



# ABSTRACT BOOK

Podium and Poster Presentations Featuring  
PROMIS<sup>®</sup>, Neuro-QoL<sup>™</sup>, ASCQ-Me<sup>®</sup>, and NIH  
Toolbox<sup>®</sup>

2019 HealthMeasures User Conference

June 4-6, 2019

Prentice Women's Hospital Conference Center

250 E. Superior Street, Chicago, IL

For additional information, visit [www.HealthMeasures.net](http://www.HealthMeasures.net)

Steven J. Schuetz, MD

Northwestern University, Feinberg School of Medicine  
s-schuetz@northwestern.edu

Rina S. Fox, PhD, MPH; Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine; Chicago, IL, USA;

Claudia B. Soliz P., BS; PROMIS®-Bolivia Organization; La Paz, Bolivia;

Maribel Marmol C., MD; Instituto Chuquisaqueño de Oncología; Sucre, Bolivia;

Edgar F. Apaza O., BA; PROMIS®-Bolivia Organization; La Paz, Bolivia;

Mayra V. Alaca M., BA; PROMIS®-Bolivia Organization; La Paz, Bolivia;

PROMIS®-Bolivia Collaborative; La Paz, Bolivia

### **An evidence-based approach to implementing the Patient Reported Outcomes Measurement Information System in a Bolivian public cancer center**

**Introduction & Purpose:** The Patient-Reported Outcomes Measurement Information System (PROMIS®) is a comprehensive set of tools to measure self-reported physical, mental, and social health in people ages 5–90. While PROMIS® has been widely used in high-income countries, it has not been implemented in Latin America despite the availability of high-quality Spanish-language translations. The purpose of the present study was to implement PROMIS® symptom screening in a multicultural, low-resource public cancer center in Bolivia, the Instituto Chuquisaqueño de Oncología (ICO).

**Methods:** An evidence-based implementation strategy was developed by incorporating the Expert Recommendations for Implementing Change into the EPIS phases of implementation (Exploration, Preparation, Implementation, Sustainment). An implementation protocol was developed based on the International Society for Quality of Life Research standards, with local contextualization based on an analysis of implementation barriers and facilitators using the Consolidated Framework for Implementation Research (CFIR) domains. All patients receiving cancer care at ICO between June 1, 2018 and March 1, 2019 were considered for inclusion in the present pilot program. Participants were eligible to complete PROMIS® symptom screening if they were over 18 years old and self-reported basic Spanish proficiency; participants were excluded if they had a cognitive or physical impairment that precluded participation. PROMIS® computer-adapted tests of anger, anxiety, depression, fatigue, and pain interference were completed up to seven times by eligible participants before each clinic visit, with a frequency of no more than once every 3 weeks.

**Results:** A total 958 patients attending 1973 clinic visits were screened for eligibility. Of these, PROMIS® was completed at 1386 (70.2%) of 1736 eligible clinic visits. All five PROMIS® domains were evaluated at 1226 visits (88.5%), and a subset of those domains were evaluated at the remaining 11.5% of visits. Leading reasons for battery non-completion included lack of time and patient physical disability. Severe symptoms (T-score > 70) of anger (9.5%), anxiety (10.8%), depression (7.1%), fatigue (2.9%), and pain interference (4.8%) were identified in this population.

**Conclusions:** Multiple frameworks exist to facilitate evidence-based and sustainable implementation of patient-reported outcome assessment using PROMIS® in diverse settings. Preliminary results of the present study demonstrate the applicability of these frameworks to a low-resource setting, and support the feasibility of implementing PROMIS® symptom screening in a low-resource ambulatory cancer

center. Future objectives include evaluating the acceptability and appropriateness of this protocol, as well as the differential functioning of PROMIS® items in this population.

Janel Hanmer, MD, PhD

University of Pittsburgh Medical Center

hanmerjz@upmc.edu

Jennifer Freel, Susanne Liposky. Both at University of Pittsburgh Medical Center

**UPMC Patient Reported Outcomes Center's governance structure, integration with Epic, and clinical operations training.**

The UPMC health system has successfully collected patient reported outcomes (PROs) through Epic since 2012. PRO questionnaires are generally assigned by visit type, though some questionnaires are assigned by patient characteristics (diagnoses, age, etc). Patients can complete questionnaires via the MyUPMC patient portal, by an office-based tablet computer, or with the help of rooming staff. Implementation of PRO collection includes training of UPMC clinicians and staff on the use of this information and how it supports shared medical decision making.

After a proliferation of many different measures (up to 73 active at one time) across many different clinics (currently 235), the UPMC health system has developed and implemented several different policies to improve scalability and success of new PRO projects. One such policy is that all new projects must include a core set of PROMIS metrics which is now administered by Computer Adaptive Testing. Other questionnaires are allowed if they are not redundant with the core PROMIS metrics, are validated, and provide clinically actionable information.

We will present the policies, materials, and our experiences associated with launching successful clinic-based PRO projects and our understanding of the barriers that can derail implementation. We will discuss our project application process, discovery, copyright, team structure, operational training, and reports.

Allison W. McIntyre  
University of Rochester Medical Center  
allisonw\_mcintyre@urmc.rochester.edu  
Judith Baumhauer, MD, MPH, Kathleen Fear, PhD

### **Implementing Workflow Change for Successful Collection of Patient-Reported Outcomes in Clinical Practice: Deconstructing the 'Golden Triangle'**

The 'Golden Triangle' business model defines the three components of successful project launch and organizational transformation as People, Process and Technology. Each tip of the triangle is equally as important, however, when attempting to transform a health system and implement standard of care, clinical collections of Patient-Reported Outcomes (PRO), the triangle must transform to allow a clear understanding of what the People need and how the Process works before developing new health technologies to collect PRO data. Once the health technology is in place it is also important to solicit feedback and have the ability to adjust to ever-changing needs in a healthcare system.

Collecting information from patients, staff and providers about what they would like to see in developing health technologies is a first step. This data allows an organization to focus energy and resources on systems that will produce a higher return on investment, both personal and financial. It is important to note that the needs and wants of different groups will often overlap and may even contradict each other. For example, providers who want to collect 15 PRO instruments at every clinic visit will, most certainly, cause survey fatigue in their patients. Using information provided by focus groups, surveys, and personal observation is crucial as health organizations develop new technologies.

Patients tend to be concerned about how long a PRO takes to complete and the content of the questions. Staff typically wants to make sure that the collection of PRO won't interfere with an already busy workday and providers need easily accessible data that can be shared in a meaningful way with their patients. If the health technology developed to collect, visualize and store PRO data addresses these issues, the result is greater satisfaction from a perceived increase in value. Educational components are also necessary to ensure that all parties are kept informed and remain invested in the clinical collections of PRO.

Understanding the Processes involved in a health system is another very important piece of the puzzle. Healthcare is a fast-paced industry and any disruptions can lead to negative associations with health technology that is intended to support more efficient collections of PRO. Working with patients, staff and providers to develop and customize the delivery of new health technology in different ambulatory and in-patient settings is necessary to succeed. With their input, a PRO collection platform might use information from the electronic medical record to assign PRO and/or use bar code technology to streamline access to the PRO, thus minimizing extra effort required to launch the PRO platform. Integrating PRO into the electronic medical record in a location that is convenient for providers to access and utilize is another way health technology can be optimized to improve PRO collections.

Developing new health care technology to collect patient-reported outcomes that is based on the needs of the people who will ultimately use the technology and the processes already in place is the path to successful change in workflow and implementation of a clinical collection of PRO.

Benjamin D. Schalet

Northwestern University, Feinberg School of Medicine

b-schalet@northwestern.edu

John Devin Peipert, Northwestern University, Feinberg School of Medicine; Terry Nakazono, UCLA Geffen School of Medicine; William Cunningham, UCLA Geffen School of Medicine

### **Examination of the PROMIS-29 Health Scales in Patients with Human Immunodeficiency Virus (HIV)**

**Background:** Advancements in anti-retroviral therapy has significantly increased the length of life for HIV+ persons. This achievement has turned the focus away from mortality and toward outcomes like health-related quality of life (HRQOL) and symptoms. Living longer, HIV+ persons are increasingly likely to develop comorbid conditions that compromise HRQOL, especially physical aspects of HRQOL. In addition, there is still significant stigma associated with HIV, which has significant negative impacts on mental HRQOL. To date, few studies have examined the psychometric properties of the Patient Reported Outcomes Measurement Information System (PROMIS) measures in HIV+ persons. To help fill this gap, we examined the PROMIS-29 among a sample of HIV+ persons.

**Method:** Data were sourced from the baseline assessment of a randomized controlled trial aiming to increase care retention among HIV+ patients: Effectiveness of Peer Navigation and Contingency Management on Retention in HIV Care (CHAMPS). Medical comorbidities and the PROMIS-29 profile measure were assessed with 450 patients prior to the peer navigation intervention. First, we examined distributions of the PROMIS-29 scales (physical function, anxiety, depression, fatigue, sleep disturbance, ability to participate in social roles, pain interference) and compared to the US general population normative mean value of 50 (standard deviation = 10). We examined the dimensionality of the PROMIS-29 with exploratory factor analysis, pre-specifying a two-factor solution. Each PROMIS-29 scale T-score was entered into the factor model. Scales with loadings of >0.40 were retained. We then estimated PROMIS-29 scores for patients with comorbid conditions. Differences > 1/3 of a standard deviation (3 T-score points) were interpreted as meaningful.

**Results:** Mean T-scores for several PROMIS-29 scales appreciably trended below the US population mean: physical function (mean = 45.3, range = 22.9-56.9), anxiety (mean=54.4, range=40.4-81.5), satisfaction with social roles (mean = 46.9, range=28.9-63.7), pain interference (mean=54.7, range=41.6-75.6). Factor analyses found two factors, one representing physical HRQOL [physical function (loading=0.90), pain interference (loading=0.80), satisfaction with social roles (loading=0.60)] and mental HRQOL [anxiety (loading=0.90), depression (loading=0.90), fatigue (0.60), sleep disturbance (loading=0.40)]. The correlation between the two factors was -0.40. Contributing evidence of construct validity, patients reporting a clinical diagnosis of depression had higher PROMIS depression T-scores in comparison to patients who did not (55.8 vs. 47.9), and patients reporting a clinical diagnosis of anxiety had higher PROMIS T-scores than those who did not (57.6 vs. 49.8). Other comorbidities were associated with physical HRQOL PROMIS scales. Patients with hypertension had higher pain intensity (57.3 vs.53.1) and lower physical function (42.8 vs. 46.8) T-scores than those without. In addition, patients with diabetes had higher pain interference (57.6 vs. 54.2) and lower physical function (41.6 vs. 45.9) T-scores in comparison to patients without diabetes.

**Conclusions:** The PROMIS-29 scales evidenced validity among persons with HIV. Notably, in factor analyses, physical and mental HRQOL was negatively correlated among this sample, which differs from

results in the US national sample, where the correlation was 0.69, indicating a different relationship between physical and mental health among persons with HIV.

Cameron Moore  
Northwestern University Feinberg School of Medicine  
cameron.moore@northwestern.edu  
Rina S. Fox, Rylee Brower, Sophia Mazurek, & Jason C. Ong

### **Characterizing Health-Related Quality of Life in Narcolepsy: A Mixed-Methods Study**

**Introduction:** Narcolepsy is a sleep disorder characterized by chronic, excessive sleepiness which currently has no cure. Poor psychosocial functioning and diminished quality of life have been reported in the literature, but the applicability of measures included in the Patient-Reported Outcome Measurement Information System (PROMIS) has not yet been explored for people with narcolepsy (PWN). The purpose of this mixed-methods study was to use quantitative data collected from PROMIS as well as qualitative data from focus groups to characterize the health-related quality of life (HRQL) in PWN.

**Methods:** Twenty-nine adults (93% female, mean age = 31 years) with established medical care for narcolepsy (Type I = 58.6%) completed PROMIS computer adaptive tests for depression, anxiety, fatigue, sleep disturbance, sleep-related impairment, pain interference, and physical function, as well as the Patient Health Questionnaire-9 (PHQ-9). Focus groups were conducted using live videoconferencing, which were led by a psychologist who queried participants on various issues related to the impact of narcolepsy symptoms on psychosocial and occupational functioning, the effectiveness of current treatments to improve HRQL, and suggestions for the development of a psychosocial intervention. Thematic analysis was used to code qualitative data.

