Text Analysis and Uncovering Demographic Differences in Family Caregiver Experiences

May Khine, Andrea Robang, Mark Lekina

QSS 20. Final Class Status Update. 03.07.2022

Outline

- ▶ Motivation
- ▶ Research questions
- ▶ Data
- ▶ Methods
- ▶ Results
- ▶ Next steps

Motivation

- ▶ **Goal**: To uncover demographic differences in family caregiver experiences during the COVID pandemic.
- ▶ We want to identify common themes and changes across different demographics w.r.t. responses about mental health needs and existing services.
- ▶ By analyzing variation between demographic groups, we can potentially determine:
    - ▶ whether (and how) peoples' needs and opinions of mental health services change;
    - ▶ whether their requests have been met, or whether their suggestions have been acted upon.

Research questions

- ▶ What keywords and themes are most prevalent in responses from individuals with IDD and their families highlighting still-standing mental health needs and services?
- ▶ How do they vary with racial groups, age, and nature of disability over the course of the pandemic?

Data sources

- ▶ *Primary data source*: START Family Experience Interview Survey (FEIS)
- ▶ Tracks the responses of participants with START activity and their evaluation of mental health services.
- ▶ Emphasis was placed on the last two open-ended columns:
    - ▶ ...any particular service that your family member needed that was not available?
    - ▶ What advice would you give to service planners regarding the mental health service needs of persons with IDD and their families?
- ▶ *Secondary data source*: Dartmouth Data Set (January 2019 to December 2020)
- ▶ We intend to use the data on race, age, and type of disability to find connections between different demographic groups and their given responses.

Data sources

Respondent ID # (SIRS Local ID)	In the past year, how difficult has it been caring for your family member?	Over the past year, how much information did you receive from your family member's mental health professionals (psychiatrist, therapist, case manager, etc.) regarding his/her illness?	How much assistance did you get from mental health professionals regarding what to do if there were to be a crisis involving your family member?	How much information did you get from mental health professionals regarding whom to call if there were to be a crisis involving your family member?	How much were you encouraged by mental health professionals to take an active role in your family member's outpatient treatment?	During the past year, how much did mental health professionals respond to your concerns about your family member?	How much did mental health professionals take into account your ideas and opinions about your family member's treatment?	When you had your family member's permission, how much did mental health professionals involve you in his/her treatment?	During the past year, how much did outpatient service providers recognize the burdens that family members like you face?	...

Methods: data acquisition

▶ All the data we used for the project was provided through the class, which in turn was acquired through semi-structured interviews with caregivers of participants conducted by START using the *Family Experiences with Severe Mental Illness Scale*.

▶ *Overall survey* : 1940 rows, 57 columns
Open-ended questions : 2/57 (1 conditional)
1st question (service) : 1239 answered Yes, 658 answered No (65%)
2nd question (advice) : 915 answers (usable)

Methods: data cleaning

▶ Reformatting column names for easier reference
▶ Splitting the dataset into scaled and open-ended columns
▶ Removing stopwords, non-alphabetical tokens, and tokens with less than 4 characters

Reformatting column names for easier reference

```python
1  # Replace pattern with apostrophe (')
2  data.columns = data.columns.astype("string")
3  data.columns = data.columns.str.replace(pattern, "'")
4
5  # Replace "\n" with a space(" ")
6  data.columns = data.columns.str.replace("\n", " ")
```

Splitting the dataset into scaled and open-ended columns

```
1 # define columns in each category
2 scaled_cols =
      [0,11,25,26,27,28,29,30,31,32,35,36,37,38,39,42,43,44,45
3 open_cols = [0,54,55,56]
4
5 # rated questions
6 scaled_data = data.iloc[:, scaled_cols].copy()
7
8 # open ended question
9 open_data = data.iloc[:, open_cols].copy()
```

Removing stopwords, non-alphabetical tokens, and tokens with less than 4 characters

```python
1  ## Convert advice column to lowercase and a list
2  service_advice = open_data_need.Advice.str.lower().
       to_list()
3
4  # call function to preprocess each response
5  cleaned_listings = np.array([process_step(one_listing)
       for one_listing in
6                          service_advice])
7
8  # ## Add processed strings back to dataframe
9  open_data_need["proc_advice"] = cleaned_listings
```

Data cleaning

	ID	In_Need	Service	Advice
1	1903	Yes	EMERGENCY RESPITE SERVICES & CRISIS INTERVENTION	provde referrals for emergency respite services; job support for id consumers
5	6744	No	NaN	resommendations pertinent to treatment
7	8799	Yes	would liek to have a medication eval	would like consultation with family before announcing prognosis to consumer
9	10045	Yes	community habilitation/big brother	nothing. they are doing the best that they can. it would be good for him to get out of the house more. we need help with transportation to help us get back into respite.
10	10063	Yes	play therapy. it is now available but took some time.	take advantage of all that is avaialble
...	...	...	...	...
1933	unknown	No	NaN	i think they do a good job if the person with the disability lets them.
1934	unknown	No	NaN	they are doing a good job.
1936	unknown	Yes	respite	medications may be helpful, keeping people busy helps a lot!
1937	unknown	Yes	Respite, after school support, psychiatric support, and in-home supports	take the time to get to know and give him a chance
1939	unknown	Yes	Respite services.	to have better access to the information and places that provide services

Methods: analysis or visualization

▶ Topic modelling - *Latent Dirichlet Allocation*(LDA) model
▶ Parameters:
  ▶ num_topics = 4
  ▶ passes = 6
  ▶ alpha = 'auto'
  ▶ per_word_topics = True

Results: Topics and associated common words

▶ As expected, we did not find significant variation between common words in each topic, i.e. same common words with different frequencies across topics.

▶ This is because we haven't compiled a list of stopwords specific to these responses.

Results: Topics and associated common words

topic	top_words
0	[provid, support, servic, famili, need, avail]
1	[famili, avail, servic, support, need, provid]
2	[need, famili, provid, servic, avail, support]
3	[servic, need, famili, provid, support, avail]

Next steps

- ▶ Compile stopwords that are specific to the open ended questions in the FEIS dataset
- ▶ Further topic modeling of the columns of interest mentioned above
- ▶ Possibly incorporate TF-IDF (term frequency-inverse document frequency) into topic modeling
- ▶ Incorporate demographic data from the Dartmouth Data Set, i.e link back results from topic modeling to demographic differences
    - ▶ Merge the two datasets by ID
    - ▶ Find top words from different subsets of demographic data
- ▶ Visualizing the results