

## ACT Project 2

Natural language processing of clinical free text:  
methodological approaches to qualitative  
research with people with dementia

Will I. Bowers, KPWHRI

Lily N. Shapiro, KPWHRI



---

## Introductions

- Project 2—Alzheimer’s Disease subtypes
- Aim 3 Team!
  - Janelle S. Taylor, PhD
  - Marlaine Figueroa Gray, PhD
  - Kelly Ehrlich, MS
  - Tiina Maripuu, MA
  - Jenna Mae Peterson, MA
  - Norma B. Coe, PhD
  - Chuxuan Sun, MPA
  - Sundary Sankaran, MPH
  - Robert B. Penfold, PhD
  - Paul K. Crane, MD, MPH

# Presentation Outline

- I. Previous qualitative research in ACT
- II. Background of Natural Language Processing (NLP)
- III. Project 2, Aim 3 NLP Approach
- IV. Strengths and Challenges
- V. Results
- VI. Future work



# Qualitative Research in ACT

- Previous R21—Health Outcomes for Patients with Dementia without Family Caregivers (PI: Janelle Taylor)
- Attending qualitatively to clinical free text to learn about the lived experience of people with dementia
  - New use of ACT's incredibly rich data
  - More inclusive research

## Research Article

### **Kinless Older Adults With Dementia: Qualitative Analysis of Data From the Adult Changes in Thought Study**

Janelle S. Taylor, PhD,<sup>1,\*</sup> Marlaine S. Figueroa Gray, PhD,<sup>2</sup> Corinne M. Mar, PhD,<sup>3</sup> Paul K. Crane, MD, MPH,<sup>4</sup> Hitomi Kariya, MPH,<sup>5</sup> Callie Freitag, MA, MS,<sup>6</sup> Priyanka Taneja, BS,<sup>7</sup> Arvind Ramaprasan, MS,<sup>2</sup> Bettina Shell-Duncan, MPH, PhD,<sup>8</sup> Ann M. O'Hare, MA, MD,<sup>9</sup> Clara Berridge, PhD, MSW,<sup>10</sup> Elizabeth K. Vig, MD, MPH,<sup>4,11</sup> Stephanie G. B. Wheeler, MD, MPH,<sup>4</sup> Manu Thakral, PhD,<sup>12</sup> Rene J. Hawkes, BS,<sup>2</sup> and Eric B. Larson, MD, MPH<sup>2</sup>

*Journal of Aging Studies* 65 (2023) 101140



ELSEVIER

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

*Journal of Aging Studies*

journal homepage: [www.elsevier.com/locate/jaging](http://www.elsevier.com/locate/jaging)



Expanding the ethnographic toolkit: Using medical documents to include kinless older adults living with dementia in qualitative research

Lily N. Shapiro<sup>a,\*</sup>, Marlaine Figueroa Gray<sup>a</sup>, Callie Freitag<sup>b</sup>, Priyanka Taneja<sup>c</sup>, Hitomi Kariya<sup>b</sup>, Paul K. Crane<sup>b</sup>, Ann M. O'Hare<sup>d</sup>, Elizabeth K. Vig<sup>d</sup>, Janelle S. Taylor<sup>e</sup>



# Using NLP to expand on this method

## The Challenge

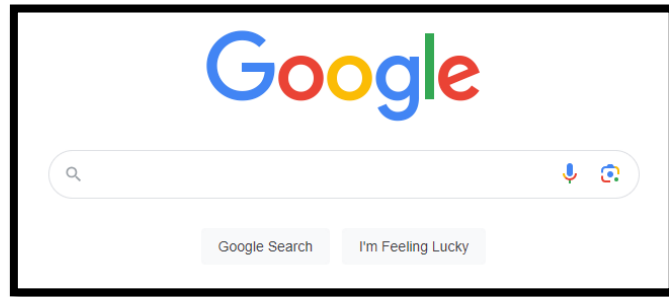
- Too much material
- Much of it is not relevant

## Why Natural Language Processing (NLP)?

- Ability to pre-process notes into categories based on keywords or people or note type
- Ability to focus on “enriched” notes (more likely to be relevant)
- Ability to create longitudinal corpora based on a sequence of notes for individuals