**Results:** Clinically significant levels ( $t$ -score  $> 60$  or  $t$ -score  $< 40$ , scale-dependent) were reported on the PROMIS scales for depression ( $t = 64.8$ , SEM = 2.7), anxiety ( $t = 66.3$ , SEM = 2.6), fatigue ( $t = 68.3$ , SEM = 2.5), and sleep-related impairment ( $t = 66.9$ , SEM = 2.5) but not on sleep disturbance ( $t = 57.7$ , SEM = 2.83), pain interference ( $t = 53.8$ , SEM = 2.69), or physical function ( $t = 43.6$ , SEM = 2.16). Elevated scores for depressive symptoms were reported on the PHQ-9 ( $M = 15.79$ ,  $SD = 3.85$ ). Qualitative data from focus groups revealed strong consensus that public awareness of narcolepsy was generally lacking, which resulted in many PWN feeling misunderstood. There was also agreement among PWN that symptoms of narcolepsy impacted psychosocial functioning, with many unable to find the energy to maintain relationships and many others reporting feeling depressed due to self-doubt, loss of self-identity, or diminished capacity for professional achievement. Some participants also felt dismissed by health care providers who were not knowledgeable about narcolepsy.

**Conclusion:** This study is the first to report PROMIS data for PWN. This quantitative data, along with the qualitative findings from the focus groups, highlights the debilitating effect of chronic excessive sleepiness and the burden of managing narcolepsy symptoms on psychosocial functioning. In particular, the negative stigma of narcolepsy, low self-image and self-efficacy, and the feeling of being dismissed by health care providers were key factors which contributed to poor psychosocial functioning and reduced HRQL. The findings demonstrate the applicability of PROMIS instruments to PWN, and illuminate specific targets for psychosocial interventions to improve HRQL in this population.

**Support:** This project was funded by a grant from Wake Up Narcolepsy.

Christopher Abeare  
University of Windsor  
cabeare@uwindsor.ca

Laszlo Erdodi and Isabelle Messa (University of Windsor); Doug Terry ( Department of Physical Medicine and Rehabilitation, Harvard Medical School, Boston, Massachusetts, USA - Spaulding Rehabilitation Hospital, Boston, Massachusetts, USA - MassGeneral Hospital for Children Sports Concussion Program, Boston, Massachusetts, USA - Home Base, A Red Sox Foundation and Massachusetts General Hospital Program, Boston, Massachusetts); Will Panenka (British Columbia Neuropsychiatry Program; Vancouver, British Columbia, Canada - Department of Psychiatry, University of British Columbia, Vancouver, British Columbia, Canada); Grant Iverson ( Department of Physical Medicine and Rehabilitation, Harvard Medical School, Boston, Massachusetts, USA - Spaulding Rehabilitation Hospital, Boston, Massachusetts, USA - MassGeneral Hospital for Children, Sports Concussion Program, Boston, Massachusetts, USA - Home Base, A Red Sox Foundation and Massachusetts General Hospital Program, Boston, Massachusetts); and Noah Silverberg (Division of Physical Medicine & Rehabilitation, University of British Columbia, Vancouver, British Columbia, Canada - Rehabilitation Research Program, Vancouver Coastal Health Research Institute, Vancouver, British Columbia, Canada)

### **Development of Embedded Performance Validity Indicators in the NIH Toolbox Cognitive Battery**

Performance validity assessment is central to the assessment of cognitive abilities. However, NIH Toolbox Cognitive Battery (NIHTB-CB) currently lacks embedded validity indicators (EVIs). Further, there has been little to no research evaluating the influence of performance validity on NIHTB-CB scores. In this prospective multicenter study of participants with mild traumatic brain injury (mTBI), we sought to develop EVIs for the NIHTB-CB. Data were collected from 98 adults with mTBI consecutively referred for neuropsychological assessment. Neurocognitive functioning was assessed with NIHTB-CB. The Medical Symptom Validity Test served as the criterion for performance validity. Classification accuracy of three categories of EVIs was examined (7 Subtest-Based, 6 Derivative EVIs, and 8 Item-Based). The signal detection profile of individual EVIs varied greatly, but multivariate models had superior classification accuracy. Failing  $\geq 3$  Subtest-Based EVIs at the liberal cutoff or  $\geq 2$  at the conservative cutoff produced a good combination of sensitivity (.55-.59) and specificity (.89-.93) to psychometrically defined invalid performance. Combining the Subtest-Based and Derivative EVIs improved sensitivity (.59-.73) at comparable specificity (.88). However, combining all 21 EVIs into a single index provided no incremental advantage (.59-.73 sensitivity at .88-.89 specificity). In conclusion, the newly developed EVIs within the NIHTB-CB effectively discriminated valid from invalid response sets. Aggregating EVIs within the same category into validity composites improved signal detection over univariate cutoffs. Combining Subtest-Based and Derivative EVIs resulted in the optimal balance of classification accuracy and user-friendliness, providing a flexible and psychometrically sound method of performance validity assessment utilizing existing scores on NIHTB-CB subtests.

Colleen E Crangle

University of Louisville / Converspeech LLC

crangle@alumni.stanford.edu

Sallie Dannenberg, A Balanced Approach to Counseling; John Hamel, PhD, LCSW

### **Using PROMIS in batterer intervention programs: Towards program evaluation**

We describe a novel application of PROMIS measures in a batterer intervention program. Domestic violence is a serious and persistent health and social problem in the United States with close to 7 million women and 5 million men experiencing intimate partner violence annually. Convicted offenders are typically court-ordered to attend a 52-week batterer intervention program. However, these programs appear to be ineffective when measured by re-offence rates. Program attrition rates are high, and thought to be a prime reason for the limited evidence of effectiveness. Program evaluation, therefore, now seeks rather to identify specific characteristics of offenders that affect their ability to complete a program and to measure changes in these characteristics throughout the program. It is widely believed that emotion processing influences violent behaviors in abusive relationships, with preliminary studies showing structural brain differences in emotion processing and emotion regulation areas between male batterers and other violent criminals, and presenting evidence that when provoked to anger, violent offenders exhibit increased brain activity involved in emotion along with decreased brain activity involved in emotion regulation.

In light of these studies, as part of on-going service evaluation at A Balanced Approach to Counseling, an agency in Santa Clara County, California, that offers a court-certified 52-week program, emotion-related PROMIS HealthMeasures have been used to collect longitudinal data from offenders during their participation in the program. Paper-based PROMIS forms on Anger (5-item) and Depression (8-item) were filled out at the start of every weekly class. The following results were obtained with this data collection, revealing specific challenges and suggesting ways to address them.

Over the period February to August 2017, n=111 people completed n=103 Depression measures (average 10 weeks) and n=33 Anger measures (average 4.7 weeks). Two problems were evident. Many fewer Anger forms were completed. Of these 61% appeared to have been non-cooperatively filled out, with 1 selected for every item every week, and similarly 35% of the Depression forms. Clients subsequently reported considerable concern that any indication of ongoing anger might be reported to the court, hence the greater non-compliance on that measure. The non-cooperative responses in general may be attributable to clients believing “they don’t care about my feelings” in court-ordered programs that focus, by law, on holding offenders accountable for their violence. Furthermore, paper forms did not have close to real-time data input, resulting in lack of compliance not being noted in time.

To better integrate the PROMIS measures with the batterer intervention program and address these problems, we began a pilot test to administer PROMIS measures on tablets using Google Forms to present one item at a time, with each client having his/her own tablet, the forms being completed together at the start of the class, and only non-personally-identifying information collected. The 8-item Self Efficacy for Managing Emotions measure was the first used. We report on the initial three months of this amended data collection and our efforts to test if it is feasible and informative to use emotion-related PROMIS measures in this population and setting.

Hua Zheng, PhD

Univ of Massachusetts Medical School

hua.zheng@umassmed.edu

Danielle Lavalley, PharmD PhD (Univ Washington), Christina Bond, Patricia Franklin, MD (Northwestern Feinberg)

## **Integrating Patient-reported Measures and Predictive Data in a Shared Decision Report for Knee and Hip Arthritis Treatment Decisions**

### Introduction

Patient-reported outcome (PRO) data are often collected for research or quality reporting. However, the real-time use of PROs to facilitate patient-clinician shared decision making is new and poses new informatics challenges. For example, PRO and decision tools are often built in electronic health record (EHR) systems that are limited to local, potentially incomplete, data without national norms or statistical computation functions. To test real-time use of PRO-based, shared decision reports, the A.S.K. (Arthritis care through Shared Knowledge) study integrates real-time PROs, FORCE-TJR national registry data, and predictive analytics to generate individual shared decision reports. We describe the methods and outcomes of designing/generating individual shared decision-reports to empower patient-clinician treatment decision-making.

### Methods

An iterative process defined a direct-to-patient, web-based, data capture system to assess patient-reported symptoms and risk factors prior to the first clinician visit. The FORCE-TJR registry generated national PRO norms for pre-surgical patients and multi-variable prediction models for post-surgical PRO (pain, function) outcomes. Patients and clinicians evaluated multiple iterations of report content and design for ease of data interpretation. The informatics team translated the preferred visual and functional specifications to a web-based data reporting system to be used in the clinic.

### Results

Overall, patients preferred report backgrounds of red (advanced OA), yellow, green (healthy) and simple bar graphs to place PRO scores in context with norms. Clinicians and patients agreed on clinical and PRO measures to support shared decisions about surgical and non-surgical treatments. The final office reports include trended pain and function scores over time, comparisons with national norms, patient risk profiles, and individualized predictions of likely pain relief and functional gain after surgery. Of note, patients preferred prediction models of likelihood of being pain free with specific tasks, while clinicians preferred predictions of PRO scores.

### Conclusion

A direct-to-patient, web-based, data capture system captures patient-reported symptoms and risk factors prior to the first office visit. The shared decision report integrates national data with predictive models and PROs submitted by arthritis patients at time of clinical decisions. The ASK report is now being evaluated in a cluster randomized trial of 3000 patients. Early interviews confirm that patients and

clinicians value the individual predictive reports. The system can be readily deployed to interface with EHRs to return the data and decision reports for storage.

#### Implications

New informatics tools to generate evidence-based, individualized PRO reports have the potential to improve patient-clinician shared treatment decision making.

Prakash Jayakumar

University of Texas at Austin; Dell Medical School; The Commonwealth Fund, NYC

prakash.jayakumar@austin.utexas.edu

Prakash Jayakumar, Harkness Fellow in Health Care Policy and Practice Innovations; The Value Institute / Department of Surgery and Peri-operative Care The University of Texas at Austin. Teun Teunis

Department of Plastic Surgery. University Medical Center. Ana-Maria Vranceanu Massachusetts General Hospital and Harvard Medical School Behavioral Medicine Program Department of Psychiatry Sarah

Lamb Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Medicine University of Oxford. Mark Williams Department of Sport and Health Sciences, Oxford Brookes University. David Ring

Department of Surgery and Peri-operative Care The University of Texas at Austin. Stephen Gwilym

FRCS(Orth) Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Medicine

University of Oxford. Nuffield Orthopaedic Centre.

## **Construct Validity and Precision of different Patient Reported Outcome Measures during Recovery after Upper Extremity Fractures**

### Background

Patient perceptions of their limitations due to illness and injury can be quantified using patient reported outcome measures (PROMs). Assessments of construct validity (i.e. using correlations and factor analysis) and precision (i.e. floor and ceiling effects) of commonly used physical limitation PROMs in upper extremity trauma populations varying by area of focus (i.e. general health, region-specific or joint-specific measures) and mode of administration (i.e. fixed scale or computer adaptive test) are lacking.

### Questions / Purposes

Question 1. What is the magnitude of correlation between different PROMs within a week, 2-4 weeks and 6-9 months after shoulder, elbow, and wrist fracture? Question 2. What are the underlying constructs being measured by these PROMs using factor analysis? Question 3. What are the floor and ceiling effects of these instruments?

### Methods

A total of 744 patients recovering from an isolated shoulder, elbow or wrist fracture completed physical limitation PROMs that differed by area of focus and mode of administration at baseline (i.e. their initial office visit after diagnosis in the emergency department), 2-4 weeks after injury, and at final assessment 6-9 months after injury. Measures were evaluated over the course of recovery for construct validity (using correlations and factor analysis), and precision (using floor and ceiling effects).

### Results

Physical limitation PROMs were intercorrelated at all time points, and the magnitude of correlation increased over time. Factor analysis identified 'capability' (the ability to perform or engage in different activities) and 'biopsychosocial health' as separate and distinct underlying constructs being measured by these PROMs. Joint-specific and general health PROMs demonstrated high ceiling effects 6-9 months after injury.

