oxygen, hospital, day, breath, heart, increase, rate, send, admit, discharge, taste, pneumonia, release, back, loss, sob, cough, drop
Topic 3	pain, start, see, symptom, appt, go, headache, visit, give, long, breathing, leg, lung, issue, asthma, depression, blood, chest, problem, tell
Topic 4	smell, taste, fatigue, issue, loss, migraine, headache, food, bad, much, start, pain, memory, day, week, first, get, symptom, come, continue
Topic 5	day, get, go, work, take, time, home, hospital, family, blood, heart, well, walk, parkview, feel, fatigue, symptom, low, doctor, trouble
Topic 6	day, get, work, go, feel, start, much, home, good, fatigue, bad, sleep, body, walk, week, difficulty, breathing, back, bed, extreme

Topic 7	see, rec, asthma, wue, lab, ent, order, sinus, headache, ct, ddx, worsen, level, dupixent, sid, first, sleep, sharaan, possible, echo
Topic 8	fatigue, pain, loss, cough, brain, taste, headache, smell, chest, muscle, sob, fog, start, memory, fever, short, joint, symptom, appetite, breath
Topic 9	pain, chest, cough, visit, headache, sob, increase, symptom, high, pneumonia, severe, steroid, hospital, see, parkview, inhaler, lung, bp, admit, hr
Topic 10	headache, symptom, cough, pain, fatigue, fever, sob, day, ache, chest, continue, body, night, diarrhea, nausea, go, fog, work, chill, sweat
Topic 11	feel, get, antibody, pain, taste, go, day, smell, lose, blood, body, loss, negative, good, time, normal, energy, current, diagnose, vaccine
Topic 12	get, week, take, help, back, go, place, day, home, doctor, work, shot, month, long, vent, number, note, headache, continue, worsen
Topic 13	cough, go, back, breath, headache, fatigue, day, fever, taste, smell, come, pain, work, body, feel, chest, week, get, symptom, negative
Topic 14	go, work, start, much, walk, get, day, take, tired, night, bad, time, home, hour, make, shake, inhaler, see, pt, back
Topic 15	pain, right, leg, left, head, back, side, leave, arm, neck, numbness, chest, feel, start, foot, tingle, hand, ear, blood, heart
Topic 16	normal, blood, pain, heart, see, visit, chest, liver, show, ct, continue, result, diagnose, hospital, follow, symptom, order, increase, prednisone, pneumonia
Topic 17	fog, brain, pain, fatigue, breath, issue, headache, start, memory, heart, shortness, anxiety, low, sleep, loss, dizziness, high, pressure, chest, symptom
Topic 18	smell, taste, fatigue, brain, fog, symptom, headache, day, body, get, feel, bad, ache, severe, back, sleep, much, diagnose, pain, time
Topic 19	feel, get, much, day, time, work, go, bad, able, good, start, symptom, also, sleep, normal, activity, walk, fatigue, life, hour
Topic 20	fatigue, brain, headache, symptom, loss, fog, smell, start, dizziness, taste, continue, muscle, day, issue, severe, nausea, begin, resolve, bad, pain
Topic 21	pain, leg, start, bad, back, go, foot, much, also, headache, arm, loss, knee, leave, get, left, head, hair, eye, neck
Topic 22	symptom, see, day, cough, pulmonary, issue, sob, headache, bad, get, ear, asthma, function, go, start, inhaler, help, np, feel, steroid
Topic 23	smell, day, taste, week, start, symptom, headache, work, continue, back, much, feel, time, brain, pain, go, vaccine, fatigue, return, fog
Topic 24	smell, taste, day, loss, headache, fatigue, lose, start, begin, food, come, time, resolve, continue, back, thing, return, symptom, notice, fog
Topic 25	pain, joint, headache, fog, muscle, brain, fatigue, bad, cough, chest, ache, taste, sob, day, start, weakness, smell, go, body, fever

Example of Coding Top Documents (Taken from Topic 2)

