

DISTINGUISHING ALZHEIMER'S AND DEMENTIA COMMUNITIES BASED ON MEDHELP HEALTH FORUMS

Kristen McGarry

LHS 712: Final Project Presentation

Dementia



“A GENERAL TERM FOR A DECLINE IN MENTAL ABILITY SEVERE ENOUGH TO INTERFERE WITH DAILY LIFE.”

In 2012, the Dementia rates in Americans 65+ ~ 8.8%

Types of Dementia:

1. Alzheimer's Disease
2. Vascular Dementia
3. Lewy Body Dementia
4. Frontotemporal Dementia

“A NUMBER OF SYNDROMES CHARACTERIZED BY DIVERSE BEHAVIORAL, COGNITIVE, AND EMOTIONAL IMPAIRMENTS”

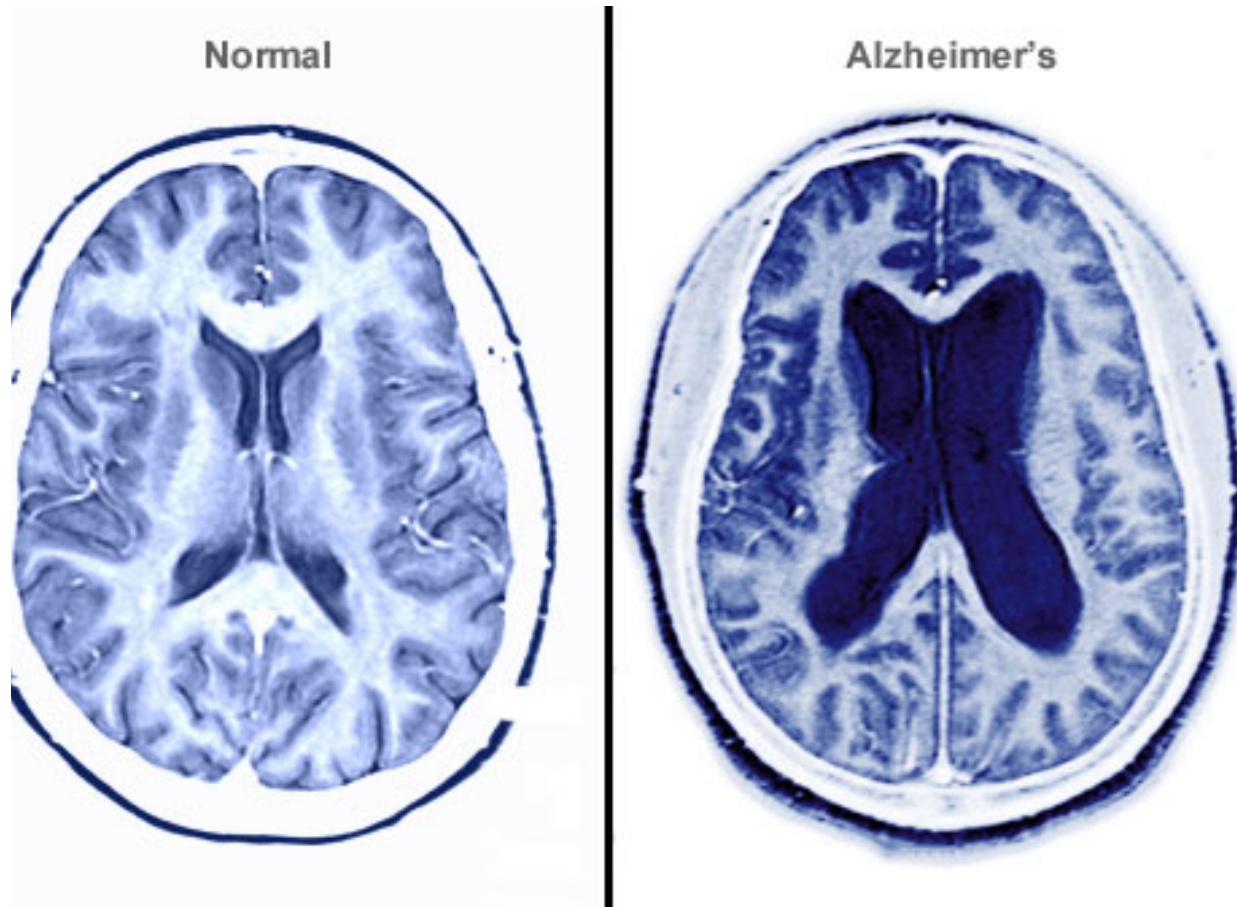
References:

1. [IMAGE]<https://www.psychologytoday.com/blog/the-distracted-couple/2014/11/adhd-and-dementia-what-s-the-connection>
2. Alzheimers.net, 2017

Alzheimer's

Every 66 seconds,
someone develops
Alzheimer's disease

6th Leading Cause of
Death



“A TYPE OF DEMENTIA THAT CAUSES PROBLEMS WITH MEMORY, THINKING AND BEHAVIOR. SYMPTOMS USUALLY DEVELOP SLOWLY AND GET WORSE OVER TIME, BECOMING SEVERE ENOUGH TO INTERFERE WITH DAILY TASKS.”

**IN THE US MEDICARE, MEDICAID AND OUT-
OF-POCKET EXPENDITURES FOR
ALZHEIMER'S AND OTHER DEMENTIA CARE
TOTAL \$200 BILLION (2002)**

ALZHEIMER'S ↔ DEMENTIA

“Is Alzheimer Disease and dementia the **same thing?**”

–MedHelp Forum Question

“From what the Drs. told us , at this time there is no sign of Tumor or Seizures **or- they say- Dementia/Alzheimers.**”

- MedHelp Forum Question

“I get the impression from what I have read on the internet that **dementia is general catchall word** for debilitating mental disorders. Alzheimers, PTSD, etc. Until a specific disorder is diagnosed, they tend to call it dementia”

– MedHelp Forum Comment

Study Question:

What are the similarities and differences between Alzheimer's and dementia online health communities, given that these terms are often used interchangeably or not easily distinguished?

DATA SOURCE: MEDHELP HEALTH FORUMS

Benefits of Online Health Forums:

- An abundance of information available online
- Bridges the medical literacy gap by facilitating discussion at a non-medical level
- Connections between people with similar situations
- Builds a support system
- Allow researchers to go directly to the people

Forums:

- Dementia
- Alzheimer's
- Genetics
- Memory Loss
- Caregiver



METHODS

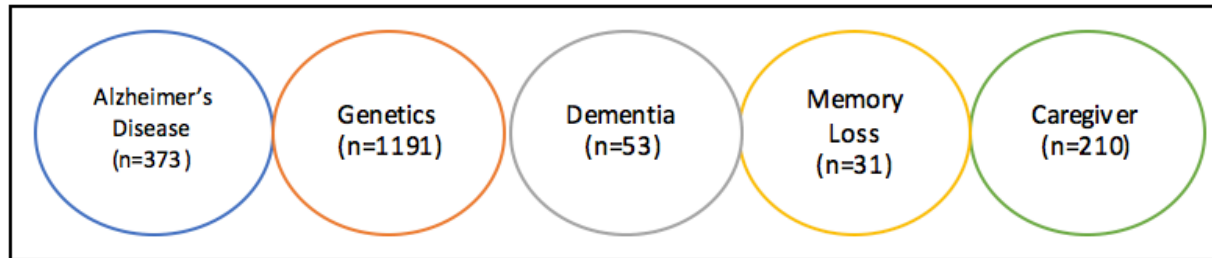
1. Process Data from MedHelp Health Forums into Dementia, Alzheimer's, and Both categories
2. Analyze Word Frequencies
 - ✓ Attempt 1: Questions Only
 - ✓ Attempt 2: Questions & Comments
3. Bag of Words
4. Text Classification

METHODS

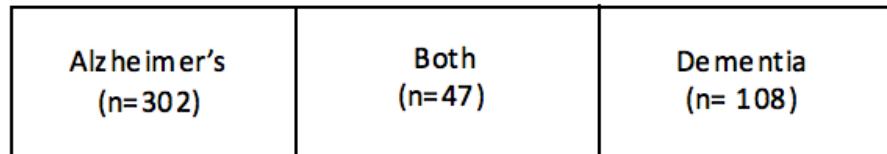
1. Process Data from MedHelp Health Forums into Dementia, Alzheimer's, and Both categories
2. Analyze Word Frequencies
 - ✓ Attempt 1: Questions Only
 - ✓ Attempt 2: Questions & Comments
3. Bag of Words
4. Text Classification