# What is Natural Language Processing (NLP)?

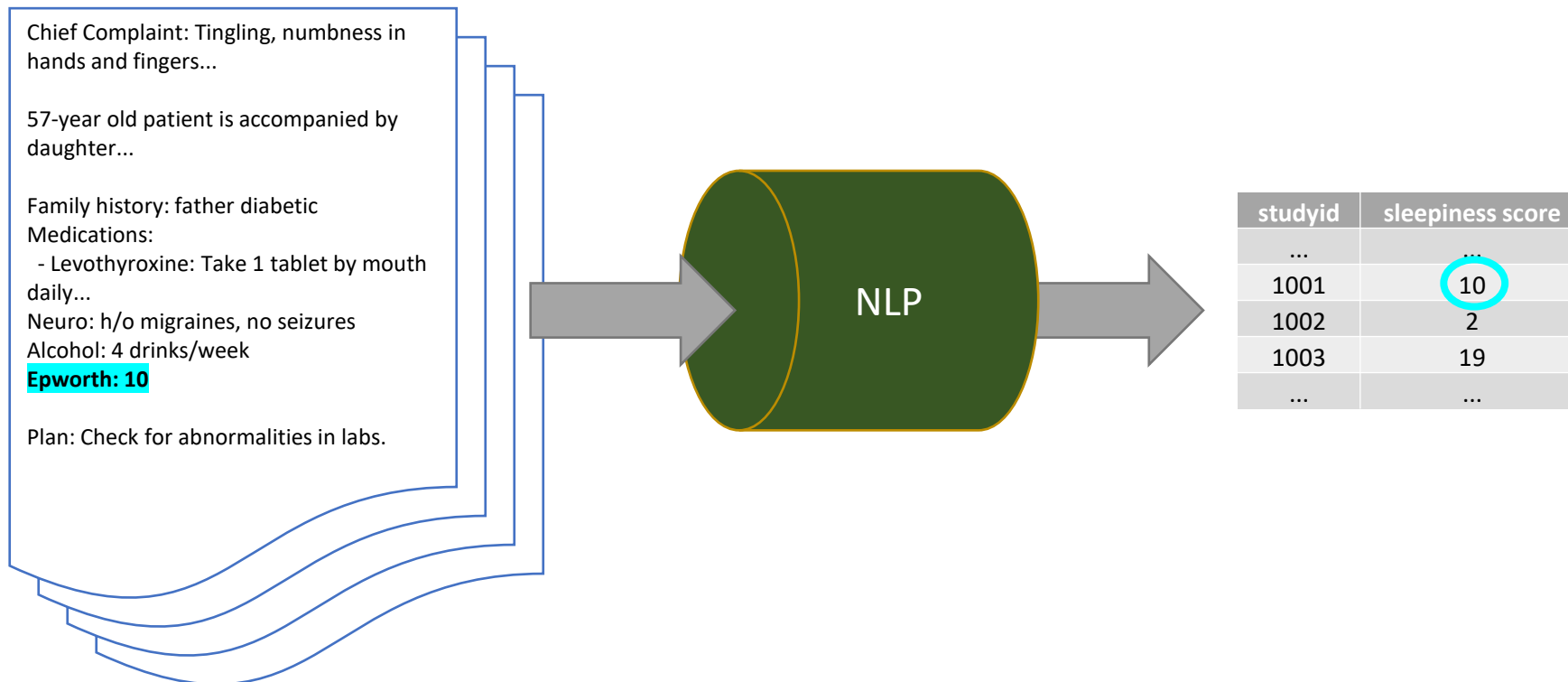
Enables computers to make use of human language.



# What is NLP in health research?

## CONCEPT EXTRACTION:

Extracting insights from clinical free text, especially what is incomplete or missing in structured data.



# Rule-based methods

- Manually develop patterns to find & extract target language
  - Use of regular expressions ('regex'; a language for advanced string-searching)

[...] She has **strong** social support. [...]

Regular Expression

```
(?:good|sufficient|strong|extensive)\W*(?:support\Wsystem|social\Wsupport)
```

[...] She has a **good** support system. [...]