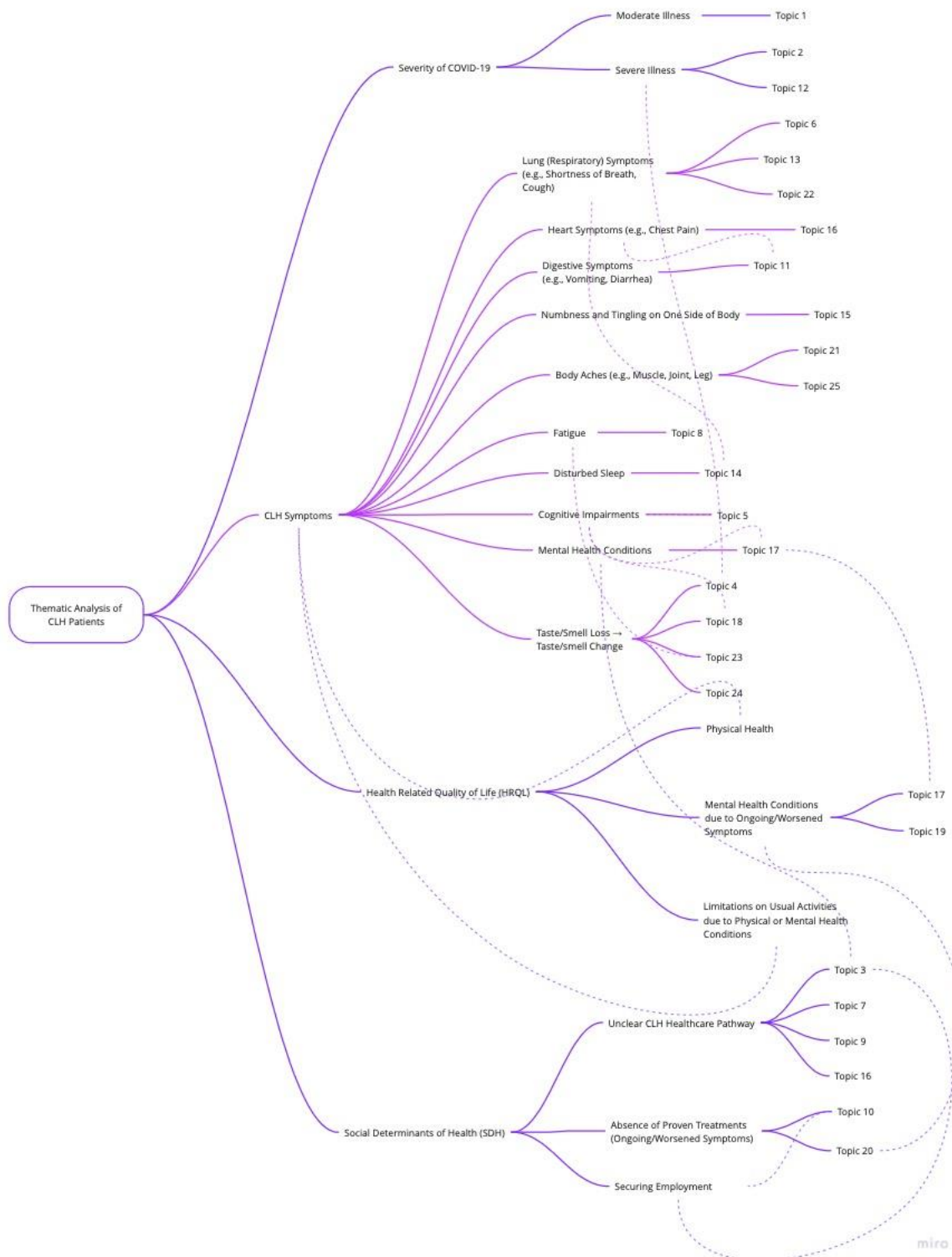
Patient's Journey/Major Events	Codes/Themes
10/7/21: tested positive for covid 10/8/21: went to hospital for infusion and paced on oxygen, admitted 10/21/21: released to inpatient rehab for weakness 11/3/21: released to home health on 2L oxygen 11/23/21: discharged from home health, still on oxygen current: on 2L of oxygen at all times	<ul style="list-style-type: none"> - Severe illness during COVID-19 infection <ul style="list-style-type: none"> o Hospitalized o Low blood oxygen - CLH symptoms <ul style="list-style-type: none"> o Weaknesses o Continuous low blood oxygen - Post-COVID-19 recovery care <ul style="list-style-type: none"> o Inpatient rehab o Home health