### Conclusions

There is substantial correlation between PROMs that assess physical limitations based on anatomical region, and general health after upper extremity fractures and this relationship strengthens during recovery. Regardless of the delivery modality or area of focus, PROMs largely appear to represent two underlying constructs, 'capability' and 'biopsychosocial health'. Computer adaptive tests may be favoured over fixed scale measures for their efficiency and limited censoring.

Prakash Jayakumar

University of Texas at Austin; Dell Medical School; The Commonwealth Fund, NYC  
prakash.jayakumar@austin.utexas.edu

Teun Teunis, University of Utrecht; Ana-Maria Vranceanu, Harvard Medical School; Sarah Lamb,  
University of Oxford; David Ring, UT Austin; Steve Gwilym, University of Oxford

## **The Relationship between the Magnitude of Limitations and Patient Experience during recovery from an Upper Extremity Fracture**

### Background

The relationship between magnitude of limitations (measured by Patient Reported Outcomes Measures or PROMs) and satisfaction with care and services (measured by Patient Reported Experience Measures or PREMs) over the course of recovery after injury is unclear. The purpose of this study was to assess the relationship among a range of PROMs and PREMs in patients with shoulder, elbow and wrist fractures at 3 time points: the initial office visit, 2-4 weeks later and 6-9 months post-injury.

### Methods

We enrolled 744 adult patients with an isolated shoulder, elbow, or wrist fracture and invited them to complete PROMs and PREMs at their initial visit within the outpatient surgical practice (maximum 1 week post fracture), between 2 and 4 weeks, and between 6 and 9 months after injury. Correlational analysis was performed utilising a range of the most widely used PROMs and PREMs at each time point. PROMs varied by type (i.e. joint-specific health, region-specific health, general health) and mode of administration (i.e. fixed scale, computer adaptive test). PREMs assessed satisfaction with care providers and satisfaction with hospital services.

### Results

There was moderate to high correlation between PROMs and PREMs 6-9 months after injury. The strength of the correlation increased exponentially from the first week to the 6-9 month evaluation. These correlational trends were observed with all forms of PROM. Patients reporting greater limitations after injury were also less satisfied with their care and services.

### Conclusions

The increasing alignment of PROMs and PREMs as recovery after an upper extremity fracture progresses suggests that restored physical function improves perceptions of satisfaction over time. Fracture care should address both patient satisfaction and their limitations during recovery in order to maximise positive patient outcomes.

Robert Schloesser  
Sheppard Pratt Health System  
rschloesser@sheppardpratt.org

## **Implementation and Utilization of Patient-Reported Outcomes at the Behavioral Health Point-of-Care**

Sheppard Pratt Health System (SPHS) is the largest non-profit provider of mental health and substance use services in the country. As a nationwide resource, SPHS provides 2.3 million services each year across a comprehensive continuum of care, spanning both hospital- and community-based services. Sheppard Pratt Health System (SPHS) treats over 80,000 patients annually including over 12,000 inpatient admissions at two hospitals.

During the past two years SPHS has systematically implemented patient-reported outcome measures (PROM) including PROMIS and NeuroQoL measures as part of standard of care.

In this presentation, we will present data from these efforts focusing on 12 diverse clinical programs including adult and child and adolescent inpatient units, outpatient clinics, day programs and specialized program for patients with eating disorders and neuropsychiatric conditions. As of February 2019 over 5000 patient have participated in PROM and >3000 patients are actively enrolled in longitudinal assessments as part of their long-term outpatient care, resulting in over 50,000 completed PROMIS assessments. Over 200 clinicians are utilizing PROM data and are actively involved selection of the utilized measures, measurement frequencies and design and development of reporting tools.

In this presentation, we will discuss:

- Development of novel technological solutions that allow PROM data-capture and reporting at the point of care. Our technology is based on open-source frameworks and tools including the REDCap web application, R programming language and D3/ggplotly data visualization framework. Our tools are fully integrated with our electronic medical record and patient scheduling systems to minimize the burden caused by systematic PROM implementation for all point-of-care stakeholders.
- Our processes for implementation of PROM utilizing agile project management principles to facilitate multidisciplinary team work and continuous improvement to improve stakeholder engagement.
- Our approach to common language and graphical reporting of PROM data to improve interpretability and utilization for clinical decision making.
- Results of ongoing collection including metrics on collection efficiency, response rates, completion time as well as psychometric data. As of February 2019 we have consistently achieved over 80% of survey completion rates in programs with defined length of stay and high participation/low refusal rates at outpatient programs. Psychometric data shows expected higher degree of reporting of symptoms associated with psychiatric disorders (e.g. depression, anxiety, anger) in our patient populations and we will present data over the course of treatment and stratification by different patient populations.

- We will discuss future plans for utilization of PROM data and predictive analytics for treatment stratification and clinical decision support.

Ann Kerlin  
Luther Rice College & Seminary  
ann.kerlin@lutherrice.edu

### **Using PROMIS and Neuro-QoL in Residential Drug Treatment Facilities for Both Client Assessment and Program Evaluation**

This presentation covers the design, implementation, and results of two years of data collection using four PROMIS measures, and one Neuro-QoL measure on clients seeking residential drug rehabilitation. Assessments were given at entry, and information forwarded to administrators and counselors on staff. Assessments were repeated every 3 months. These measures are short and easy to use in this setting, and we collected data using Survey Monkey. Clients were logged into the survey using a desktop computer, and results were downloaded and scored using the free Assessment Center tool at Health Measures.com. These tools made for affordable program evaluation and client progress tracking in a non-profit environment.

Ann Marie Kerlin  
Luther Rice College & Seminary  
ann.kerlin@lutherrice.edu

### **PROMIS and Neuro-QoL in Addictions Treatment and Program Evaluation**

Addictions are an epidemic in our nation. My particular interest are Christian treatment protocols, where one's faith and desire for change are linked. In most treatment centers, whether religiously affiliated or not, there is a dearth of research on outcomes and that was my initial interest. I had been collecting data on a residential treatment center using some instruments that are well-known but expensive to implement (the SCL-90R for one), and decided to use the PROMIS measures and Neuro-QoL when I was introduced to them working on a medical study.

This project included the following assessments from HealthMeasures:

- PROMIS Short Form v.1.0 – Depression 8a
- PROMIS Short Form v.1.0 – Anxiety 8a
- PROMIS Short Form v. 2.0 – Emotional Support 8a
- PROMIS Short Form v.1.0 – Alcohol Use 7a
- NEURO-QoL Short Form v. 2.0 – Cognitive Function

Additionally, the Drug Abuse Screening Test-10, the Experiences in Close Relationships –Relationship Structures (ECR-RS) was used to measure attachment styles with people and with God, and the Transgression-Related Interpersonal Motivations Scale--12-Item Form (TRIM-12) was used to examine forgiveness.

Demographic data, including a history of adverse experiences was also gathered for this project.

Data collection was completed initially by using the free version of Assessment Center and then moved to SurveyMonkey (non-existent budget).

Because the measures here are free, program evaluation type of assessments on a variety of clinics could be managed fairly inexpensively, although the data analysis and someone to run the project would require a financial investment. Results of the findings on residential treatment will be presented; data was gathered at entry, at 3 months, and again at 6 months. The results of the project are in the process of being submitted for publication.

Barry Dewitt

Carnegie Mellon University

barrydewitt@cmu.edu

Louis Matza (Evidera), Katie Stewart (Evidera), Janel Hanmer (University of Pittsburgh)

### **The Neuro-QoL<sup>®</sup> Utility (NQU) Scoring System**

**Purpose:** To produce a societal preference-based scoring system using the following six domains of the Neuro-QoL<sup>®</sup>: depression, fatigue, ability to participate in social roles and activities, cognitive function, upper extremity function, and lower extremity function. The scoring system is called the Neuro-QoL Utility (NQU) scoring system.

**Methods:** A general population sample from the UK (n=203) was recruited for utility elicitations using the standard gamble technique, which were administered in-person. Elicited preferences were used to estimate single-attribute scoring functions for each of the six Neuro-QoL domains, which were then combined using multi-attribute utility theory to produce a single summary multi-attribute scoring function. A sample of UK multiple sclerosis (MS) patients (n=61) completed a similar in-person survey, used for convergent and known-groups validation.

**Results:** The six single-attribute scoring functions were fit using isotonic regression with linear interpolation. The sample's preferences demonstrated interactions among the health domains, favoring the multiplicative multi-attribute functional form over the linear additive form. The multiplicative multi-attribute NQU scoring function estimates utilities for Neuro-QoL multi-attribute health profiles on a scale where 0 is the utility of being dead and 1 is the utility of full health, which are also the lowest and highest possible utility scores, respectively. The MS patients were assigned systematically lower NQU scores than the general population (mean difference = 0.12 (0.08);  $p < 0.001$ ). Among the MS patients, the NQU had a Spearman correlation with the EQ-5D-5L and Health Utilities Index of 0.66 ( $p < 0.001$ ) and 0.63 ( $p < 0.001$ ), respectively. Those reporting more disability from their MS had systematically lower NQU scores (e.g., splitting the group by number of relapses, the mean NQU score of those below the median number of relapses scored 0.04 lower than those above).

**Conclusion:** Health-related quality of life preference-based scores are used to track and compare groups in populations, clinical health studies, and cost-effectiveness analyses. We developed a societal preference-based scoring system, the NQU scoring system, that is based on six Neuro-QoL domains. The NQU shows good convergent and known-groups validity. Studies collecting patient-reported outcomes focusing on neurological health can now incorporate a preference-based measure into their analyses.

Christina Bond, MS

Patient Research Group

christinabond@patientresearchgroup.com

Danielle Lavalley, PharmD PhD (Univ of Washington), Hua Zheng, PhD (Univ of Massachusetts), Patricia Franklin, MD MBA MPH (Northwestern Feinberg)

## **Implementing PRO-based Patient-Clinician Shared Decision Reports in the Learning Healthcare System**

### Introduction:

Learning health systems require transformative informatics and operational protocols to place patients and their data at the center of care decisions. If implemented successfully, informed patient-clinician shared decisions have the potential to improve healthcare quality and value. Osteoarthritis (OA) patients' discussion of surgical treatment options with clinicians can benefit from these data. However, optimal use of real-time shared decision reports is contingent on implementation strategies that integrate into clinical workflow. We will discuss the development and deployment of tailored workflow strategies and their impact on PRO-based decision report use.

### Methods:

OA patients and clinicians informed the content and visual display of information required for a shared decision report in the A.S.K.; Arthritis care through Shared Knowledge study. Patient-reported outcomes and clinical risk data are integrated with national and local health data to generate normative and predictive analytics. New office protocols were developed and staff trained to assure complete data were collected in the home or office and the report was available prior to the office visit. Implementation solutions were tailored to 13 unique office environments (26 clinicians) and 1456 patients completed both pre-visit PRO surveys and a post-visit phone interview by an independent researcher.

### Results:

The web-based data capture and reporting system automatically scored PROs and risk factors, generated predicted outcomes, and presented results in real-time, individual reports for use by 100% of patients in this sample. Implementation challenges for patients and clinics included the need for: (1) new pre-visit procedures for routine distribution of web-based PRO assessments and timely data collection, (2) office-based technology for patients to report the PRO and risk data when not completed via email, and (3) office staff training to generate the real-time report and make the report available for the patient and clinician at the visit. To date, 57% of PRO surveys were completed prior to arrival at the office (43% in office), 58% of patients reported receiving a printed report in the office, and 45% remembered discussing the report with the clinician.

### Conclusion:

Novel, real-time PRO data can be integrated with national data and predictive analytics to generate shared decision reports to improve patient care. However, new patient engagement and office implementation and training strategies are required to collect uniform data, generate reports, and use these data during patient-clinician visits.

### Implications:

Evidence-based, individualized shared decision reports generated at the point of care can facilitate patient-clinician shared treatment decisions, which can accelerate the learning healthcare system. However, today's information technology and office logistics must be adapted to support efficient and effective implementation.

Courtney K. Blackwell  
Northwestern University Feinberg School of Medicine  
ckblackwell@northwestern.edu

### **Don't throw the baby out with the bathwater: Creating developmentally-appropriate PROMIS early childhood parent-proxy instruments**

Currently, the Patient-Reported Outcome Measurement Information System (PROMIS®) provides pediatric instruments beginning at age 5 via parent-proxy reports and beginning at age 8 via child self-reports. While such measures cover a wide range of childhood and adolescence, they miss one of the most critical time periods for physical, neurological, and social development. Indeed, this arbitrary age cliff leaves a noticeable absence of brief, scientifically rigorous and clinically-relevant assessments of young children and fails to reflect the state of the science in developmental measurement, which increasingly demonstrates the reliability and validity of such constructs as young as 1 year of age. As part of the NIH-funded Environmental influences on Child Health Outcomes (ECHO) research program, the ECHO Person-Reported Outcomes (PRO) Core engaged in a novel adaptation of the PROMIS methodology for the development and validation of PROMIS Early Childhood (EC) parent-proxy instruments across the physical, mental, and social health domains.

