Applying Natural Language Processing (NLP) to Verbatim Patient-Reported Outcomes

[1]Jennifer L. Purks, BS [1]Michael Harris, MA [1]Karen E. Anderson, MD and [1]Ira Shoulson, MD

[1] Georgetown University, Washington DC, United States

MedStar Health Research Institute

Introduction

• REACH2HD examined the safety and benefits of PBT2, an experimental modulator of metal ionophores, on cognitive impairment, the major and untreatable source of disability in early HD.[1,2]

• Huntington Disease Patient Reported Outcome of Problem (HD-PROP) captures bothersome problem verbatim descriptions reported by individual patients.

• Natural Language Processing (NLP) extracts relationships and meaning from large text-based resources.[3] One tool is word clouds (images composed of words from a text where the size of the word indicates frequency or importance).

Objective: Apply Natural Language Processing (NLP) to analyze verbatim patient-reported outcomes (PROs) among Huntington Disease (HD) research participants in the Phase 2 REACH2HD randomized-controlled trial

Methods

• The HD-PROP were administered to the 109 REACH2HD participants at baseline (BL), Week 12 (W12), and Week 26 of experimental treatment (W26) (randomly assigned to PBT2 250 mg/day (n=36), PBT2 100 mg/day (n=38), or placebo (n=35).

• NLP, specifically word clouds, was applied to the HD-PROP problem and consequence data in the REACH2HD trial at BL, W12, W26, and all time points to extract relationships and meaning from large text-based resources.

Table 1: Most commonly reported single word from Word Cloud with Example Verbatim Reported Problem Responses

Word	Example Verbatim Reported Problem Responses
Balance	"Not riding bike- **balance**." "**Balance**." "hmm good question I'd say **balance** probably." "I fall quite often, I fell three times in the past month, my **balance**" "Worsening of **balance**" "Being unsteady and walking problems. 'That is hard' Being off **balance**"
Memory	"The fact that I lose my **memory** - I worry about that." "My **memory** is not what I would like it to be. I struggle with names of relatives, acquaintances." "I think the lack of **memory** and forgetting to do things like taking my medications." "somewhat failing **memory**: doesn't seem like a crisis" "**memory** and thinking" "**memory**"
Movements	"I guess it's the **movements**. I can't get comfortable and the I eventually get comfortable and I have to move again." "just **movements**" "My **movements** bother me the most." "(Chorea) the **movements**." "I miss driving. I have increased **movements** that cause me to give up my license." "The fact that my **movements** are still around even with medication."

Results

• The Problem Word Cloud showed that: "BALANCE" (69 counts), "MEMORY" (66 counts) and "MOVEMENTS" (48 counts) were the most bothersome reported problems (Figure 1).