# P2A3 NLP approach

- Task: Curate three enriched sub-corpora for qualitative analysis
  - Patient-initiated communication
  - Caregiving
  - Housing
- Distill 161,111 notes -> 500 relevant notes
- Two step approach:
  1. Extract concepts from entire corpus – pyTAKES
    - a. Derive concepts from previous work
  2. Leverage concepts to filter corpus into sub-corpora

# Step 1: Extract concepts

- Rule-based method – pyTAKES
  - Regular expressions -> allows fuzzy matching
  - Prioritizes recall (TPR) over precision (PPV)
  - Concept dictionary - 32 concepts with 6,786 concept texts

## **Patient point of view (POV)**

- “pt reports”
- “they state”

## **Caregiving & social support**

- “with patient”
- “family member”

## **Living situation**

- “AFH”
- “return to”

# Step 1a: Deriving concepts

- Bootstrap concept dictionary from analytical codes from previous work
  - Interface between qualitative and quantitative approaches
- Iteratively:
  - run pyTAKES, assess codes, refine codes, repeat

## Excerpt

“Pt notes that she feels isolated, lonely, and still working through grief and loss issues”



## Shortened Concept Text

“Grief”  
“Loss issues”  
“Isolated”  
“Lonely”  
“Pt feels”

## Qualitative Code

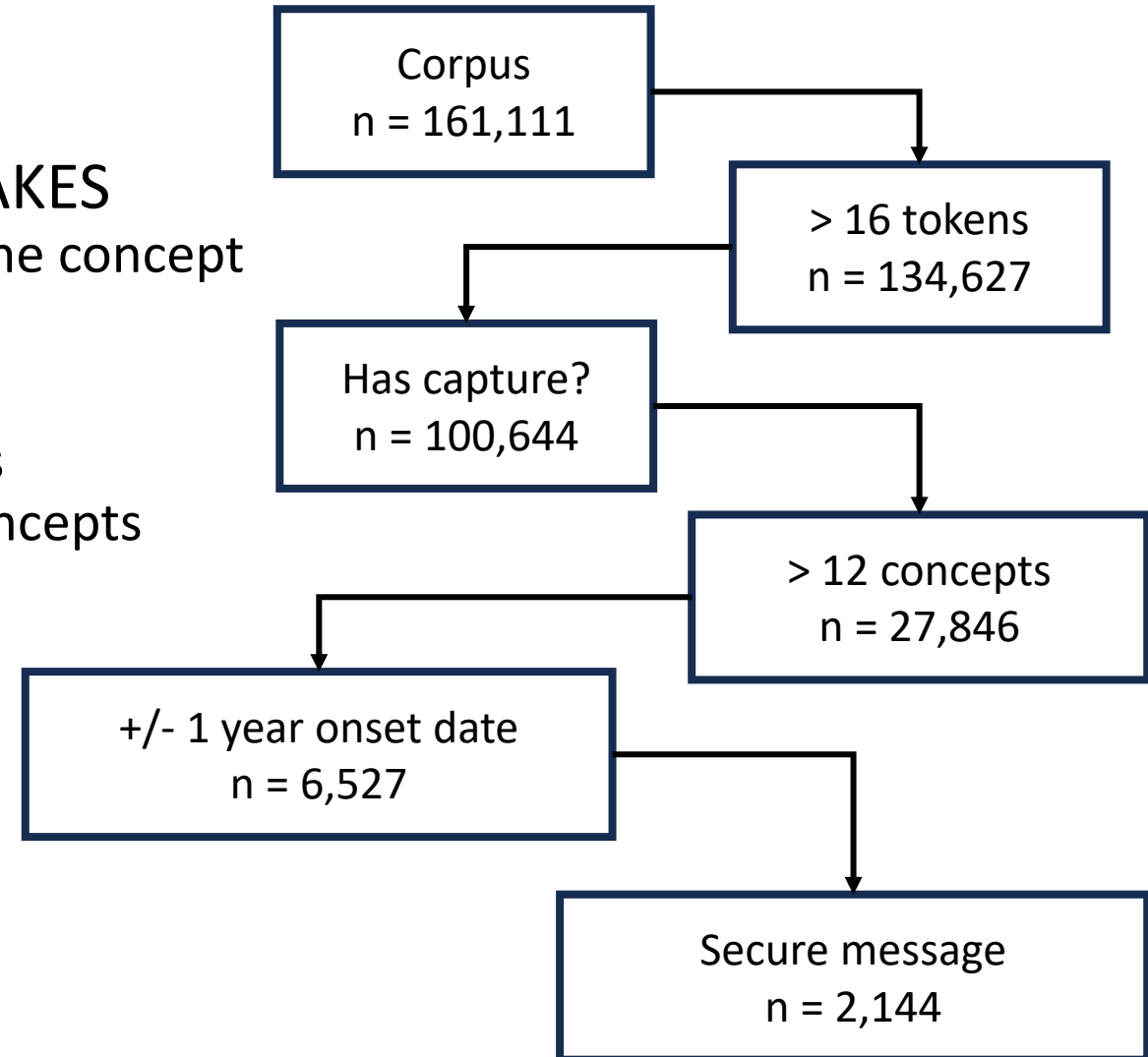
Patient emotions/affect

## Concept

Patient emotions/affect

# Step 2: Filter corpus into sub-corpora

- Distill 161,111 notes -> 500