Common practice for extending existing adult health measures to pediatric versions oftentimes results in “tweaking” constructs to fit at younger ages for lifespan coherence or simply throwing out constructs all together in the early years if they cannot be mapped directly to older youth and adults. For example, tantrums are the most discrete expression of anger/frustration in early childhood but would be inappropriate to assess in adolescence. Similarly, evaluating a 3-year-old’s well-being is unlikely to directly map onto the existing PROMIS well-being concepts of Life Satisfaction and Meaning and Purpose, but this does not mean well-being in and of itself is irrelevant for young children. Such methods of adapting measures often result in substantial error and reduced validity as behaviors are expressed differently at different developmental periods; this is particularly critical in early childhood when capacities emerge and are consolidated at an accelerated pace.

Alternatively, the current project brought together state-of-the-art PROMIS methods with developmental specification theory to develop a series of parent-proxy instruments for use with parents of 1- to 5-year-olds that are both meaningful developmental expressions and reasonably consistent with extant PROMIS constructs that span the life course. This presentation will begin with an overview of the key considerations used to develop these new instruments, describe how the new measures conceptually align with existing PROMIS parent-proxy measures for 5- to 17-year-olds, and showcase the psychometric validation results and final item banks/scales of the 11 new PROMIS EC measures: global health, anger/irritability, anxiety, depressive symptoms, positive affect, family relationships, peer relationships, sleep health, physical activity, engagement, and self-regulation.

Deborah Mattila and Andrew Westin  
CIBMTR / National Marrow Donor Program  
dmattila@nmdp.org  
University of Pittsburgh and Northwestern University

**Building a novel and flexible electronic patient-reported outcome collection system for cellular therapy research**

The Center for International Blood and Marrow Transplant Research (CIBMTR) is a research collaboration between the National Marrow Donor Program/Be The Match and the Medical College of Wisconsin. CIBMTR maintains an observational database of patients receiving cellular therapies and collects patient, disease, and transplant characteristics routinely reported by hematopoietic cell transplant (HCT) centers. CIBMTR also conducts and supports clinical trials related to HCT and other cellular therapies. Historically, survey or patient-reported outcome (PRO) data to support clinical trials or pilot projects were collected by phone or paper, and merging PRO data with clinical data in the observational database was difficult. In 2018 CIBMTR established a new electronic PRO (ePRO) data collection system to address consolidated data analysis needs and to move towards routine PRO data collection in the registry. PROMIS measures form the foundation of this ePRO system.

The presenters in this session will share how CIBMTR's ePRO system was built using the Assessment Center API with Qualtrics Surveys as the patient interface and Salesforce as its customer relationship manager (CRM) to manage and track study participants. Presenters will show, through diagrams and screenshots, how the ePRO system components are integrated for flexible collection of HealthMeasures and other PRO data in a seamless participant experience. The presenters will also share how PROMIS and other PRO data are combined with clinical data already collected for the outcomes registry and clinical trials, and applications of PROMIS and ASCQ-Me domains in HCT and other cellular therapy research.

Elizabeth Fallon

American Cancer Society

Elizabeth.Fallon@cancer.org

Robert L. Stephens, American Cancer Society; Tenbroeck G. Smith, American Cancer Society; David Cella, Northwestern University; Susan Gapstur, American Cancer Society

### **Measurement Invariance of the PROMIS Global Health Measure in a Large Prospective Cancer Cohort**

**Background :** Prospective cohort studies designed to study cancer etiology can be leveraged to study the impact of demographic, behavioral, genomic, clinical and psychosocial factors on quality of life among cancer survivors. Foundational to this research, however, is establishing the measurement properties of quality of life measures within the cohort's sample. Therefore, this study aimed to replicate the Patient Reported Outcome Measurement Information System (PROMIS) Global Mental Health (GMH) and Global Physical Health (GPH) factor structures, and test for measurement invariance across sex, age group, and over time in the Cancer Prevention Study-II Nutrition Cohort (CPS-II NC). **Method:** CPS-II NC participants with complete data for all GMH and GPH items on the 2009 and 2011 questionnaires comprised the analytic sample (N = 77,521). The PROMIS GPH and GMH factor structures were evaluated using confirmatory factor analysis at both 2009 and 2011. Subsequent analysis examined measurement invariance across sexes, four age groups (< 72.3 yrs, 72.3 to < 77.3 yrs, 77.3 to < 82.3 yrs, ≥ 82.3 yrs), and over time (2009 - 2011) separately for GMH and GPH using configural, weak, and strong factor models. All analyses used the weighted least squares estimator, with a mean and variance adjusted chi-square (WLSMV). Evaluation criteria included common fit indices (i.e., chi-square, root mean square error of approximation (RMSEA), comparative fit index (CFI), Tucker-Lewis index (TLI)). Models with a large chi-square ( $p < .05$ ) were examined for misfit.

**Results:** The GMH measurement model was a good fit to the data in 2009 [m (sd) = 50.01 (9.98);  $X^2 = 215.02$ ,  $p < .001$ ; RMSEA = .05, 90% CI (.05, .06); CFI = 1.00, TLI = 1.00] and in 2011 [m (sd) = 50.00 (9.99);  $X^2 = 174.62$ ,  $p < .001$ ; RMSEA = .05, 90% CI (.04, .05); CFI = 1.00, TLI = 1.00]. Additionally, the GPH measurement model was a good fit to the data in 2009 [m (sd) = 50.00 (9.99);  $X^2 = 302.42$ ,  $p < .001$ ; RMSEA = .04, 90% CI (.04, .05); CFI = 1.00, TLI = .99] and in 2011 [m (sd) = 50.00 (9.99);  $X^2 = 404.30$ ,  $p < .001$ ; RMSEA = .05, 90% CI (.05, .06); CFI = 1.00, TLI = .99]. Both the GMH and GPH measures displayed invariance across sexes (RMSEA values ≤ 0.06, CFI/TLI values ≥ 0.95), age groups (RMSEA values ≤ 0.06, CFI/TLI values ≥ 0.95) and over time (RMSEA values ≤ 0.06, CFI/TLI values ≥ 0.95). There was no evidence of misfit in any of the models.

**Discussion:** The PROMIS GMH and GPH measures maintained their psychometric properties within the CPS-II NC sample. Their functional equivalence across sex, age, and time ensures that measurement error does not substantively contribute to observed associations between GMH/GPH and demographic, behavioral, genomic, clinical or psychosocial factors.

Erin Peregrine Antalis

University of Michigan

erperegr@med.umich.edu

Denise Reinke, M.S., N.P. M.B.A., Department of Internal Medicine, University of Michigan

Laurence H Baker, D.O., Department of Internal Medicine, University of Michigan

Anao Zhang, PhD, LCSW, School of Social Work, University of Michigan

**Using PROMIS tools to assess change in patients' experiences of anxiety, depression, and pain among cancer survivors in a survivorship clinic.**

**BACKGROUND:** Sarcomas account for less than 1% of adult solid malignant cancers, but over 20% of all pediatric solid malignant cancers. The Sarcoma Survivorship Clinic was established in 2015 to evaluate high risk sarcoma patients for the long-term and late effects of cancer and its treatment on survivor's physical, social, and emotional well-being.

**METHODS:** Patients are seen annually. Each patient electronically completes NIH's PROMIS (Patient-Reported Outcomes Measurement Information System) questionnaires (on Anxiety 8a, Depression 8a, Pain Interference, Sleep Disturbance 6a, Physical Function). The PROMIS measures are used primarily as a clinical tool to support patient care. Patients are counseled on measures to improve their mental, physical and social health in the survivorship clinic annual appointments. Survivorship care also entails diagnosing and managing hypertension, obesity, lipid disorders, diabetes. Survivorship care plans are distributed which include screening schedules, test results, and recommendations to reduce risks associated with cancer therapies, especially chemotherapy. All responses are recorded in a confidential REDCap database. All patients signed an IRB approved informed consent.

**RESULTS:** A total of 67 patients have had an enrollment visit into the cohort ('baseline'), between October 2014 to 2018, with 31 and 21 patients completing 12 and 24 month visits. The median age at diagnosis was 26.5 years (range 2-67), and the median age at baseline clinic visit was 42 years (range 18-82). 60% are female sex, 99.5% identify as white/Caucasian, and 91.5% identify as non-Hispanic. 63% had a soft tissue sarcoma and 37% had a bone sarcoma. Treatment history includes chemotherapy (86%), surgery (94%), and radiation (54%). Patients reported positive changes in pain (3.1), depression (3.2), physical function (4.6), and sleep (2.9) between baseline and 24 months, with a modest change in anxiety (1.8).

**CONCLUSION:** Our preliminary findings from this ongoing study suggest that survivors report a positive change over time in their experience of pain, depression, physical function, and sleep. Next steps will be to assess the influence of annual disease-specific survivorship care on the overall health and well-being.

Janel Hanmer

University of Pittsburgh Medical Center

hanmerjz@upmc.edu

Stefanie Altieri Dunn, Jennifer Freel, Andrew Bilderback, Johanna Bellon, Parthasarathy Thirumala. All at UPMC

**PROMIS domain scores are associated with utilization in an outpatient sample of Neurology patients.**

**OBJECTIVES:** To evaluate associations between PROMIS scores and hospital utilization in outpatient Neurology clinic patients in a large medical system.

**METHODS:** Eight PROMIS domains have been routinely collected in Neurology practices across a large medical system: Anxiety, Cognitive Function, Depression, Fatigue, Pain Interference, Physical Function, Sleep Disturbance, and Ability Participate in Social Roles. We used the first scores collected from any patient between May and December 2018 to examine associations to inpatient stays and emergency department (ED) visits within 30 days before or after PROMIS data collection. Inpatient stays excluded inpatient rehab, inpatient hospice, and maternity admission. ED visits excluded transfers, direct observation visits, and labor and delivery.

PROMIS scores were analyzed by 5-point intervals between scores of <35, 35-39.9, 40-44, . . . , ≥65. Pearson's chi-squared tests were used to determine differences between PROMIS scores and inpatient stays or ED visits for each PROMIS domain.

**RESULTS:** We analyzed 13084 unique patients from 16 outpatient clinics. The average age was 56 and 59% were female. Approximately 4% of patients had at least 1 inpatient stay. Of these stays 313 occurred in the 30 days before data collection and 240 occurred in the 30 days after. Significant differences ( $p < 0.05$ ) between PROMIS scores and inpatient stays before data collection were observed for Cognitive Function, Pain Interference, and Physical Function. Significant differences between PROMIS scores and inpatient stays in the 30 days after data collection were observed for Anxiety, Cognitive Function, Fatigue, Pain Interference, Physical Function, and Social Roles. Lower function or higher symptom burden was associated with higher rates of inpatient stays after data collection. In contrast, before data collection, lower pain interference scores were associated with more inpatient stays. For Cognitive Function and Physical Function, an inpatient stay in the 30 days before data collection was associated with less functioning except at the highest level of function.

About 10% of patients had at least one ED visit. Of these visits, 899 occurred in the 30 days before PROMIS data collection and 527 occurred in the 30 days after. Significant differences ( $p < 0.001$ ) between PROMIS scores and ED Visits were observed for all PROMIS domains for visits both before and after data collection. Lower function or higher symptom burden was associated with higher rates of ED visits for both 30 days before and after PROMIS data collection.

**CONCLUSIONS:** We found significant associations between PROMIS domain scores with both inpatient stays and ED visits in patients visiting outpatient Neurology clinics. Associations were stronger between PROMIS scores and inpatient stays occurring 30 days after data collection compared to 30 days before. All PROMIS domain scores were associated with ED visits in both 30 days before and after data collection. This study provides evidence of construct validity for PROMIS scores in this population and

evidence that PROMIS scores may provide useful information in clinical analytics and identifying patients for targeted interventions, such as reducing inpatient admissions.

Jeffrey E. Cassisi, Ph.D.

University of Central Florida

jeffrey.cassisi@ucf.edu

Le-Chu Su, M.D., Emily Ross, M.A., & Helize Vivier (all University of Central Florida)

### **Impact of a Gluten Free Diet on Celiac Disease as Measured by the NIH PROs**

Celiac disease is an autoimmune disorder which affects the digestive process of the small intestine. Celiac disease is estimated to affect at least 1% of the population, or approximately 3 million people in the United States. Celiac disease results in significant GI distress, but it has been associated with cognitive decline and a variety of psychological symptoms and cognitive deficits such as “brain fog.”