Results

The analysis produces four major themes:

1. **Severity of COVID-19:** CLH patients in this study having mild, moderate, and severe illnesses during their COVID-19 infection show different CLH symptoms and post-COVID-19 recovery care needs
2. **CLH symptoms:** CLH patients report a diverse range of symptoms (e.g. neurological symptoms, mental health conditions, taste/smell change, lung symptoms, body aches, etc.)
3. **Health-related Quality of Life (HRQL):** Persistent CLH symptoms adversely affect the patients' quality of life. In particular, physical functioning, pain interferences, sleep disturbance, fatigue, mental health conditions and cognitive impairments place a great burden on patients' physical, psychological, and social life.
4. **Social determinants of health (SDH):** CLH patients continue to struggle in receiving appropriate care, returning to baseline health, and securing employment, all of which further influence the patients' health outcomes

Below is a mind map showing how each topic corresponds to the above themes.



Discussion

Severity of COVID-19

COVID-19 affects people differently. As defined in U.S. National Institutes of Health's COVID-19 treatment guidelines, COVID-19 patients can be grouped into having an asymptomatic or presymptomatic infection; mild illness; moderate illness; severe illness, and critical illness during their COVID-19 infection [15]. Several topics in this model identify patients with various levels of severity of COVID-19.

Patients with mild illness are not grouped in a particular topic but are distributed throughout the topics. These patients can experience various signs and symptoms but usually do not have more severe symptoms such as shortness of breath, dyspnea, or abnormal chest imaging [1]. They mainly report experiencing cough, headache, muscle pain, and low or no fever. They are not admitted to the hospitals or placed in specific follow-up care as they recover from the COVID-19 infection.

Topic 1 identifies patients with moderate illness when having COVID-19. Most of the patients in this group experience fever, shortness of breath, chest congestion, and tightness in the acute phase of COVID-19. Many of these patients continue to experience shortness of breath, chest pain and chest pressure post-infection. Other CLH symptoms for the group are upper respiratory symptoms including sinus infection, sinus pressure, sinus congestion, coughing, mouth sores, and sore throat; body aches; headache; and fatigue. Several of them make multiple emergency room visits during and after their COVID-19 infections.

Topics 2, 4, and 12 identify patients with severe illnesses during their COVID-19 infection. These patients are admitted to the hospital or even the ICU due to low blood oxygen. One patient was even hospitalized for 30 days. Many of them need to use ventilators. Some of them develop pneumonia and experience blood clots, heart attacks and coma while hospitalized. Such severe illnesses usually end up making the patients feel very weak and exhausted with limited mobility. Shortness of breath is also a prevalent CLH symptom among these patients. These patients are often sent to 'home health', 'home healthcare', 'nursing home', or other home therapies post-COVID-19.

Patients with different level of severity of COVID-19 report different CLH symptoms. They also show different needs for recovery and rehabilitation resources. Some make multiple emergency visits, some need to be hospitalized for an extended period, and some require additional long-term follow-up care post-COVID-19. With the healthcare system already strained after years of the pandemic, there is a need for better health system management for different groups of patients to mitigate the burden of post-COVID-19 condition.

CLH Symptoms

CLH involves a wide range of new, returning, or ongoing health problems. The list of CLH symptoms found among this group of CLH patients are:

- Lung (respiratory) symptoms (Topic 6, 13 and 22)
 - o shortness of breath, cough, asthma, allergy
- Cardiovascular Symptoms (Topic 16)
 - o chest tightness, chest pain, increased heartbeat, irregular heartbeat
- Digestive symptoms (Topic 11)
 - o vomiting, diarrhea, constipation, high blood sugar
- Numbness, tingling, swelling and tremor on one side of body (Topic 15, 21)
 - o Upper body (arm and hand) and lower body (leg, feet and tow)
 - o Painful vein
 - o Blood clot
- Body aches (Topic 21 and 25)
 - o Neck, shoulder, back, hand, hip, leg, muscle, joint
 - o Weak muscle, joint swelling and losing flexibility
- Fatigue (Topic 8)
- Neurological symptoms (Topic 5, 17, 18)
 - o Cognitive dysfunction: brain fog, brain confusion, memory loss, difficulty concentrating, lack of focus, forgetfulness, mixing words and speech
 - o Headache
 - o Burning sensation on brain
- Mental health conditions (Topic 17)
 - o Anxiety, depression, sadness
 - o Lack of motivation, loss of appetite
 - o Panic attack, weeping, laughing and fight episodes
 - o Disturbed sleep and sleep problems
- Taste/smell loss and taste/smell change (Topic 4, 18, 23 and 24)
 - o Taste/smell loss turn to taste/smell change
 - o Parosmia (distorted smell and taste): taste of dirty dish, food patients used to like smell different now
 - o Increased sensitivity to certain smell and cannot tolerate certain food smell
 - o Phantosmia (phantom smell): bad smell, chemical smell, smell of gasoline, dog poop and burnt skin
 - o Irritation and burning sensation in throat, nose and eyes
- Symptoms that get worse after physical or mental effort

Health-related Quality of Life (HRQL)

Beyond the list of CLH symptoms, this study provides insights into how the symptoms above affect the patients' overall HRQL. HRQL is defined as health-related conditions that affect patients' quality of life. It includes the patient's general health status, physical health, mental health, and the limitations in performing everyday functions due to their physical and mental health condition [5,19]. CLH patients having to suffer a decreased quality of life due to CLH symptoms is an essential indicator of CLH having a heavier disease burden than other influenzas-like diseases [1,5].

Physical Health

Many of the CLH symptoms are physical discomforts. This includes respiratory symptoms, cardiovascular symptoms, digestive symptoms, numbness and tingling of one side of the body, body aches, fatigue, headache, cognitive impairments, taste/smell loss, and taste/smell change. As much research shows an impact of physical health on patients' quality of life [8,12,16], CLH patients continuing to experience these physical dysfunctions for an extensive period after the initial illness can significantly impact their quality of life. What's worse is that they sometimes also need to face worsening symptoms.

Mental Health

One recurring theme in this study is CLH patients reporting mental health conditions throughout their journey in fighting COVID-19 and CLH. Common CLH mental health conditions include anxiety, depression, sadness, stressfulness, loss of appetite, loss of motivation, feeling lost, weeping, laughing and fighting episodes, and sleep problems. These conditions can be neurological as well as psychological [11]. The bi-directional relationship between the physical and mental symptoms can further complicate recovery, creating a vicious cycle. Adding together these mental health conditions can hugely influence one's perception of their quality of life, sometimes even more so than physical discomforts [6].

Limitations in Performing Normal Activities

CLH patients also report their daily life affected by the above physical and mental health conditions in several topics.

Lung (Respiratory) Symptoms. Patients can sometimes experience difficulty breathing for no reason (topic 6). A patient mentions having an asthma attack after returning to work (topic 13).

Numbness, Tingling, Swelling and Tremor on One Side of Body. Losing some or all sensations in the affected part of the patients' body can greatly affect their ability to perform the daily task as they may not be able to feel light touch, pain, temperature, or vibration [20].

Pain Interference. Lower body joint and muscle pain affect patients' ability to walk. Abdominal pain makes patient vomit (topic 25). Severe pain can also interfere with sleep.

Fatigue. With fatigue, patients cannot perform tasks they are once able to. Even light physical activities like hanging Christmas ornaments can leave patients exhausted (topic 8). Some find themselves cannot do hours of work (topic 19). Some patients experience weak

muscles and lose flexibility (topic 21 and 25). Some find it challenging to stay awake (topic 8). These affect patients' ability to perform daily tasks and spend time with family (topic 19). This leaves the patients feeling lost, which can sometimes also worsen their mental conditions.