- Leverage concepts extracted from pyTAKES
  - 100,644 (62.47%) of notes have at least one concept
- Pull 'levers' to filter sub-corpus
  - Minimum note length – 16 tokens
  - Minimum number concepts – 12 concepts
  - Minimum number repeat concepts – 2 concepts
  - Date range
  - Concept
  - AD subtype
  - Note type
- Finally, sample 500 notes



# Example: Curate caregiving sub-corpus

...Reviewed Allergies and Medications with pt...

...Cataract removal is one of the most common operations performed in the United States...

...moved from own condo to assisted living...doing well...

...The patient has family living nearby including a daughter...

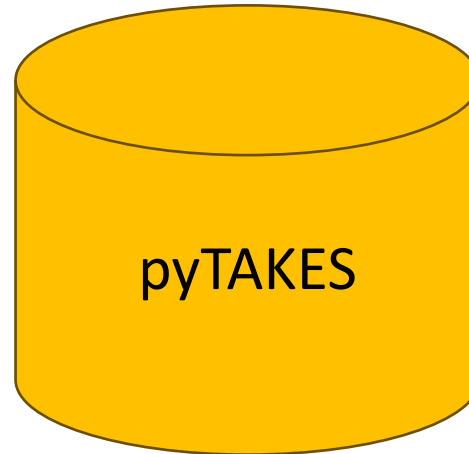
## Snippet

...Reviewed Allergies and Medications with pt...

...Cataract removal is one of the most common operations performed in the United States...

...moved from own condo to assisted living...doing well...

...The patient has family living nearby including a daughter...



## Concept

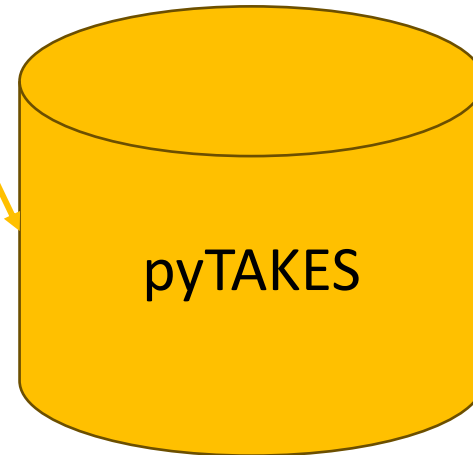
## Snippet

...Reviewed Allergies and Medications with pt...

...Cataract removal is one of the most common operations performed in the United States...

...moved from own condo to assisted living...doing well...

...The patient has family living nearby including a daughter...



## Concept

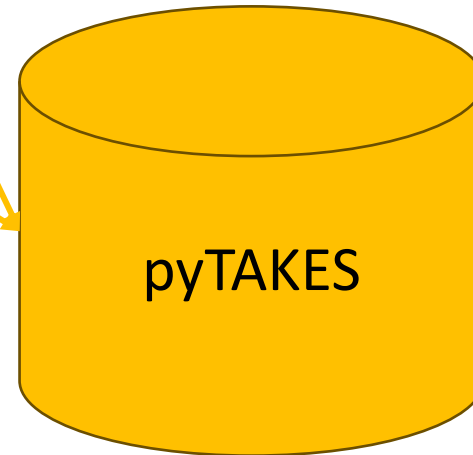
## Snippet

...Reviewed Allergies and Medications with pt...