When a person with celiac disease consumes gluten, a protein found in wheat, rye and barley, the individual’s immune system attacks the small intestine which reduces the absorption of nutrients into the body. Once diagnostic test results are collected and there is confidence the diagnosis of Celiac Disease (through serology, biopsy, and phenotypic analysis), the primary treatment is to place patients on a Gluten Free Diet (GFD).

We will be presenting our treatment protocol as applied to two new cases of probable Celiac Disease in the UCF Health GI Clinic. This will include description of Patient Characteristics, Clinical/Lab Measures, and Measures of Patient Reported Outcomes (PROMIS-GI, NIH Toolbox, and Neuro-QoL).

The NIH PRO Measures used here demonstrate that we can assess both specific GI and systemic symptoms with detail in the outpatient setting without disrupting clinic flow. Careful assessment of symptoms demonstrates that lifestyle changes (implementing a GFD) has benefits on general functioning (both cases) and specific GI symptoms (second case). Especially exciting is that in both cases, cognitive functioning as measured by the NIH Toolbox improved within 6-months of implementation of the GFD. This suggests that any cognitive decline associated with the disorder is reversible with the GFD. The change in cognitive flexibility occurred in both a young and elderly patient.

Causality cannot be determined with Case Studies or Clinical Series, but they can suggest areas for future research. Many mechanisms of action for the GFD are possible: 1.) Decrease in immune response and recovery of the gut epithelium in the small intestine may improve both nutrient absorption and availability to the CNS (direct effect on cognition). 2.) Improved nutrient absorption may reduce fatigue which in turn increases cognitive, emotional, and social functioning (fatigue as a mediator). 3.) Implementation of a GFD may increase a patient’s sense of self-control and efficacy which in turn decreases the interference of cognitive and social functioning by mood (self-efficacy and depression as mediators).

Jin-Shei Lai

Northwestern University

js-lai@northwestern.edu

Sally Jensen, Northwestern University; Joel Charrow, Ann & Robert H. Lurie Children's Hospital of Chicago; Robert Listernick, Ann & Robert H. Lurie Children's Hospital of Chicago

### **Symptom Burden Reported by Children and Adolescents with Neurofibromatosis Type 1 Associated Plexiform Neurofibroma**

**Objective:** Children with neurofibromatosis type 1 are likely to experience poorer health-related quality of life (HRQOL) than their peers. Though NF-1 children who have plexiform neurofibroma (pNF) have been considered to be experiencing relatively inferior HRQOL, most studies were performed in NF-1 children irrespective of pNF status. To fill this gap, we evaluated HRQOL reported by children with pNF. Recent research utilized interviews with pediatric patients, parents and clinicians to develop a conceptual framework including five domains important to include when assessing pNF outcomes: pain, social functioning, physical function impact, stigma, and emotional distress; instruments to measure these domains are available in PROMIS and Neuro-QoL. This paper reported results from our study to evaluate QOL reported by children with pNF using the PROMIS® and Neuro-QoL.

**Study Design:** 140 children with pNF age 8-17 completed a battery of patient-reported outcome measures from the PROMIS (Anxiety, Depression, Psychosocial Stress Experiences, Fatigue, Pain Interference, Meaning and Purpose, Positive Affect, Peer Relationships, Physical Function-Mobility) via computerized adaptive tests (CAT) and an 8-item Stigma from the Neuro-QoL via online platform. T-scores for each measure were compared to U.S. population norms. Test-statistics and analysis of variance (ANOVA) were used to evaluate factors impacting children's perceived symptom burden.

**Results:** Participants reported significantly worse mean scores than the population norms on eight of 10 domains. Patients reported Fatigue (mean=50.3) and Pain Interference similar to that which was reported by the general population. Children with at least one family member having a diagnosis of neurofibromatosis type 1 and those having pain reported significantly worse symptoms and functioning on all domains. Boys reported significantly worse pain interference, stigma, meaning and purpose, mobility function, and upper extremity function than girls. Children with chronic itch, those diagnosed under 5 years old, numbers of plexiform neurofibromas, numbers of cafe-au-lait spots, cognition (attention and learning problems) all were found to be significantly associated with one or more domains measured by this study.

**Conclusions:** Results showed that children with pNF reported significant symptom burden, highlighting the importance of monitoring children's quality of life over time in clinical research and practice. Significant factors contributing to the symptom burden were identified, which can assist in developing interventions to alleviate the negative impacts from pNF. Future research should evaluate the replicability of these findings and evaluate the validity of the PROMIS and Neuro-QoL measures in relation to clinical characteristics among children with pNF.

Jonathan Silverberg, MD, PHD, MPH  
Northwestern University Feinberg School of Medicine  
Jonathanisilverberg@gmail.com

Jin-Shei Lai, PhD, Robert W. Kantor, MD, Prarthana Dalal, BS, Catherine Hickey, BS, Sara Shaunfield, PhD, Karen Kaiser, PhD, Helena Correia, David Cella, PhD

### **Development, validation and interpretation of the PROMIS Itch Questionnaire**

Current patient-reported outcome measures for itch are limited and may not capture its full impact on quality of life. We sought to develop, calibrate and validate banks of questions assessing the burden of itch as part of Patient-Reported Outcomes Measurement Information System® (PROMIS®). A systematic process of literature review, content-expert review, individual interviews, and testing in a sample of 600 adults led to analyses using classical test theory and item response theory (IRT) models. Exploratory and confirmatory factor analyses were followed by IRT model and item fit analyses. Four itch-related item banks were developed: (1) general concerns, (2) mood and sleep, (3) clothing and physical activity, and (4) scratching behavior. IRT and expert content review narrowed the item banks to 25, 18, 15 and 5 items, respectively. Scores are reported on a T-score metric, with a T-score of 50 being the average score for people with chronic itch drawn from the US population. Validity of the item banks was supported by good convergent validity with moderate to strong correlations of item bank T-scores with average and peak numerical rating scale of itch (Pearson correlation,  $p < 0.0001$ ) and discriminant validity with itch intensity as judged by analysis of variance and receiver operator curve (ROC) characteristics, very good internal consistency and no significant floor or ceiling effects. Severity thresholds were selected based on maximizing the concordance probability in ROC analysis. In conclusion, the PROMIS Itch Questionnaire (PIQ) banks have excellent measurement properties and may efficiently and comprehensively assess the burden of itch.

Kaitlin Casaletto

UCSF

kaitlin.casaletto@ucsf.edu

Any Umlauf, Ian Abramson, Donald Franklin, Robert Heaton, UCSD

**NIH Toolbox Cognition Battery is sensitive to HIV-associated neurocognitive disorders: An examination across administration platforms**

The HIV Neurobehavioral Research Program (HNRP) at the University of California at San Diego has been investigating the use of the NIH Toolbox Cognition Battery (NIHTB-CB) in detecting HIV-Associated Neurocognitive Disorder (HAND). Prior research at the HNRP and elsewhere has demonstrated that somewhere between 35-50% of HIV-infected (HIV+) adults will demonstrate HAND on comprehensive neuropsychological testing, with the variability in outcome being affected by disease and treatment histories, as well as by types and severities of comorbid conditions. To date, HNRP studies have completed NIHTB-CB assessments of 613 participants. 530 of these (404 HIV+, 126 HIV-) were examined via the web-based Assessment Center, and 83 participants (57 HIV+, 26 HIV-) using the newer iPad app. HIV+ and HIV- participants had comparable demographic characteristics (mean age approximately 50 years, mean education approximately 14 years, and approximately 60% non-Hispanic Whites), except for sex (HIV+ approximately 87% males, versus 61% for HIV-). Data analyses are ongoing and the following findings will be updated. For the larger samples with web-based assessments, statistically significant differences based upon HIV serostatus were obtained on the following, demographically-corrected T-scores: Flanker ( $p < .001$ ;  $d = .43$ ), DCCS ( $p < .001$ ;  $d = .35$ ), Picture Sequence Memory ( $p = .002$ ;  $d = .33$ ), Pattern Comparison ( $p = .004$ ;  $d = .30$ ), List Sorting ( $p < .05$ ;  $d = .20$ ), Fluid Cognition Composite ( $p < .001$ ;  $d = .47$ ), and the Total Cognition Composite ( $p < .001$ ;  $d = .44$ ). As expected, there were no significant differences based upon HIV serostatus on the NIHTB-CB Picture Vocabulary, Oral Reading, or the Crystallized Cognition Composite. Using a 1-SD cutoff on the comprehensive HNRP neuropsychological (NP) test battery as the "gold standard," 47% of the total HIV+ group (web and iPad combined,  $N = 461$ ) were NP-impaired, versus 16% of the total HIV- group ( $N = 152$ ). Among both HIV+ and HIV- participants, those identified as impaired per the gold standard HNRP battery demonstrated lower performances across all NIHTB-CB T-scores, including crystallized measures. This suggests that (1) the NIHTB-CB is sensitive to HIV-associated cognitive impairment and may be useful as a screen for HAND, but that (2) consideration may be given to adjusting for an estimate of premorbid intelligence (e.g., NIHTB Oral Reading T-scores) when selecting cutoff scores for Fluid Cognition measures to classify disease-related NP-impairment (e.g., as proposed by Holdnack, Tulskey et al., 2017, *Rehabilitation Psychology*, 62(4), 496-508).

Katherine Hackett  
Temple University  
katherine.hackett@temple.edu

Robert Krikorian, PhD<sup>b</sup>; Tania Giovannetti, PhD<sup>a</sup>; Josefina Melendez-Cabrero, PhD<sup>c</sup>; Aneela Rahman, BA<sup>f</sup>; Emily E. Caesar, BA<sup>d</sup>; Jaclyn Chen, BSE<sup>e</sup>; Hollie Hristov, FNP<sup>f</sup>; Alon Seifan, MD, MS<sup>g</sup>; Lisa Mosconi, PhD<sup>f</sup>; Richard S. Isaacson, MD<sup>f</sup>; <sup>a</sup> Department of Psychology, Temple University, Philadelphia, PA, USA <sup>b</sup> Department of Psychiatry & Behavioral Neuroscience, University of Cincinnati College of Medicine, Cincinnati, OH, USA <sup>c</sup> Department of Neurology, Weill Cornell Medicine, San Juan, Puerto Rico, USA <sup>d</sup> Loyola-Stritch School of Medicine, Chicago, IL, USA <sup>e</sup> Stony Brook University School of Medicine, New York, NY, USA <sup>f</sup> Department of Neurology, Weill Cornell Medicine and NewYork-Presbyterian, New York, NY, USA <sup>g</sup> Compass Health Systems, Miami, FL, USA

### **Validity and Utility of the NIH Toolbox for assessment of prodromal Alzheimer's and dementia**

**INTRODUCTION:** The NIH Toolbox Cognition Battery (NIHTB-CB) is a computer-based protocol not yet validated for clinical assessment in memory disorders populations. The Weill Cornell Memory Disorders/Alzheimer's Prevention Clinic has used the NIHTB-CB since October 2014 as part of a larger clinical and cognitive protocol for patients seeking risk reduction and treatment services for Alzheimer's disease. Assessing the validity, utility and feasibility of the NIHTB-CB in a real-world clinic setting is an important next step in exploring the scalability and applicability of the NIHTB-CB in various healthcare settings.

**METHODS:** We administered the NIHTB-CB and traditional neuropsychological tests to 247 Memory Disorders & Alzheimer's Prevention Clinic patients with subjective cognitive decline (SCD), mild cognitive impairment (MCI), mild dementia due to Alzheimer's disease (AD), and normal cognition (CN). Principal component analysis (PCA), partial correlations, and univariate general linear model (GLM) tests were performed to assess construct validity. Discriminant function analyses compared classification accuracy. Certain features of NIHTB-CB protocol were adapted to improve ease of administration and test completion among individuals with more severe cognitive impairment, notably by forgoing administration of the Picture Sequence and List Sorting tasks and supplementing a delayed recall trial within the Rey Auditory Verbal Learning Task. Validity of the NIHTB-CB was explored both with and without these adaptations.

**RESULTS:** PCA identified three conceptually coherent factors: Memory (MEMNIH), Executive Function (EFNIH), and Crystallized Intelligence (CINIH). These factors were strongly associated with corresponding traditional tests and differed across diagnostic groups as expected. Both NIHTB and traditional batteries yielded strong overall discriminative ability (>80%). Inclusion of a delayed recall subtest substantially improved the validity and utility of the NIHTB-CB.

