Using natural language processing in clinical research:
The promise and challenges of translating into practical tools*

Session Chair: Catherine Diaz-Asper, Ph.D.1
Discussant: Brita Elvevåg, Ph.D.2
1 Marymount University
2 University of Tromsø – The Arctic University of Norway

There is much excitement about potential applications of machine learning (ML) and natural language processing (NLP) to detect clinically significant changes in cognitive and mental state. However, the path from lab to clinic is thorny, and a clear map is needed so as to minimize the risks associated with using complex artificial intelligence (AI) models. The four speakers of this symposium - with diverse expertise in neuropsychology, clinical practice, cognitive science, artificial intelligence (AI), natural language processing, and technology transfer - will evaluate what it really will take for the algorithms and models to be translated into practical tools within psychology and psychiatry. The first speaker (Diaz-Asper) will draw on her experience of leveraging NLP methods to detect early cognitive decline in the elderly at risk for Alzheimer’s disease based on only brief samples of speech captured in real world settings. The second speakers (Cohen & Rodriguez) will focus on the need for a completely new approach to psychometrics that both enables the field to move beyond modeling just one data channel to integrate multimodal features and also explicitly build models that are culturally sensitive. The third speaker (Chandler) will challenge current practice of using aggregated data without first establishing ergodicity, namely that group level characteristics generalize to each individual, and the talk will showcase this approach using interactive data visualization techniques. The fourth speaker (Foltz) will consider how these methods and technology in general can best be translated to benefit the clinical world.

Keywords: natural language processing, speech, machine learning, artificial intelligence, cognitive decline, serious mental illness, clinical decision support system

Combining high tech and low tech approaches to increase accessibility to cognitive screening in the elderly.

Catherine Diaz-Asper (Marymount University).

Elderly populations face a myriad of barriers to accessing healthcare, which can result in lost or delayed opportunities for diagnosis and treatment. This presentation showcases how a combination of low cost and easily available technology can be combined with recent advances in natural language processing (NLP) and machine learning to identify cognitive decline associated with Alzheimer’s disease (AD) from spoken language alone. Ninety-one community-dwelling older adults diagnosed with mild AD, amnestic mild cognitive impairment (aMCI), or considered cognitively healthy completed a telephone interview in which parts were recorded for automated analysis using NLP and machine learning. Participants completed the interview in its entirety and rated the process as non-threatening and enjoyable. Recordings of the interviews were of relatively poor quality, yet NLP methods were nonetheless able to differentiate the three groups from one another with greater accuracy than traditional dementia screening methods and a human expert who read transcripts of their speech. Our approach capitalizes on the telephone as a widely available and familiar device for older people and utilizes cutting edge automated analysis techniques, resulting in a sensitive screening tool for cognitive decline based on only brief samples of speech captured in real world settings. Such an approach has the potential to ease the burden of in-person cognitive screening and potentially reach many more people than are currently evaluated.

Integrating multimodal features to improve model generalization, cultural sensitivity and interpretability in computational psychiatry and psychology.

Alex Cohen (Louisiana State University), Zach Rodriguez (Louisiana State University).

1 This article was published [to be completed by publisher].
2 (0000-0003-2323-7169)
3 This research was supported in part by the Research Council of Norway [grant number 231395] awarded to Brita Elvevåg, and the National Institutes on Aging [grant number R03AG052416] awarded to Catherine Diaz-Asper.
4 The authors declare no conflicts of interest.

Correspondence concerning this article should be addressed to Catherine Diaz-Asper Ph.D., Dept. of Psychology, Marymount University, 2807 N. Glebe Rd, Arlington VA, 22207, USA. Email: cdiazasp@marymount.edu
The notion that digital phenotyping efforts can be improved by integrating data from multiple channels (i.e., “multimodal” data integration) is well regarded. This is particularly true in clinical psychiatry and psychology, where assessment often requires consideration of verbal, nonverbal and other contextual data. As yet, the methods by which multimodal data integration should be accomplished, how they can be evaluated, and even an operational definition of it are not well developed. This presentation provides a framework for applying a broad range of multimodal data (e.g., speech, language, facial expression, geolocation, ecological momentary assessment) for understanding psychiatric states in patients with Serious Mental Illness (SMI). We begin by defining multimodal data integration within the context of computational psychiatry/psychology. Next, we argue that “multimodal” integration is critical for improving the generalization, cultural sensitivity and interpretability of models of psychiatric SMI states; three components that can serve as outcomes for model evaluation. Next, we consider methodological and computational approaches that can accommodate multimodal data. We conclude by considering next steps for multimodal data for research and clinical applications. We punctuate this presentation with data from four separate studies using data from smart phone sensors to measure various SMI states. Digital phenotyping has the potential for reshaping how assessment of SMI is conducted. Multimodal assessment is a critical, and complicated, component of this endeavor.

**Generalizability versus personalized modeling in automated mental state assessment: Investigations with information visualization techniques.**

Chelsea Chandler (University of Colorado, Boulder).

Assumptions in behavioral and medical research are typically made on aggregated data, under the premise that the data is *ergodic*, meaning that group level characteristics generalize down to each individual. Before results in domains that perform analyses on group-level behavioral data are determined to apply to their constituents, such populations must be tested for ergodicity. This talk offers an empirical first step towards investigating group ergodicity via the use of interactive data visualization techniques. Specifically, a multimodal dataset of human behavioral data (e.g., speech, language, touch, timing) was collected with a mobile phone application where participants with and without mental illness completed behavioral tasks over the course of multiple days and their interactions were recorded. A dashboard of three interconnected graphs was created to visualize the data longitudinally (i.e., how the data change over time) and multimodally (i.e., how the data change when investigating single modes versus entire datasets) both interindividually and intraindividually, and includes features to show data interactions with other important neuropsychological variables. In investigating group level consistency of the data, it was found that while the majority of the data does tend to be ergodic from the perspective of certain subsets of data types, there are outliers that behave less similar to their own group's aggregated data and tend towards the other's. Thus, the use of data visualization techniques can be a critical safeguard in behavioral research as it offers a simple and transparent way to explore any deviations from group trends in individuals for whom inferences made from group data may be catastrophically imprecise.

**Practical considerations for applying AI in real-world assessment of mental health.**

Peter Foltz (University of Colorado, Boulder).

There has been a long history of analyzing language to extract information about people’s mental states. Within the last 20 years, computational methods have become much more widely available, more powerful, and easier for researchers to use. Many current methods go beyond just counting basic features, and analyze deep conceptual semantic features in language and subtle differences in speech, syntax and language structure. At the same time, data of patient and clinician discourse has become more available, allowing not just categorizing patients based on broad classifications, but also in-depth analyses of computational language features that can serve as digital biomarkers. Assessing these biomarkers can improve our understanding of the neuropsychological bases of disorders as well as potentially faster, more personalized, and improved assessment and treatment delivery. While results from many recent research studies show great promise of applying NLP for assessments, there are several considerations that need to be addressed in order to translate promising lab studies to practical usage. This talk will describe these considerations and give specific examples from the transition of AI research to implementation. The talk will focus particularly on aspects of applying AI in ways that will permit trustworthiness by patients and clinicians, generalizability of results, avoiding bias and promoting equity in the use of the approaches. Finally, the talk will describe how these considerations can be best addressed in order to translate these methods to bring the greatest benefit to the field.