...Cataract removal is one of the most common operations performed in **the United States**...

...moved from own condo to assisted living...doing well...

...The patient has family living nearby including a daughter...



pyTAKES

## Concept

Patient  
POV



## Snippet

...Reviewed Allergies and Medications with pt...

...Cataract removal is one of the most common operations performed in **the United States...**

...moved from own **condo** to assisted living...**doing well...**

...The patient has family living nearby including a daughter...

pyTAKES

## Concept

Patient  
POV

Living  
Situation

Patient  
Emotion

+

## Snippet

...Reviewed Allergies and Medications with pt...

...Cataract removal is one of the most common operations performed in **the United States...**

...moved from own **condo** to assisted living...**doing well...**

...The patient has **family living nearby** including a **daughter...**

pyTAKES

## Concept

Patient  
POV

Living  
Situation

Patient  
Emotion

Social  
Support

Subject

+

+

## Snippet

## Concept

## Keep?

...Reviewed Allergies and Medications with pt...

...Cataract removal is one of the most common operations performed in **the United States...**

Patient  
POV

...moved from own **condo** to assisted living...**doing well...**

Living  
Situation

Personal  
History

...The patient has **family living nearby** including a **daughter...**

Social  
Support

Subject

Snippet

Concept

Keep?

...Reviewed Allergies and Medications with pt...

...Cataract removal is one of the most common operations performed in **the United States...**

...moved from own **condo** to assisted living...**doing well...**

...The patient has **family living nearby** including a **daughter...**

Patient POV

Living Situation

Personal History

Social Support

Subject

Contains any concept?

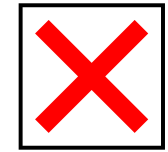


Snippet

Concept

Keep?

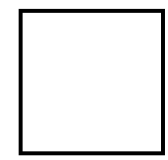
...Reviewed Allergies and Medications with pt...



...Cataract removal is one of the most common operations performed in **the United States...**

Patient POV

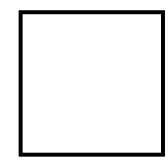
Contains any concept?



...moved from own **condo** to assisted living...**doing well...**

Living Situation

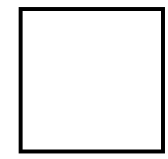
Personal History



...The patient has **family living nearby** including a **daughter...**

Social Support

Subject



## Snippet

## Concept

## Keep?

...Reviewed Allergies and Medications with pt...

...Cataract removal is one of the most common operations performed in **the United States...**

...moved from own **condo** to assisted living...**doing well...**

...The patient has **family living nearby** including a **daughter...**

Patient  
POV

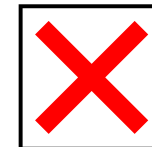
Living  
Situation

Personal  
History

Social  
Support

Subject

>= minimum  
(2) concepts?

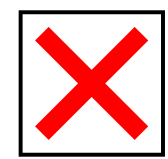


Snippet

Concept

Keep?

...Reviewed Allergies and Medications with pt...

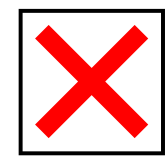


...Cataract removal is one of the most common operations performed in **the United States...**

Patient POV



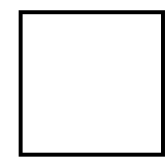
>= minimum (2) concepts?



...moved from own **condo** to assisted living...**doing well...**

Living Situation

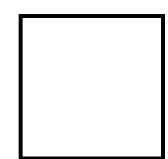
Personal History



...The patient has **family living nearby** including a **daughter...**

Social Support

Subject



## Snippet

## Concept

Keep?

...Reviewed Allergies and Medications with pt...