**DISCUSSION:** The NIHTB-CB is an efficient and valid method to assess neurocognitive domains pertinent to aging and dementia and has utility for applications in a memory clinic setting. Most patients had little trouble acclimating to the computer-based test procedure. Some NIHTB-CB subtests were too difficult for individuals with mild dementia, for which inclusion of a delayed list learning task was more appropriate. The NIHTB-CB was administered by technicians, scored by the program software and interpreted by licensed neuropsychologists and neurologists who were familiar with each individual's clinical case. NIHTB fully adjusted scores along with measures of crystallized intelligence were leveraged

to characterize cognitive abilities relative to premorbid intelligence, and were helpful for diagnostic feedback.

Kathleen Fear, PhD

University of Rochester Medical Center

kathleen\_fear@urmc.rochester.edu

Chris Dasilva, University of Rochester Medical Center; Judith F Baumhauer MD MPH, University of Rochester Medical Center; Michael Bass, Northwestern University

### **Evaluating clinical impact of a standard error reduction (SER) based stopping rule for PROMIS CAT instruments**

The University of Rochester Medical Center has been collecting PROMIS data since February, 2015, and in that time, patients have recorded more than 1.8 million scores across 16 PROMIS adult domains, 5 parent proxy domains and 7 pediatric domains. Most of these scores were generated from PROMIS CAT instruments, which dynamically administer questions until a standard error of 3 or less is reached with at least 4 questions answered. While the average patient burden is 4-6 questions using the approach, some patients never reach the standard error threshold and as a result answer up to 12 questions, the maximum. An alternative to the standard stopping rule is one based on standard error reduction (SER). The SER algorithm calculates the expected standard error reduction from answering additional questions. If the expected reduction is too low, then the evaluation is completed, irrespective of how many questions have been answered. This approach has the potential to greatly reduce the number of patients answering a high number of questions, with the potential tradeoff of increasing standard errors and reducing precision.

**Objective:** To establish the impact of an SER-based stopping rule on a clinical population, we examine score similarity (how well do scores generated using the standard stopping rule agree with those generated by SER?); score quality (how does the quality of PROMIS scores, individually and in aggregate, change when using SER vs the standard method?); and patient burden (how much, if at all, does SER reduce the amount of patient effort required to produce a score compared to the standard method?).

**Methods:** To test this alternative stopping rule, we re-scored all existing evaluations using the SER algorithm. For each evaluation, a new score and standard error was generated, and the number of questions that would have been answered as well as whether the evaluation would have been completed were recorded.

**Results:** SER showed strong agreement with scoring based on the standard stopping rule across adult domains. The overall score distributions were essentially identical, and a majority of SER scores and standard errors in each adult domain overlapped the corresponding original scores and standard errors by 80% or more. Root mean square error was increased by an average of 0.4. The number of items administered decreased by at least one item in 10 of 14 adult domains, with an expected time savings of up to 16 seconds per evaluation, nearly a 25% decrease.

**Conclusion:** SER reduces patient burden while preserving key characteristics of the overall score distribution while only reducing precision minimally.

Marsha Treadwell, PhD  
UCSF Benioff Children's Hospital Oakland  
mtreadwell@mail.cho.org

Julie Kanter MD, University of Alabama Birmingham; Matthew Smeltzer PhD, University of Memphis;  
Robert Gibson PhD, Augusta University; Jane Hankins MD, St. Jude Children's Research Hospital;  
Raymona Lawrence DrPh, Georgia Southern University; Ellen Werner PhD, National Heart, Lung and  
Blood Institute

### **ASCQ-Me and PROMIS in the Sickle Cell Disease Implementation Consortium Needs Assessment: Follow-up**

**Background:** The Sickle Cell Disease Implementation Consortium (SCDIC) is the first research program within the National Heart, Lung and Blood Institute to use implementation science to promote the adoption of research findings into healthcare. The goal of the SCDIC is to identify and address barriers to quality care in sickle cell disease (SCD) within eight funded sites across the U.S. for individuals with SCD ages 15 – 45 years. In its first phase, the SCDIC conducted a needs-based community assessment that included measures of quality of care and patient reported outcomes measures (PROMs) for adolescents and adults with SCD. In Phase II, we will use implementation science to drive the design of intervention studies that will be conducted to address issues of importance identified from the needs assessments.

**Objective:** To describe preliminary findings using a uniform system of common data elements, including PROMs, for data collection within the SCDIC needs assessment.

**Methods:** SCDIC representatives identified measures from the phenX Toolkit (demographics, acute healthcare utilization, SCD genotype and SCD Self-Efficacy), and measurement to assess patient experiences of care (ASCQ-Me Quality of Care) and patient health metrics (ASCQ-Me Pain Episode Frequency and Severity, PROMIS Pain Interference) that we used across sites. De-identified data was housed at the data coordinating center.

**Results:** The needs assessment enrolled 440 adolescents and adults with SCD (mean age 27.8 ± 9 years; 56% female; 97.5% African American; 95% non-Hispanic; 70% HbSS or HbSβ0 thalassemia). While 38% of patients reported 3 or more treat-and-release ED visits for pain in the past six months, 75% reported severe pain episodes at home for which they did not seek healthcare. The mean PROMIS Pain Interference T-score was 59.2 ± 9.9 and the mean Sickle Cell Self-Efficacy Scale score was 30.8 ± 7.7. On the ASCQ-Me Quality of Care, 67% of 379 participants reported they had delayed or avoided going to the ED when they thought they needed care in the past 12 months. The majority of participants (84.2%) were “usually or always” satisfied with their usual care providers and scheduled appointments, but almost half (49.1%) were “never or only sometimes” satisfied with ED care received in the previous 12 months. Thirty-six percent of this sample gave their overall healthcare the highest possible ratings and 34% gave it the lowest.

**Conclusions:** The comprehensive HealthMeasures systems, as well as the phenX Toolkit provide us with the foundation for understanding the needs of adolescents and adults with SCD. This foundation underpins the development and evaluation of strategies that take a multi-modal, multi-sector approach to address the longstanding and pervasive disparities that adolescents and adults with SCD face in

accessing quality healthcare. Further, we are populating a registry (current enrollment n = 2250 of the targeted 2400) that includes clinical and utilization information, and PROMs. The SCDIC Registry utilizes common data elements and identified measures from the phenX Toolkit, PROMIS, ASCQ-Me and Neuro-QoL and we have generated potential research questions that could be answered from the registry data.

Maureen Cottom, MBA, PMP  
Shriners Hospitals for Children  
mcottom@shrinenet.org  
Kristi Guarino and Ashley Sellers, Shriners Hospitals for Children

### **Integrating Pediatric PROMIS Measures into Clinical and Technology Operations of Shriners Hospitals for Children Health System**

Shriners Hospitals for Children (SHC) has integrated Pediatric PROMIS (Physical Function, Pain Intensity, Pain Interference, Peer Relationships, and Upper Extremity Function) measures into 11 SHC locations. PROMIS measures are a standard of care tool used in operational workflow and technologically integrated into the system's electronic medical record. Across these 11 hospitals, we have captured over 45,000 PROMIS assessments and achieved a 92% completion rate during the period April 2017 through November 2018. Providers report PROMIS administration has increased patient-provider communication, given ability to uncover issues otherwise missed, and a way to quantifiably measure change in patients' health-related quality of life.

Mehmet Kazgan  
cliexa, Inc.  
mkazgan@cliexa.com

### **Utilizing PROMIS Measures for Pain Management**

cliexa® addresses the gap in disease management that has created significant challenges in the healthcare system: the relationships between clinicians, payers and patients. A continual trend of ineffective tools and cumbersome implementation has restricted physicians and hospitals from fully leveraging and realizing the clinical and financial benefits of patient-reported data. By enabling patients and physicians to track disease activity, the platform can leverage validated assessment tools in an easy-to-use interface that allows for high-quality screening. The system has most recently integrated IoT device data and built up a back-end compliance systems to automate the process of identifying patients who may qualify for specific billing codes.

cliexa-EASE utilized PROMIS measures for pain management tracking for the 2019 AHRQ Step Up Application Challenge and was awarded third place for the established logic and use of the tools. The platform works in three key ways: 1) Logging and tracking clinically validated assessment data from PROMIS measures (using both short form and CAT functionality); 2) creating a pipeline for medication reconciliation and complication tracking; 3) allowing the patient to connect to wearable devices, payer data and their patient portals. The tool leverages both the advanced logic of the PROMIS measures as well as wellness, medication, complication and prior treatment or hospitalization data to provide clinicians with the highest quality view of the patient's condition. In user testing, 83% of patients indicated that the assessment process and navigation was "simple or basic" allowing them to move through the application with ease, in part due to the PROMIS assessment structure.

The PROMIS measures offer a number of benefits for a digital tool because of both the validation, as well as, the innovative Computer Assisted Testing API. The API allowed us as a system to deliver the questions and answers in the validated format, while the CAT format ensures that patients would not be spending time on inapplicable questions, allowing for a higher volume of screening in a short amount of time. This approach is a unique way for clinicians to further embrace patient reported data and will likely increase patient engagement.

Additionally, cliexa works with a number of clinics utilizing a wide range of metrics and has identified the standardization of tools as a challenge when communicating across health systems or between clinicians. With PROMIS and the Health Measures, the measurement of common biomarkers via digital tools can be standardized and thus cross-compared.

Michael Bass

Northwestern University

michael-bass@northwestern.edu

Scott Morris, Illinois Institute of Technology & Richard Neapolitan, Northwestern University

### **Multidimensional Computer Adaptive Testing with PROMIS Item Banks**

The Patient Reported Outcomes Measurement Information System (PROMIS) initiative developed a broad array of high-quality PRO measures. PROMIS employs Item Response Theory (IRT) and Computer Adaptive Testing (CAT) in order to reduce the number of items administered during clinical and research-based assessments. The IRT/CAT implementation in PROMIS is unidimensional in that there is a separate set of questions administered for each measured trait. However, there are often correlations among traits. Multidimensional IRT (MIRT) and multidimensional CAT (MCAT) provide items concerning several correlated traits, and should ameliorate patient burden. We developed an MIRT model and MCAT software using existing PROMIS item banks and compared the efficiency of the MCAT approach to the unidimensional approach. In addition to investigating model efficiency, we also considered the impact of alternative stopping rules and item selection algorithm have on overall efficiency measured by administered item count in simulation studies. We will also demonstrate computational complexity and its effect on software responsiveness in a real-world implementation of MCAT, using the Assessment Center API.

Molly McGuire

Texas Scottish Rite Hospital for Children

molly.mcguire@tsrh.org

Chiu-Feng Yap, Texas Scottish Rite Hospital; Brenda Matthews, Texas Scottish Rite Hospital; Harry Kim, Texas Scottish Rite Hospital

## **Planning for a multicenter longitudinal pediatric implementation of PROMIS standard of care data collection**

### Introduction

The International Perthes Study Group (IPSG) is a consortium of 40+ surgeons dedicated to the study of children impacted by Perthes Disease. Perthes, which is typically diagnosed between ages 6 and 8, is the painful avascular necrosis and subsequent regrowth of the hip's femoral head. Treatment for Perthes includes containment, or even surgical procedures. The IPSG validated the Patient-Reported Outcomes Measurement Information System (PROMIS) for Perthes in prior work, but the feasibility and operational framework for implementing PROMIS as standard of care systematically across IPSG member sites has not been described. The purpose of this study was to assess clinical feasibility, document obstacles, and disseminate a set of recommendations for the IPSG members regarding collection of standard of care PROMIS data.

### Methods

The lead site conducted informal interviews with PROMIS outcomes coordinators and experts familiar with the themes of PROMIS iterations, sub-domain options, validation methodologies, implementation issues, and scoring interpretation. Subsequently, the lead site developed and trialed PROMIS as standard of care in an outpatient Perthes clinic at Texas Scottish Rite Hospital for Children in Dallas, Texas for an 8-month period. Volunteer and clinic staff administering the measures underwent training prior to launch. PROMIS data collection included short form versions of the Physical Function, Anxiety, Depressive Symptoms, Fatigue, Pain Interference, and Peer Relationships sub-domains, assigned to either the patient (self-report) or the parent (proxy report) depending upon the patient's age. Surveys were automatically assigned and collected via the electronic health record during Perthes-related clinic visits for new patients or follow-up. Timestamps were collected separately.