PROCESSING DATA & ATTEMPT I

 MedHelp
Forums
(n=1858)



Separated on
Key Words
(n=457)



1. Tokenization
2. Stemming
3. Stemmed Word Frequency

- Key Words:
 - UMLS concepts matching "alzheimer" (i.e. amyloid, tau) and "dementia"
- Questions Only
- Manually Removed Irrelevant Tokens post stemmed word frequency (Prepositions, Adverbs, Typos ect.)
- NLTK

FREQUENCY RESULTS: QUESTIONS



ATTEMPT I: WORD FREQUENCIES FROM QUESTIONS ONLY

Words	Word Frequencies					
	Alzheimer's Questions		Dementia Questions		Both Questions	
	Unadjusted	Adjusted (per 10,000)	Unadjusted	Adjusted (per 10,000)	Unadjusted	Adjusted (per 10,000)
alzheimer	216	58.2	3	1.6	70	90.7
dementia	3	0.8	142	75.2	67	86.8
home/house	71	19.1	50	26.5	10	13.0
brain	65	17.5	27	14.3	19	24.6
memor	4	1.1	67	35.5	24	31.1
test	47	12.7	26	13.8	9	11.7
sleep	48	12.9	24	12.7	4	5.2
neuro	21	5.7	28	14.8	16	20.7
mri	33	8.9	22	11.7	8	10.4
famil	47	12.7	14	7.4	0	0.0
prescri	17	4.6	8	4.2	10	13.0
gene	13	3.5	2	1.1	6	7.8
depress	25	6.7	24	12.7	12	15.6
cell	13	3.5	9	4.8	9	11.7
diseas	90	24.3	31	16.4	19	24.6

sympt	43	11.6	33	17.5	15	19.4
surger	18	4.9	7	3.7	3	3.9
doctor/dr	43	11.6	104	55.1	28	36.3
hospital	35	9.4	40	21.2	8	10.4
blood	44	11.9	21	11.1	6	7.8
nurs	41	11.1	57	30.2	3	3.9
job	8	2.2	22	11.7	2	2.6
confus	37	10.0	21	11.1	5	6.5
cost	1	0.3	9	4.8	1	1.3
alcohol	9	2.4	3	1.6	1	1.3
diabet	5	1.3	10	5.3	4	5.2
drug	26	7.0	28	14.8	10	13.0
hospic	9	2.4	25	13.2	0	0.0
chronic	12	3.2	10	5.3	3	3.9
amyloid	4	1.1	0	0.0	2	2.6
tau	0	0.0	0	0.0	0	0.0
protein	10	2.7	3	1.6	3	3.9
fac1	0	0.0	0	0.0	0	0.0
mitochondria	0	0.0	0	0.0	1	1.3
forget	42	11.3	23	12.2	22	28.5

**Adjusted for document length

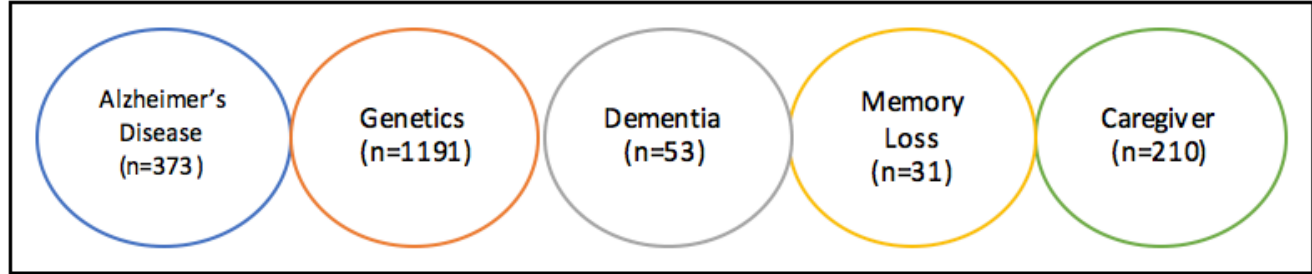
SUMMARY: FINDINGS OF ATTEMPT I

- Slight differences in word frequency
- Small sample size (n=457 questions), so difficult to determine statistical significance