Taste/Smell Change. Taste/smell change can in some ways be more debilitating than loss of taste/smell. Distorted taste/smell and increased sensitivity to certain smells affect what kind of food the patients can eat (topic 11), which often leads to poor appetite. Patients with phantom smell often describe unpleasant odors (topic 4 and 18). All of these can be incredibly debilitating and depressing experiences for CLH patients.

Disturbed Sleep. CLH patients can also suffer from disturbed sleep due to other physical and mental symptoms, nervousness, and nightmares. Such a condition makes the patient even more tired and can potentially worsen their CLH symptoms.

Cognitive Impairments. Brain fog and difficulty focusing make it hard for patients to make decisions. Mixing word and speech requires the patients to formulate speech and maintain a conversation (topic 5). One patient even reported being let go of work due to their brain fog and memory issues (topic 17).

Social Determinants of Health (SDH)

The role of SDH in CLH is another crucial theme from this analysis. SDH is defined as non-medical factors influencing health outcomes, such as social, economic, and political factors [21]. Some researches show that SDH can account for between 30-55% of health outcomes [21]. Three years into the pandemic, the cause, symptom patterns, and patient profile associated with long COVID remain challenging to isolate. Such a lack of clear definitions and guidelines for treating long COVID affects the doctor's ability to care for patients and the patients' ability to get healthcare and social support for their new disabilities. And the unclear CLH healthcare pathway, absence of proven treatment, and lack of job security all become key factors in complicating and worsening the recovery from CLH.

Unclear CLH Healthcare Pathway

Topics 3, 7, 9, and 16 highlight patients' journeys in seeking appropriate care for treating their persistent symptoms. Many patients are left to navigate the healthcare system on their own. They report the need to make multiple appointments and get referred to various doctors. The process usually involves patients being tested for specific symptoms. For instance, a patient in topic three reports taking chest and chest CT scans to detect possible blood clots and a lung CT scan for pulmonary functionality and was told to do breathing treatments. These tests and follow-up appointments are with different doctors and on different days. Another patient in topic 22 mentions the journey of being referred to a pulmonologist for a pulmonary function test, then doing a colonoscopy with another doctor, cardiology with another, seeing a urologist, a spine specialist, and endocrinology. All of the above appointments are with different doctors and are done on separate trips. Patients in topic 16 mention having multiple ER visits and blood tests since COVID-19. Other tests patients need to go through include:

- Pulmonary function test
- Chest CT scan (for acute chest disease)
- Echocardiogram

- Cardiac MRI
- Endoscopy
- Abdominal and pelvic CT
- Colonoscopy
- MR Cholangiopancreatography (MRCP) for evaluating the liver, gallbladder, bile ducts, pancreas, and pancreatic duct
- Lip biopsy for Sjogren's syndrome
- Epstein-Barr virus (EBV) test
- Brain MRI
- Stress test
- CSF IgG index (for measuring the levels of IgG in the cerebrospinal fluid)

There are no specific orders for taking these tests among the patients. While some patients report positive results in certain tests, many report 'normal' or 'negative' results for the above tests. We can see from these examples that the lack of an established healthcare pathway for CLH patients leaves them having to wait between appointments and go through multiple doctor visits to look for possible explanations for their discomfort. This often results in lengthy diagnostic odysseys that can complicate and worsen the patients' already debilitating symptoms.

Absence of Proven Treatments

Many patients also discuss the frustration of seeing no improvements in their CLH symptoms. Some patients also experience symptoms fluctuating throughout the day and from day to day. Worse, CLH patients often experience symptoms 'coming back' (topic 13), so they have to keep returning to the doctors for treatments. Topic 23 recognizes patients turning to essential oils in the hope of treating their distorted taste/smell. However, the complexity and variety of symptoms experienced by each patient make it extremely difficult to create a treatment plan, leaving patients with an unclear timeline for recovery. At the same time, topic 20 identifies patients reporting mental health problems from feeling lost, anxious, stressed, and fearful that they will never resolve the symptoms and will not be able to return to pre-COVID conditions. And that can have a significant adverse effect on CLH patients' conditions.