### Results

Of 192 eligible visits that occurred from April 20th – December 20th, 133 surveys were administered and completed, indicating a 69.30% completion rate. Only 14 were collected during new patient visits, while the majority (89.58%) were completed during follow-up visits at the clinic. Self-report surveys made up 75.19% of the data collected and required 7 minutes and 8 seconds to complete, compared to proxy report surveys, which required 6 minutes 43 seconds. In a given month, 17 surveys were completed on average, a number which increased moderately over time. Severe scores in the depressive symptoms (n=2) and anxiety (n=8) sub-domains respectively would have triggered follow-up with psychology or social work. Between the first and second surveys of patients with multiple visits (N=31), we observed 108 minimally important differences (>3 points) across sub-domains. Challenges related to implementation mentioned by interviewees included execution of institutional contracts, workflow development, and consensus on sub-domains of interest. Both academic institutions and hospital

interviewees however, expressed that automatic scoring, CAT capability, and the ability to complete surveys without research staff involvement all lead to the generation of more data, make standard of care implementation worthwhile.

#### Conclusion

The lead site was able to institute changes in real-time to address technical/logistical errors in standard of care implementation of PROMIS in a clinic setting. These adjustments, upon dissemination to the multicenter sites, will facilitate standardization of data collection, administration, scoring, and interpretation not only for research, but for improved patient care.

Natacha Akshoomoff  
University of California, San Diego  
nakshoomoff@ucsd.edu

### **Use of the NIH Cognition Toolbox with Pediatric Samples**

In this presentation I will present an overview of the studies our group has conducted with typically developing children and adolescents (ages 3-21). Issues to be discussed are the age-related changes we see in typically developing children across the various Toolbox tests, administration issues, emergence of different factors with development, and preliminary data from the ABCD study. In addition, I will present some longitudinal data that present challenges for interpretation.

Raluca Petrican

Rotman Research Institute/University of Toronto

raluca.petrican@gmail.com

Hedvig Söderlund \*, Uppsala University; Namita Kumar, Baycrest Centre for Geriatric Care, Toronto, Ontario, Canada; Zafiris J. Daskalakis, Centre for Addiction and Mental Health, Clarke Division, Toronto, Ontario, Canada; Alastair Flint, University Health Network, Toronto, Ontario, Canada, and University of Toronto; Brian Levine, Rotman Research Institute and University of Toronto

## **Using the NIH Toolbox Cognition and Emotion Batteries to Predict Neural Markers of Psychopathology**

Existing theoretical frameworks suggest that psychopathology can be regarded as one extreme of an individual differences continuum. If the distinction between “normality” and “abnormality” is quantitative rather than qualitative in nature, then successful therapeutic interventions may work by boosting or suppressing the expression of specific, naturally occurring neurocognitive profiles relevant to optimal versus sub-optimal functioning. Consequently, studies of typically developing individuals can provide critical insights into the mechanisms conferring vulnerability versus resistance to psychopathology. To test these hypotheses, we focused on electroconvulsive therapy (ECT), a widely used and effective treatment for refractory depression, whose therapeutic neural underpinnings remain poorly understood. Here, we targeted a core cognitive deficit associated with depression, which is reliably ameliorated through ECT, specifically, the ability to learn visuospatial information. Thus, we pursued two goals. First, we tested whether ECT can “normalize” the functional brain organization patterns associated with visuospatial memory and whether such corrections would predict post-ECT improvements in learning visuospatial information. Second, we investigated whether, among healthy individuals, stronger expression of the neural pattern susceptible to adjustments through ECT would predict reduced incidence of depression-relevant cognition and affect. Thus, in a task fMRI study with a clinical and a healthy comparison sample, we characterized two functional connectome patterns: one that typifies trait depression (i.e., differentiates patients from healthy individuals) and another that is susceptible to “normalization” through ECT. The latter was specific to an autobiographical memory retrieval context, in which participants were asked to visualize as vividly as possible past personal events. As predicted, its stronger post-ECT expression was linked to improvements in visuospatial learning, as measured by the brief visuospatial memory test (BVMT-R). Subsequently, using task fMRI data and scores on the DSM-oriented scales, as well as on the NIH Toolbox Cognition and Emotion (Negative Affect) Batteries, we tested whether the two identified neural profiles would be linked to depression-relevant mentation and affect in healthy young adult participants in the Human Connectome Project (HCP) (N= 333). Results revealed that the functional brain organization of healthy participants with greater levels of subclinical depression and anxiety, more negative current emotional experience and poorer fluid cognition, particularly as applied to visuospatial information, shows greater similarity to the trait depression neural profile and reduced similarity to the ECT-correctable neural profile, as identified in the patient sample. The aforementioned brain-behavior associations emerged most strongly in learning-relevant task contexts (working memory, perceptual relational processing). Our findings are thus compatible with the hypothesis that some successful therapeutic interventions, such as ECT, work by “normalizing” the expression of adaptive traits which are spontaneously observed in the general population. As such, our results call attention to the need to develop assessment tools sensitive enough to detect subclinical variations in psychopathology-relevant traits. In this respect, our findings

suggest that the NIH Cognition and Emotion toolboxes may be useful in the early identification of individuals who are at chronic or acute risk for developing psychopathology, as well as in monitoring the progress of different therapeutic interventions.

Rina S. Fox

Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine  
rina.fox@northwestern.edu

Richard C. Gershon: Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine; Jerry Slotkin: The Center for Health Assessment Research and Translation, University of Delaware; Jennifer J. Manly: cTaub Institute for Research on Alzheimer's Disease and the Aging Brain, Department of Neurology, College of Physicians and Surgeons, Columbia University; Ellen M. Roney: Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine

### **The Spanish version of the NIH Toolbox: Overview and Development**

The NIH Toolbox for Assessment of Neurological and Behavioral Function (NIH Toolbox) was developed with particular sensitivity to the growing Hispanic/Latino populations in the United States, both English and Spanish speaking. During the development process, a Cultural Working Group (CWG) was convened to ensure that all included measures were culturally and conceptually appropriate for use with Hispanics/Latinos, among other underrepresented groups, in both English and Spanish. The CWG also reviewed all English-language NIH Toolbox measures in depth to identify barriers to cross-cultural validity and to ensure appropriateness for use with Hispanics/Latinos, as it was anticipated that many members of this group would elect to complete the NIH Toolbox in English. In addition, a Spanish Language Working Group was convened to conduct a translatability assessment of all English-language NIH Toolbox measures. Language-specific content within the Sensation, Motor, and Cognition Batteries consists primarily of test administrator demonstration and script recitation. As such, these measures were translated following a modified version of the Functional Assessment of Chronic Illness Therapy (FACIT) translation methodology applicable for use in more simplified translations. In instances where the potential for language interpretation could impact scores, such as with the Emotion measures and a limited number of survey measures from the other domains, a more rigorous translation and cultural adaptation process was used. Additionally, given that language development and usage differs greatly by culture, the Spanish versions of the Cognition Battery language measures (i.e., Picture Vocabulary Test, Oral Reading Recognition Test) were developed independently of their English counterparts. Tests were normed on a national sample ranging in age from 3-85 years, with individuals who preferred reading in Spanish. Of note, given that less than 2% of Spanish-speaking children in the United States between the ages of 8 and 17 speak Spanish as their primary language, it was anticipated that very few school-aged children would elect to complete study participation in Spanish versus English. Therefore, only Spanish-speaking children between the ages of 3 and 7, and adults between the ages of 18 and 85, were included in the norming sample. Certain conditions should be noted when implementing the Spanish-language version of the NIH Toolbox. For example, because the English and Spanish versions of the language measures were developed independently, their scores cannot be compared or combined within a single sample. Further, certain test administration requirements (e.g., requiring participants to place their hand in a specific location between trials) applicable to a subset of the included measures may be unfamiliar to individuals from various cultural backgrounds, and therefore additional instruction and reinforcement may be required. Additionally, due to differences in articulation time across languages, the Spanish-language version of the NIH Toolbox requires a slightly longer administration time than the English-language version. Despite these considerations, the Spanish-language version of the NIH Toolbox provides a much needed set of tools that can be selected as appropriate to

complement existing protocols being conducted with the growing Hispanic/Latino population in the United States.

Shelley Collins  
Shriners Hospital for Children-Salt Lake City  
scollins@shrinenet.org

## **Evaluating the social health of pediatric patients using PROMIS Peer Relationships: Social work intervention thresholds.**

### Introduction

In 2016, Shriners Hospital for Children-Salt Lake City implemented a process to routinely measure pediatric outcomes. The outcomes packet is administered at every clinic visit and consists of four PROMIS domains including the Peer Relationships short form. Children  $\geq 8$  complete a self-assessment. Caregivers complete proxy assessments for children aged 5-7 and for children who are unable to self-report. Submitted results are immediately available in the electronic medical record.

To identify children who may need support, social workers initially were notified of scores two standard deviations below the mean of 50 ( $\leq 29.99$ ) and they attempted to meet with the family during the appointment or with a telephone follow up.

After a few weeks of encounters with patients reporting scores  $\leq 29.99$ , the care team expanded encounters to include children scoring  $\leq 34.99$  to explore needs in this additional population.

### Methods

Social workers were notified of scores  $\leq 34.99$ . Encounter notes were evaluated and placed in categories:

#### Level 0-Low value encounter

- Child/proxy misunderstood the questions/scale
- Child had no interactions with peers over the last 7 days
- Child is delayed but well connected with resources

#### Level 1-Valuable encounter

- Child/family experiencing some social/mental health issues, some counseling provided in clinic
- Child is delayed, social workers evaluate needs, some advice about resources given

#### Level 2-High value encounter

- Child/family experiencing social/mental health issues requiring referral and follow-up
- Child is delayed, referrals and new resources given

### Results

Data was separated into self-report and proxy-report categories. The mean age of children who self-reported was 12; and the mean age of children with a proxy report was 9. The total number of self-

assessments was 1,907; 1.6% (31) scored  $\leq 29.99$ , and 2.8% (54) scored 30-34.99. The total number of proxy assessments was 506; 4% (20) scored  $\leq 29.99$ , and 5.3% (27) scored 30-34.99.

Social work notes were evaluated on 47 patients who scored  $\leq 34.99$ . 36% (17) had a Level-0 encounter, 32% (15) had a Level-1 encounter, and 32% (15) had a Level-2 encounter. Of 85 children with a score  $\leq 34.99$ , 29% (25) children had a recorded diagnosis of cognitive delay, autism spectrum disorder, or ADHD.

In the proxy-assessed population, 22 patients who scored  $\leq 34.99$  had a recorded social work encounter; 36% (8) had a Level-0 encounter, 45% (10) had a Level-1 encounter, and 18% (4) had a Level-2 encounter. Of 47 children with a proxy-assessment score of  $\leq 34.99$ , 79% (37) had a recorded diagnosis of cognitive delay, autism spectrum disorder, or ADHD.

### Conclusions

- Scores  $\leq 34.99$  are a small percentage of overall scores
- A higher percentage of children in the self-report population had Level-2 encounters
- Scores  $\leq 34.99$  in the self-reported population were less likely to be associated with cognitive delays, autism, or ADHD

Social workers appreciated the ability of PROMIS to identify children who may need support and felt the most valuable encounters were with teenagers. Our care team has modified the original social work threshold policy, currently any child over the age of 10 who self-reports and has a score of  $\leq 34.99$  is contacted by a social worker.

Sherif M. Badawy, MD, MS

Division of Hematology, Oncology and Stem Cell Transplant, Ann & Robert H. Lurie Children's Hospital of Chicago; Department of Pediatrics, Northwestern University Feinberg School of Medicine

sherif.badawy@northwestern.edu; sbadawy@luriechildrens.org

Leonardo Barrera, BA, MPH, Division of Hematology, Oncology and Stem Cell Transplant, Ann & Robert H. Lurie Children's Hospital of Chicago

### **Home-based Assessment of Patient Reported Outcome Measures Using a Smartphone App Platform: A Feasibility Study**

**Background:** Sickle cell disease (SCD) is the most common genetic disorder in the United States, seen in 100-120,000 Americans. SCD Complications include pain episodes, chronic anemia and long-term end organ damage. These complications result in significant declines in health-related quality of life (HRQOL) and other patient-reported outcomes (PROs) across the lifespan. However, PROs are not routinely monitored in the clinical setting or at home in SCD, and the ideal frequency of HRQOL assessment remains unclear. Prior studies suggested that frequent PROs assessments result in patient survey fatigue.

**Aims:** (1) To evaluate the feasibility and acceptability of the assessment of patients HRQOL at home using smartphones with PROMIS<sup>®</sup>-CAT measures integrated into a SCD-app; (2) To examine the effect of the frequency of HRQOL assessments on participants' completion rate over 24-week period with HRQOL evaluated every 2 weeks (Group 1) vs. every 4 weeks (Group 2); and (3) To explore participants' experience and preferences with the process and the frequency of HRQOL assessment at home using their smartphones with PROMIS<sup>®</sup>-CAT measures integrated into a SCD-app.