ATTEMPT 2: INCLUDE COMMENTS

- Increased sample size (almost doubles)
- Approximately 1:1 ratio between question asked & number of response comments

MedHelp
Forums
(n=1858)



Separated on
Key Words
(n=457)

Alzheimer's (n=302)	Both (n=47)	Dementia (n=108)
------------------------	----------------	---------------------

1. Tokenization
2. Stemming
3. Stemmed Word Frequency

Combined Questions
with Comments to
Increase Sample Size
(n=815)

Alzheimer's (n=302)	Both (n=47)	Dementia (n=108)
Alzheimer's Comments (n=245)	Both Comments (n=35)	Dementia Comments (n=78)

1. Tokenization
2. Stemming
3. Stemmed Word Frequency

ATTEMPT 2: WORD FREQUENCIES FROM QUESTIONS & COMMENTS

Words	Word Frequencies					
	Alzheimer's Questions & Comments		Dementia Questions & Comments		Both Questions & Comments	
	Unadjusted	Adjusted (per 10,000)	Unadjusted	Adjusted (per 10,000)	Unadjusted	Adjusted (per 10,000)
alzheimer	482	43.0	54	8.6	110	3.6
dementia	102	9.1	296	46.9	95	19.9
home/house	224	20.0	193	30.6	26	13.0
brain	155	13.8	87	13.8	47	5.9
memor	186	16.6	108	17.1	55	7.3
test	160	14.3	58	9.2	39	3.9
sleep	105	9.4	61	9.7	16	4.1
neuro	123	11.0	87	13.8	32	5.9
mri	55	4.9	40	6.3	15	2.7
famil	152	13.5	61	9.7	34	4.1
prescri	39	3.5	21	3.3	21	1.4
gene	58	5.2	32	5.1	13	2.2
depress	65	5.8	40	6.3	18	2.7
cell	18	1.6	13	2.1	11	0.9
diseas	249	22.2	44	7.0	35	3.0

sympt	119	10.6	48	7.6	37	3.2
surger	25	2.2	27	4.3	4	1.8
doctor/dr	362	32.3	149	23.6	99	10.0
hospit	64	5.7	56	8.9	14	3.8
blood	90	8.0	44	7.0	18	3.0
nurs	126	11.2	75	11.9	10	5.0
job	30	2.7	27	4.3	20	1.8
confus	68	6.1	39	6.2	8	2.6
cost	14	1.2	21	3.3	5	1.4
alcohol	8	0.7	3	0.5	3	0.2
diabet	18	1.6	19	3.0	9	1.3
drug	87	7.8	40	6.3	22	2.7
hospic	38	3.4	27	4.3	1	1.8
chronic	26	2.3	12	1.9	5	0.8
amyloid	6	0.5	0	0.0	2	0.0
tau	0	0.0	0	0.0	0	0.0
protein	17	1.5	3	0.5	3	0.2
fac1	0	0.0	0	0.0	0	0.0
mitochondria	0	0.0	0	0.0	1	0.0
forget	118	10.5	36	5.7	42	2.4

Words	Difference Question Only (adjusted alz- adjusted dem)	Difference Question & Comments (adjusted alz- adjusted dem)	Average Difference
alzheimer	56.7	34.4	45.5
dementia	-74.4	-37.8	-56.1
home/house	-7.3	-10.6	-9.0
brain	3.2	0.0	1.6
memor	-34.4	-0.5	-17.5
test	-1.1	5.1	2.0
sleep	0.2	-0.3	0.0
neuro	-9.2	-2.8	-6.0
mri	-2.8	-1.4	-2.1
famil	5.3	3.9	4.6
prescri	0.3	0.1	0.2
gene	2.4	0.1	1.3
depress	-6.0	-0.5	-3.3
cell	-1.3	-0.5	-0.9
diseas	7.9	15.2	11.5
sympt	-5.9	3.0	-1.4
surger	1.1	-2.0	-0.5
doctor/dr	-43.5	8.7	-17.4
hospit	-11.7	-3.2	-7.5
blood	0.7	1.1	0.9
nurs	-19.1	-0.7	-9.9
job	-9.5	-1.6	-5.5
confus	-1.1	-0.1	-0.6
cost	-4.5	-2.1	-3.3
alcohol	0.8	0.2	0.5
diabet	-3.9	-1.4	-2.7
drug	-7.8	1.4	-3.2
hospic	-10.8	-0.9	-5.9
chronic	-2.1	0.4	-0.8
amyloid	1.1	0.5	0.8
tau	0.0	0.0	0.0
protein	1.1	1.0	1.1
fac1	0.0	0.0	0.0
mitochondria	0.0	0.0	0.0
forget	-0.9	4.8	2.0