...Cataract removal is one of the most common operations performed in **the United States...**

...moved from own **condo** to assisted living...**doing well...**

...The patient has **family living nearby** including a **daughter...**

Patient  
POV

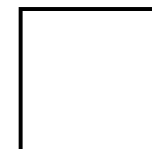
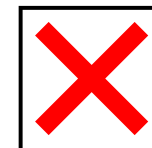
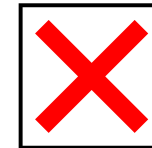
Living  
Situation

Personal  
History

Social  
Support

Subject

Contains 'Social Support' Concept?



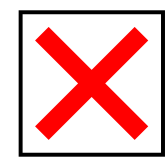


# Snippet

# Concept

# Keep?

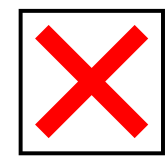
...Reviewed Allergies and Medications with pt...



...Cataract removal is one of the most common operations performed in **the United States...**

Patient POV

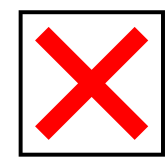
Contains 'Social Support' Concept?



...moved from own **condo** to assisted living...**doing well...**

Living Situation

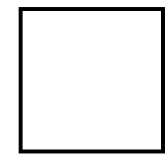
Personal History



...The patient has **family living nearby** including a **daughter...**

Social Support

Subject



## Snippet

## Concept

Keep?

...Reviewed Allergies and Medications with pt...

...Cataract removal is one of the most common operations performed in **the United States...**

...moved from own **condo** to assisted living...**doing well...**

...The patient has **family living nearby** including a **daughter...**

Patient  
POV

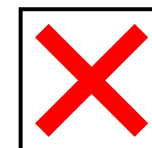
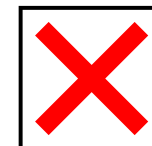
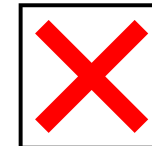
Living  
Situation

Personal  
History

Social  
Support

Subject

Contains 'Social Support' Concept?



## Snippet

## Concept

Keep?

...Reviewed Allergies and Medications with pt...

...Cataract removal is one of the most common operations performed in **the United States...**

...moved from own **condo** to assisted living...**doing well...**

...The patient has **family living nearby** including a **daughter...**

Patient  
POV

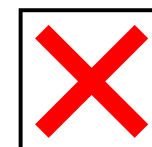
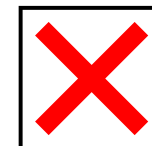
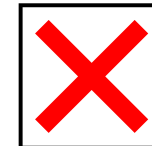
Living  
Situation

Personal  
History

Social  
Support

Subject

Contains 'Social Support' Concept?



# NLP: Strengths & Challenges

- Strengths:
  - Concept refinement is straightforward
    - Concepts & results are natural language, **NOT** abstract numbers or matrices
  - Results are highly interpretable
- Challenges:
  - Drawing nuanced distinctions between concepts
    - "ADLs: managing" vs. "ADLs: difficulty managing"
  - Significant expert involvement required

# Results

- Three enriched sub-corpora focused on relevant domains.
  - Patient-initiated communications
    - When do patients (and caregivers) reach out to the healthcare system, what kinds of concerns do they bring up, with what consequences
  - Caregiving networks, needs, challenges, successes
  - Living situation



# Results

- Richness of messages
  - Visibility of lived experience—including caregiving networks, living arrangements, daily life
  - Stories over time

Greetings,

I just returned from a visit to see [participant] and [participant's relative]. The same challenges that I had noted before seemed to appear again, in which [participant's] frustrations seem to be being taken out on those of us closer to her...and thus creating distance. We had a great talk last night, in which I told her we all want to support her, and affirm her control over her life, and feel closer to her. She now seems open to having a meeting with [social worker], [participant's relative] and myself. It would be nice if you [physician] could be there at the beginning, just to help frame it.

Thanks so much!

# Looking ahead

- Successful first pass of NLP-assisted, qualitative analysis of clinical free text
- Possible future work:
  - Longitudinal sub-corpora
  - Sub-corpora curated by AD subtype
  - Other aspects of a participant's lived experience
    - Emotions
    - Difficulty with tasks
    - Transportation
  - Making the framework publicly available



# Thank you!

Questions?

- [Lily.X.Shapiro@kp.org](mailto:Lily.X.Shapiro@kp.org)
- [Will.I.Bowers@kp.org](mailto:Will.I.Bowers@kp.org)