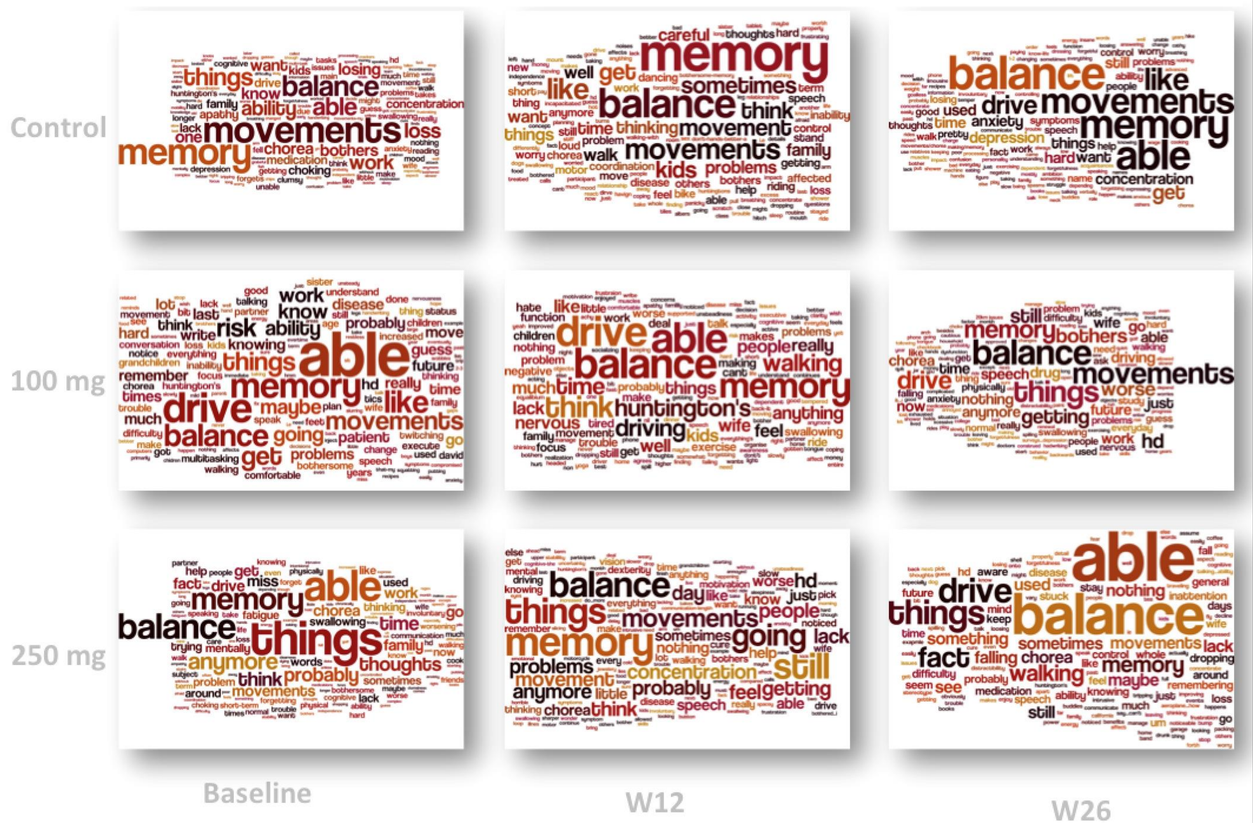
• There was a decrease from 8 "MEMORY" counts at BL to 4 "MEMORY" counts at W26 in the PBT2 250mg group (Figure 3).

• However, the Functional Consequence Word Cloud (Figure 2) did not show a clear or informative pattern.

Figure 1: Verbatim REACH2HD Problems reported Word Cloud collapsed across treatment group and time

Figure 2: Verbatim REACH2HD Consequences Word Cloud reported collapsed across treatment group and time

Figure 3: Verbatim Problems by treatment group over time



HD-PROP Questions

• A study staff member asked each participant the following questions and recorded verbatim responses. Participants were allowed to report up to eight bothersome problems.

1. What is the most bothersome problem due to your Huntington disease?

2. In what way does this problem bother you by affecting your every day functioning or ability to accomplish what needs to be done?

3. How much (severely) does this problem bother you by limiting your functioning?
   - 1 = Not at all
   - 2 = Mildly (minimally or rarely)
   - 3 = Moderately (more often than not)
   - 4 = Severely (plenty or all of the time)

Conclusion

• The NLP word clouds for verbatim-reported problems and consequences help to quantify and visually depict patterns in the HD-PROP dataset of the REACH2HD trial.

• Verbatim problems are more uniformly informed by NLP than verbatim consequences, perhaps related to the complexity of the questions and replies. This may reflect lack of insight or impaired cause and effect reasoning between problem and its functional consequence in this patient population.

• Some PBT2 250mg treatment effect on MEMORY is suggested by these results.

Future Directions

- More advanced NLP may provide more informative analysis of PROs.
  - Tri-grams (frequencies of three words reported together)
  - Text metrics (number of characters or words in a response)
  - Sentence parsing (model of grammatical sentence structure)

Acknowledgements & References

[1]Huntington Study Group Reach2HD Investigators. Safety, tolerability, and efficacy of PBT2 in Huntington's disease: a phase 2, randomised, double-blind, placebo-controlled trial. *The Lancet Neurology*, 2015 *14*(1), 39-47.
[2]Dorsey ER, Beck CA, Darwin K, et al. Natural history of Huntington disease. JAMA Neurol 2013; 70: 1520–30.
[3]Percha B and Altma RB. Learning the Structure of Biomedical Relationships from Unstructured Text. PLoS Comput Biol 2015;11(7): e1004216.

Funding Support: Prana Pharmaceuticals (Melbourne, Australia) and 2016 ANA Travel Grant.

Financial Disclosures: Ira Shoulson serves as non-executive Director of Prana Pharmaceuticals. Karen Anderson, MD is a paid consultant to Prana

PRANA BIOTECHNOLOGY Limited

GRIFFIN FOUNDATION

ANA