Employment

In some topics, patients discuss their concerns about keeping and returning to work post-COVID-19. In topic 10, a patient has to switch jobs because returning to work with fatigue and brain fog makes it impossible to continue with the previous position. Topic 19 highlights a couple of cases where the patient mentions being unable to catch his breath, and the symptom worsens at work. One patient reported not being able to do hours of work due to poor brain functions and low energy. Another patient reports feeling overwhelmed when returning to work which makes depression worse. One patient feels awful when returning to work. Those working in a more demanding job, such as a social worker, also feel emotionally and physically exhausted having to work while dealing with CLH symptoms. All of the above examples show a lack of support in enabling the patients to return to the workplace, creating additional stress as the patients are fighting the CLH symptoms.

Conclusion

CLH creates a surge of patients with chronic medical conditions and disabilities, bringing about new public health challenges. Incorporating thematic analysis into the study facilitates a well-balanced interpretation of the underlying themes of the free-form text responses of CLH patients. Apart from the list of CLH symptoms, the study is able to uncover the underlying personal, social, and economic implications of CLH, indicating that solving the CLH problem means more than treating the individual symptoms. The severity of COVID-19 affects the healthcare resources needed to treat individual patients. CLH symptoms affecting patients' physical health, mental health, and ability to perform everyday tasks can largely impact the patients' quality of life. And lastly, the role of social determinants of health in CLH calls for attention to the CLH healthcare pathway, treatment plans, and supports for CLH patients to return to work. These insights gained from the patients' perspective are valuable when developing therapies, reducing delays in patients receiving optimal treatments, and resource planning.

References

- [1] Atefeh Aghaei, Ran Zhang, Slone Taylor, Cheuk-Chi Tam, Chih-Hsiang Yang, Xiaoming Li, and Shan Qiao. 2022. Impact of persistent COVID-19 symptoms on social life of female long haulers: A qualitative study. 2022.01.21.22269671. DOI:<https://doi.org/10.1101/2022.01.21.22269671>
- [2] Olalekan Lee Aiyegbusi, Sarah E Hughes, Grace Turner, Samantha Cruz Rivera, Christel McMullan, Joht Singh Chandan, Shamil Haroon, Gary Price, Elin Haf Davies, Krishnarajah Nirantharakumar, Elizabeth Sapey, and Melanie J Calvert. 2021. Symptoms, complications and management of long COVID: a review. *J. R. Soc. Med.* 114, 9 (September 2021), 428–442. DOI:<https://doi.org/10.1177/01410768211032850>
- [3] Moumita Bhattacharya, Claudine Jurkovitz, and Hagit Shatkay. 2017. Identifying Patterns of Associated-Conditions through Topic Models of Electronic Medical Records. Retrieved November 20, 2022 from <http://arxiv.org/abs/1711.10960>
- [4] Felicity Callard and Elisa Perego. 2021. How and why patients made Long Covid. *Soc. Sci. Med.* 268, (January 2021), 113426. DOI:<https://doi.org/10.1016/j.socscimed.2020.113426>
- [5] Kathleen R. Case, Chen-Pin Wang, Meredith G. Hosek, Sarah F. Lill, Alexandra B. Howell, Barbara S. Taylor, James Bridges, Daniel J. MacCarthy, Paula Winkler, and Joel Tsevat. 2022. Health-related quality of life and social determinants of health following COVID-19 infection in a predominantly Latino population. *J. Patient-Rep. Outcomes* 6, 1 (June 2022), 72. DOI:<https://doi.org/10.1186/s41687-022-00473-8>
- [6] Janice Connell, Alicia O’Cathain, and John Brazier. 2014. Measuring quality of life in mental health: Are we asking the right questions? *Soc. Sci. Med.* 1982 120, (November 2014), 12–20. DOI:<https://doi.org/10.1016/j.socscimed.2014.08.026>
- [7] Diana Duong. 2022. The evolving picture of long COVID. *CMAJ* 194, 24 (June 2022), E850–E851. DOI:<https://doi.org/10.1503/cmaj.1096004>
- [8] John T Farrar. 2019. Pain Interference on Quality of Life Is Not Just About Physical Function. *Pain Med.* 20, 2 (February 2019), 211–212. DOI:https://doi.org/10.1093/pm/pny147_2
- [9] Trisha Greenhalgh, Matthew Knight, Christine A’Court, Maria Buxton, and Laiba Husain. 2020. Management of post-acute covid-19 in primary care. *BMJ* 370, (August 2020), m3026. DOI:<https://doi.org/10.1136/bmj.m3026>

