

Name of Organizer: Sook-Lei Liew & Steven C. Cramer

Affiliation: University of Southern California & University of California, Irvine

Phone: 713-816-8999 & 714-456-6876

Email: sliew@usc.edu & scramer@uci.edu

Title of Symposium: 'Big Data' for Rehabilitation: Promises, Pitfalls, and Future Potential

Description of submitted symposium (please limit to 2000 characters): A persistent challenge in rehabilitation research is the vast heterogeneity within clinical populations. This inter-individual variability makes it difficult to establish significance and reliably replicate findings of rehabilitation studies across smaller sample sizes. Large, diverse datasets (aka "big data"; e.g., $n > 1000$) have the potential to drive rehabilitation research forward by providing the greater statistical power needed for robustly evaluating clinical hypotheses and validating findings from smaller studies. However, collecting, organizing, and analyzing such large amounts of data comes with a number of limitations and considerations. Here we present current applications of 'big data' approaches for rehabilitation research across both retrospective and prospective collections of behavioral, neuroimaging, and clinical outcomes data. In each talk, we provide a balanced approach to this topic, highlighting both the potential of 'big data' approaches for driving the rehabilitation field forward, as well as the challenges associated with properly implementing, analyzing and interpreting the results. In doing so, we aim to educate attendees about current methodologies and available tools for conducting big data analyses in rehabilitation. We also hope to provide a tempered, realistic view of the limitations of these approaches and ways to complement this approach with experimental approaches. While several of the applications presented here focus on stroke rehabilitation, we emphasize general methodologies and applications that can be related to many rehabilitation populations. Attendees will not only gain big picture insights into how large datasets can be used to further rehabilitation research, but they will also learn practical knowledge regarding what types of information are contained in various databases, how to access or contribute to them, and how to use these resources for their own questions and purposes.

Length of time required for symposium?: 90 minutes total: 1) Brief introduction to the topic; 2) 15 minutes per talk + several minutes for questions; 3) 10 minutes at the end for general questions and discussion

Additional Presenters (Limited to 4 additional presenters, list full name and email address) Please Note: Any Non-member speakers must receive prior approval from the Program Chair.: Presenter 1 Full name: Kenneth J. Ottenbacher, PhD, OTR Email: kottenba@utmb.edu Talk Title: Large Data Tools and Databases for Analyzing Functional Outcomes in Medical Rehabilitation Presenter 2 Full name: Keith Lohse, PhD Email: krl0022@auburn.edu Talk Title: Centralized Open-Access Rehabilitation Database for Stroke (SCOAR): Information Architecture in Stroke Rehabilitation Trials Presenter 3 Full name: Sook-Lei Liew, PhD, OTR/L Email: sliew@usc.edu Talk Title: ENIGMA for Neurorehabilitation: A Large-Scale Meta-Analysis Approach to Modeling Neuroimaging, Genetics, and Behavior Presenter 4 Full name: Julie Bernhardt, PhD Email: j.bernhardt@unimelb.edu.au Talk Title: Lessons on Big Prospective Datasets from A Very Early Rehabilitation Trial (AVERT)

What is the role of each presenter?: 1) Dr. Ottenbacher is the Principal Investigator of the NIH-funded Center for Large Data Research and Data Sharing in Rehabilitation (CLDR). He will discuss what 'big data'

is (and what it is not), provide examples of successful large data projects, and introduce CLDR tools, databases and funding opportunities for rehabilitation researchers to conduct their own large data analyses. 2) Dr. Lohse is the director for the Rehabilitation Informatics Lab at Auburn University and will discuss data integration across trials using the Centralized Open-Access Rehabilitation database for Stroke (SCOAR). Archiving summary statistics from RCTs in a central repository allows for more advanced searching, generates exploratory meta-data for hypothesis generation, and highlights important methodological variables of dose, timing and patient age, that inform power-analyses and clinical trial design. 3) Dr. Liew is the head of the ENIGMA Stroke Recovery working group, an international consortium with a goal of generating large-scale ($n > 3000$) post-stroke neuroimaging and behavioral analyses. She will discuss the benefits of ENIGMA's meta-analytic approaches for pooling together large sets of neuroimaging, genetic and behavioral data (e.g., $n > 30,000$, Hibar et al., Nature) and highlight preliminary results and challenges from several rehabilitation-related ENIGMA working groups (stroke recovery, traumatic brain injury, and multiple sclerosis). 4) Dr. Bernhardt is the Leader of the AVERT Early Intervention Research Program, the largest international trial of early mobility-based rehabilitation ($n > 2000$). She will briefly discuss the AVERT trial design and results, and highlight both the unique benefits and challenges of collecting large samples of prospective data across diverse, international study sites. She will also provide lessons learned for future large-scale clinical trials.

Objective 1: On the completion of this activity, participants will be able to define what 'big data' is for rehabilitation research and list four examples of 'big data' rehabilitation studies, including the types of data used and the methods used to analyze that data.

Objective 2: On the completion of this activity, participants will be able to describe three strengths of using different types of large datasets in rehabilitation research as well as three limitations/considerations.

Objective 3: On the completion of this activity, participants will be able to identify available resources to engage in 'big data' rehabilitation research, whether through using existing resources (e.g., from the Center for Large Data Research), joining existing meta-analytic efforts (e.g., SCOAR, ENIGMA), or starting their own large dataset collections.