ATTEMPT 1 & ATTEMPT 2 FINDINGS

- **Alzheimer's:** alzheimer, diseas
- **Dementia:** dementia, home/house, memor, neuro, diseas, doctor, hospit, nurs, job, hospic
- Common Medical Words not found
 - I.e Fac I, amyloid, tau, mitochondria
- No Genetic difference between Alzheimer's & dementia

METHODS

1. Process Data from MedHelp Health Forums into Dementia, Alzheimer's, and Both categories
2. Analyze Word Frequencies
 - ✓ Attempt 1: Questions Only
 - ✓ Attempt 2: Questions & Comments
3. Bag of Words
4. Text Classification

BAG OF WORDS

- Randomly split questions into 60 Train – 40 Test
 - Bag of words on train set

	# Alzheimer's Questions	# Dementia Questions	# Both Questions
Train (60%)	182	65	28
Test (40%)	120	43	19
Total	302	108	47

- Removed Stop Words (sklearn)
- Only took letters (Regex)
- Lower Cased text
- Attempt 1: with Stemming (NLTK)
- Attempt 2: without Stemming (NLTK)

W/O STEMMING

	Alzheimer's BOW	Dementia BOW	Both BOW
BOW Length	3544	1928	1627
Unique Words	2107	692	584
% (BOW Length/Unique Words)	58.45%	42.5%	25.9%

W/ STEMMING

	Alzheimer's BOW	Dementia BOW	Both BOW
BOW Length	2772	1534	1340
Unique Words	1532	465	392
% (BOW Length/Unique Words)	55%	30.3%	29.2%

METHODS

1. Process Data from MedHelp Health Forums into Dementia, Alzheimer's, and Both categories
2. Analyze Word Frequencies
 - ✓ Attempt 1: Questions Only
 - ✓ Attempt 2: Questions & Comments
3. Bag of Words
4. Text Classification

TEXT CLASSIFICATION (SCIKIT LEARN)

Categories	Precision	Recall	F1-score
Alz	0.82	0.93	0.87
Dem	0.63	0.55	0.59
Both	0.25	0.08	0.22
Avg/total	0.73	0.77	0.74

Trained Naïve Bayes & SVM Models

- Removed Stop Words, Bag of words (no stemming)

Best Model:

- SVM
- loss='hinge', penalty='l2', alpha=1e-3, n_iter=5, random_state=34

Summary of Findings

Common Words

- Alzheimer's:
 - Alzheimer
 - Diseas
- Dementia:
 - Dementia
 - Home/house
 - Memor
 - Neuro
 - Diseas
 - Doctor
 - Hospit
 - Nurs
 - Job
 - Hospic

Overall

- Very difficult to distinguish between Alzheimer's questions, dementia questions and questions asking both
- Little unique words in bag of words
- Low classification model scores

ALZHEIMER'S ↔ DEMENTIA

LIMITATIONS & LESSONS LEARNED

- Manually excluded and included stemmed words in word frequencies
- Online health forums are full of spelling errors
- Difference between everyday language and medical terminology

“I would prefer to be able to talk with my father in his final years (even if he doesn't recognize me) rather than look at someone who resembles Nicholson in the final scene of Cuckoo's nest.”