- [10] Trisha Greenhalgh, Rosamund Snow, Sara Ryan, Sian Rees, and Helen Salisbury. 2015. Six ‘biases’ against patients and carers in evidence-based medicine. *BMC Med.* 13, (September 2015), 200. DOI:<https://doi.org/10.1186/s12916-015-0437-x>
- [11] E. Guedj, J. Y. Campion, P. Dudouet, E. Kaphan, F. Bregeon, H. Tissot-Dupont, S. Guis, F. Barthelemy, P. Habert, M. Ceccaldi, M. Million, D. Raoult, S. Cammilleri, and C. Eldin. 2021. 18F-FDG brain PET hypometabolism in patients with long COVID. *Eur. J. Nucl. Med. Mol. Imaging* 48, 9 (2021), 2823–2833. DOI:<https://doi.org/10.1007/s00259-021-05215-4>
- [12] Muhammad Abdul Hadi, Gretl A. McHugh, and S. José Closs. 2019. Impact of Chronic Pain on Patients’ Quality of Life: A Comparative Mixed-Methods Study. *J. Patient Exp.* 6, 2 (June 2019), 133–141. DOI:<https://doi.org/10.1177/2374373518786013>
- [13] Uday Jain. 2020. Effect of COVID-19 on the Organs. *Cureus* 12, 8 (August 2020). DOI:<https://doi.org/10.7759/cureus.9540>
- [14] Sandra Lopez-Leon, Talia Wegman-Ostrosky, Carol Perelman, Rosalinda Sepulveda, Paulina A Rebolledo, Angelica Cuapio, and Sonia Villapol. 2021. More than 50 Long-term effects of COVID-19: a systematic review and meta-analysis. *medRxiv* (January 2021), 2021.01.27.21250617. DOI:<https://doi.org/10.1101/2021.01.27.21250617>
- [15] National Institutes of Health. 2022. Clinical Spectrum of SARS-CoV-2 Infection. *COVID-19 Treatment Guidelines*. Retrieved December 14, 2022 from <https://www.covid19treatmentguidelines.nih.gov/overview/clinical-spectrum/>
- [16] M. Craça Pereira, Cátia Carvalho, Eleonora C. V. Costa, Ângela Leite, and Valéria Almeida. Quality of life in chronic pain patients: Illness- and wellness-focused coping as moderators. Retrieved December 14, 2022 from <https://onlinelibrary.wiley.com/doi/abs/10.1002/pchj.410>
- [17] Gareth Terry, Nikki Hayfield, Victoria Clarke, and Virginia Braun. 2017. *Thematic Analysis. The SAGE Handbook of Qualitative Research in Psychology*. SAGE.
- [18] Yuesong Zou, Ahmad Pesaranghader, Ziyang Song, Aman Verma, David L. Buckeridge, and Yue Li. 2022. Modeling electronic health record data using an end-to-end knowledge-graph-informed topic model. *Sci. Rep.* 12, 1 (October 2022), 17868. DOI:<https://doi.org/10.1038/s41598-022-22956-w>
- [19] 2018. HRQOL Concepts | CDC. Retrieved December 14, 2022 from <https://www.cdc.gov/hrqol/concept.htm>
- [20] Numbness - Brain, Spinal Cord, and Nerve Disorders. *Merck Manuals Consumer Version*. Retrieved December 14, 2022 from <https://www.merckmanuals.com/en-ca/home/brain,-spinal-cord,-and-nerve-disorders/symptoms-of-brain-spinal-cord-and-nerve-disorders/numbness>
- [21] Social determinants of health. Retrieved December 11, 2022 from <https://www.who.int/health-topics/social-determinants-of-health>