- Forum Question

“Are Alzhimers or Picks inheirited (Sorry, know that's spelled wrong)”

- Forum Question

• **Alzheimer's**

- Alz
- Alzherimers
- Alzhemers
- Alzhiemer
- Alzh
- Alzheimer
- Altzheimer's

• **Dementia**

- Demantia
- Demitia
- Dem
- Dementias
- Dementia
- Dimentia
- Dimensia

FUTURE DIRECTIONS

- Attempt with a larger dataset
- Analyze surrounding context of words to analyze vocabulary rather than specific words
- Incorporate a "spell checker" feature to account for common spelling words

QUESTIONS & ANSWERS



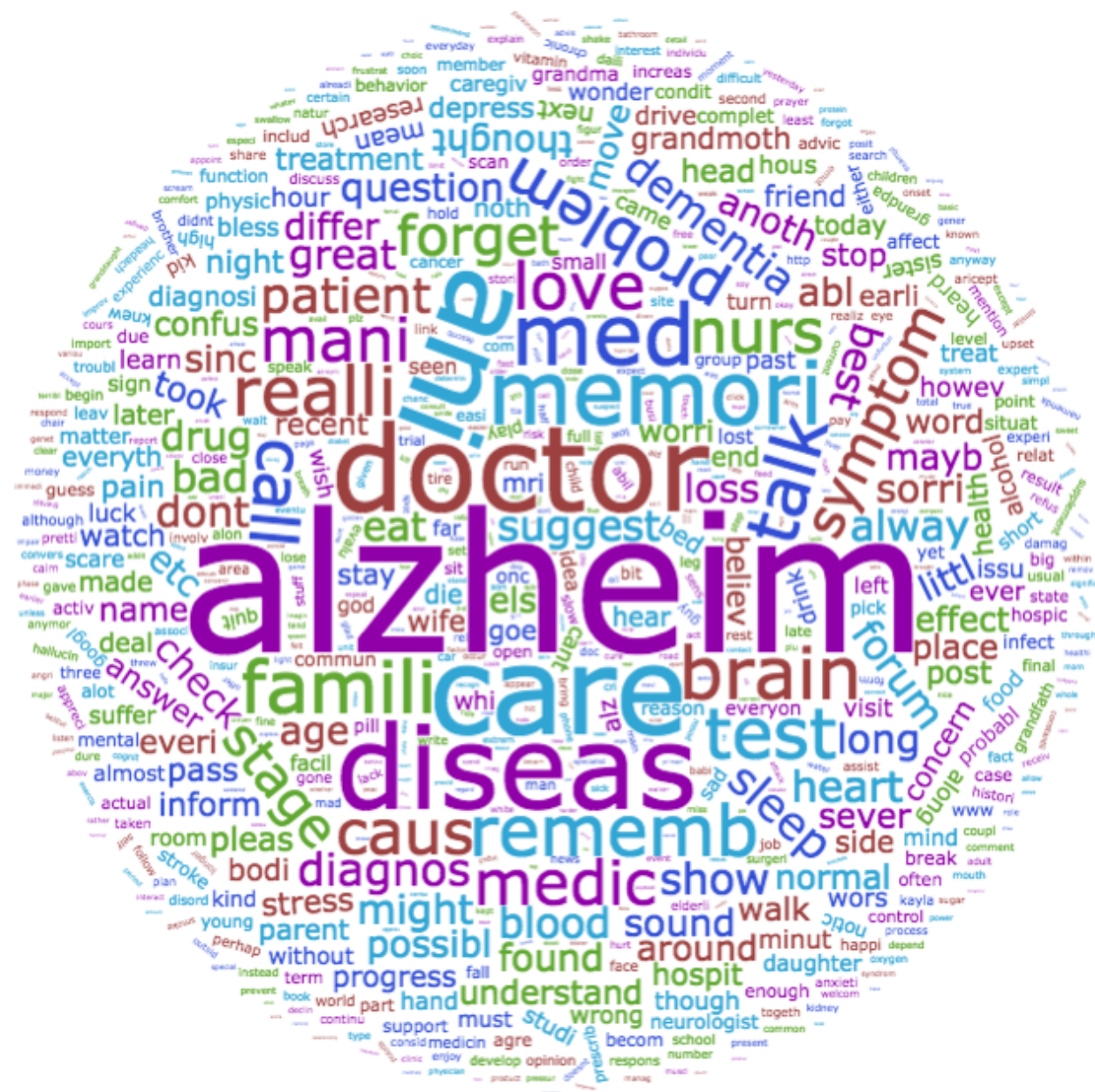
ALZHEIMER'S QUESTIONS



DEMENTIA QUESTIONS



BOTH
QUESTIONS



ALZHEIMER'S QUESTIONS & COMMENTS



DEMENTIA QUESTIONS & COMMENTS



BOTH QUESTIONS & COMMENTS