**Hypothesis:** The assessment of patients' HRQOL at home using a SCD smartphone application platform (app) is feasible and acceptable, and that less frequent assessments of HRQOL at home will have an overall higher completion rate when compared to more frequent ones.

**Methods:** In this pilot randomized trial, patients and their parents were enrolled from comprehensive sickle cell clinic at Lurie Children's Hospital of Chicago. Patients were eligible if they were 12 years or older and had a SCD diagnosis. Participants were randomly assigned to either Group A (every 2 weeks) or Group B (every 4 weeks) HRQOL assessment using PROMIS<sup>®</sup>-CAT measures using our SCD-app. At enrollment, participants had SCD-app downloaded and set-up on their smartphones and completed demographics and technology comfort questionnaire. At the end of the study, participants completed a semi-structured interview as well as usability and acceptability questionnaires.

**Results:** A total of 42 patients participated (57% males, 91% Black, age [mean±SD] 15.7±3 years old) with 94% enrollment rate. Overall, HRQOL assessment completion rate was 65% for patients and 47.9% for parents (P=0.13). Completion rates were significantly higher in Group B [every 4 weeks] compared to Group A [every 2 weeks] among patients only (71.7% vs. 59.3, P=0.005) and all participants [patients/parents] (65.4% vs. 45.5%, P<0.001), respectively. Similar findings were seen among parents with trend towards significance (Group B [58.3%] vs. Group A [37.5%], P=0.09). Participants who completed assessments using iPads had significantly higher completion rates compared to iPhones (100% vs. 45.2%, P<0.001), respectively. Similar findings were seen among participants who installed

SCD-app at home compared to those who did so in clinic (83.3% vs. 47%,  $P < 0.001$ ), respectively. Acceptability and usability scores were high among participants (86-100%).

Conclusions: The completion of HRQOL assessments at home using PROMIS<sup>®</sup>-CAT measures integrated into SCD-app is feasible and acceptable. Completion rates were significantly higher with less frequent HRQOL assessment (every 4 weeks) and using iPads. Future longitudinal studies to integrate routine HRQOL assessments in patients with SCD care are warranted.

Sherif M. Badawy, MD, MS

Division of Hematology, Oncology and Stem Cell Transplant, Ann & Robert H. Lurie Children's Hospital of Chicago; Department of Pediatrics, Northwestern University Feinberg School of Medicine

sherif.badawy@northwestern.edu; sbadawy@luriechildrens.org

Leonardo Barrera, BA, MPH; Robert Liem, MD, MS; Ann & Robert H. Lurie Children's Hospital of Chicago;

Jane Holl, MD, MPH; David Cella, PhD; Northwestern University Feinberg School of Medicine

### **Longitudinal Relationship Between Hydroxyurea Adherence and Health-related Quality of Life among Adolescents with Sickle Cell Disease**

**Background:** Sickle cell disease (SCD) is the most common genetic disorder in the United States, seen in 100-120,000 Americans. SCD Complications include pain episodes, chronic anemia and long-term end organ damage. These complications result in significant declines in health-related quality of life (HRQOL) across the lifespan. Hydroxyurea (HU), an FDA-approved medication for SCD, reduces morbidity and mortality, improves HRQoL and lowers healthcare utilization. However, adherence to HU remains suboptimal and it's unclear how non-adherence affects HRQOL over time and the direction of that relationship is not yet known.

**Aim:** To assess the longitudinal relationship of HU adherence to stakeholder selected HRQOL domains, including fatigue and depression.

**Hypothesis:** Low HU adherence is associated with impairment of HRQOL domains.

**Methods:** In this pilot randomized trial, patients were enrolled from comprehensive sickle cell clinic at Lurie Children's Hospital of Chicago. Patients were eligible if they were 8 years or older, had a SCD diagnosis, and were on HU with a stable dose for 2 months or more. Study measures included PROMIS®-CAT measures for different HRQOL domains, adherence visual analogue scale (VASdose), modified Morisky Adherence Scale 8-items (©MMAS-8), demographics surveys and other patient-reported outcomes (PROs). Measures were completed at baseline and every 3 months over 12-month period with a total of 5 visits (0, 3, 6, 9 and 12 months).

**Results:** A total of 16 patients participated (62% females, 100% Black, age [mean±SD] 16.4±3.1 years old). Using VASdose, patients reported hydroxyurea adherence (mean±SD) of 71.6±21.3% with 11 (69%) reporting low adherence (< 90%). Using ©MMAS-8, parents reported hydroxyurea adherence of 4.7±1.9 with all 16 (100%) reporting low/moderate adherence. The number of adherence barriers inversely correlated with adherence level using VASdose ( $r_s = -0.7$ ,  $P=0.003$ ) and ©MMAS-8 ( $r_s = -0.79$ ,  $P<0.001$ ), suggesting higher hydroxyurea adherence in patients with fewer barriers. Participants' satisfaction scores with positively correlated with their adherence levels ( $r_s=0.46$ ,  $P=0.07$ ). Participants' stigma scores inversely correlated with their adherence levels ( $r_s = -0.5$ ,  $P=0.04$ ), suggesting lower hydroxyurea adherence with more perceived SCD stigma. Using PROMIS®-CAT measures, participants reported median (IQR) score of 56.5 (45.5–62.2) for fatigue, pain behaviour 42.5 (23.9–52.9), pain interference 42.6 (32–62.6), physical function mobility 47.8 (44.2–52.1), physical activity 47.1 (37.6–49.2), physical stress symptoms 62.3 (54.2–66.1), psychological stress symptoms 54.9 (46.2–60.8), depression 52.8 (37.8–55.8), anxiety 44.4 (36.4–54.2), social isolation 41.8 (34.6–49.7), cognitive function 50.3 (44.7–55.3), positive affect 46.4 (37.4–52.9), meaning and purpose 46.6 (35.4–51.4), life satisfaction 43.8 (36.6–49.7), self-efficacy for medications 42.9 (38.5–45) and self-efficacy for symptom management

46.3 (44.3–49.3). Participants' hydroxyurea adherence correlated with PROMIS®-CAT scores for pain interference ( $r_s=-0.57$ ,  $P=0.02$ ), psychological stress symptoms ( $r_s=-0.74$ ,  $P=0.001$ ), depression ( $r_s=-0.67$ ,  $P=0.005$ ), anxiety ( $r_s=-0.5$ ,  $P=0.04$ ) and social isolation ( $r_s=-0.46$ ,  $P=0.07$ ), suggesting worse HRQOL with lower adherence. Participants' higher adherence rates were associated with better self-efficacy scores ( $r_s=0.53$ ,  $P=0.03$ ). Longitudinal data collection is ongoing.

Conclusions: Patients with fewer barriers to hydroxyurea adherence were more likely to have higher adherence rates. Routine assessment of HRQOL and hydroxyurea adherence as well as its related barriers could provide actionable information to improve adherence rates, HRQOL, and other important clinical outcomes.

Juhee Cho

Center for Clinical Epidemiology, Samsung Medical Center, Seoul, Korea

juheecho@jhu.edu

Youngha Kim and Junghee Yoon (Department of Digital Health, SAIHST, Sungkyunkwan University & Center for Clinical Epidemiology, Samsung Medical Center, Seoul, Korea)

## **Translation and Linguistic validation of the 6 PROMIS profile domains for adults in Korean**

### Background

The Patient-Reported Outcomes Measurement Information System (PROMIS) is more valid, reliable and precise measurement system than previously PROMs that assess self-reported patient's physical, mental, and social health. The purpose of the study was to translate and linguistically validate the six PROMIS profile domains for adults into Korean: Fatigue, pain intensity, pain interference, physical function, sleep disturbance, and ability to participate in social roles and activities.

### Methods

A total of 268 items were translated into Korean according to the Functional Assessment of Chronic Illness Therapy (FACIT) multilingual translation methodology including forward and back-translations, independent reviews of translation quality, and pilot-testing including cognitive interviews (n=61). During the cognitive interviews, participants first completed approximately XX PROMIS-K questions and were then interviewed to evaluate the conceptual equivalence of the translation to the original PROMIS English language source. Interview probes addressed comprehension, clarity and ease of judgement. Three rounds of interviews were conducted. Items that met the a priori threshold of  $\geq 20\%$  of respondents with comprehension difficulties were considered for rephrasing and retesting.

### Results

Most respondents understood the items and were able to interpret for Korean version of PROMIS profile domains. During the cognitive interviews, we found 27 items that were difficult to comprehend for at least 20% of respondents. Reasons for problem with comprehension were cultural or living environment difference, confused by examples, use of uncommon and unfamiliar words, sentence structure, misunderstanding the meaning, and so on. These items were revised and tested with additional 6 patients. All patients had no problems to understand modified items.

### Conclusions

The Korean version of the PROMIS profile domains has been linguistically validated in Korean adults. Further studies related to the validity and reliability of the Korean version of PROMIS profile domains can now be initiated.

Faraz Ahmad  
Northwestern University  
Faraz Ahmad  
faraz.ahmad@northwestern.edu

Michael A. Kallen, Kathryn L. Jackson (Northwestern University); Karen E. Schifferdecker, Kathleen L. Carluzzo (Geisel School of Medicine at Dartmouth); Peter A. McCullough (Baylor University Medical Center); Stephen E. Kimmel (University of Pennsylvania); Elliot S. Fisher (Geisel School of Medicine at Dartmouth); Susan E. Yount, David Cella (Northwestern University)

**Developing clinically-relevant, patient-centric health status measure for patients with heart failure built on PROMIS®.**

**Background:**

Bringing together generic and heart failure (HF)-specific items in a publicly-available, patient-reported outcome measure may facilitate better health status comparisons across groups and within individuals longitudinally in learning health systems and clinical research studies.

**Description of Works in Progress:**

We performed a mixed-methods study to develop and validate the PROMIS®-Plus-HF profile measure, a HF-specific instrument based on the generic The Patient-Reported Outcomes Measurement Information System (PROMIS). We created The PROMIS-Plus-HF profile measure—a complete assessment of physical, mental, and social health. The PROMIS-Plus-HF measure comprises 86 items (64 existing; 22 new) across 18 domains. We designed the measure such that a subset of domains, or the entire measure, can be used, depending on the clinical or research purpose. However, we recognize that abbreviated versions that include summary scores may be more relevant to patients, caregivers, clinicians, and researchers. Our ongoing analyses focus on developing a short form version of the PROMIS-Plus-HF profile measure using standard approaches. We are also employing novel multidimensional approaches (i.e., multidimensional item response theory (MIRT) and multidimensional CAT (MCAT)) to combine items within a domain and item sets across domains to develop overall summary physical, mental, and social health status scores for patients with HF. The goal of this presentation would be to provide a brief overview of how we developed the PROMIS-Plus-HF measure and describe the ongoing analyses to develop abbreviated versions and summary scores.

Stephen J. Suss Jr., BS  
University of Pittsburgh  
sjs143@pitt.edu

Suss S, Greco C, Dodds N, Pilkonis, P, Klem, M, Johnston, K, McFarland, C, Addington, E, Schalet, B, Moskowitz, J, Correia, H, Ong, J, Morone, N, Yu, L, Daehler, S, Jackson, K, Satyshur, M, Sauer, C, Horowitz, B, Victorson D

### **Mindful Measures: Creating Item Banks of Mindfulness using PROMIS® Methodology**

**Purpose:** Mindfulness interventions are often associated with health improvements. However, measurement of mindfulness is imprecise; not only has the construct proven historically difficult to operationalize—but prior instruments have been criticized on specific psychometric issues, such as differential item function, response shift, and ceiling and floor effects. The goal of this project, *Creating and Optimizing Mindfulness Measures to Enhance and Normalize Clinical Evaluation (COMMENCE)*, is to develop self-report measures of mindfulness and related constructs using the rigorous methodology of the National Institutes of Health (NIH) Roadmap initiative, Patient-Reported Outcomes Measurement Information System (PROMIS®). The current abstract describes the methods used in identifying the most relevant conceptual domains of mindfulness and the development of an initial pool of items.

**Methods:** Procedures for developing the model of mindfulness and related constructs included: 1) an extensive literature search to obtain existing measures and additional conceptual materials; 2) six focus groups (3 consisting of mindfulness teachers, 3 consisting of meditators from various traditions) and 54 online surveys of meditators, mindfulness researchers, and meditation teachers to gather content ideas; and 3) qualitative coding of these focus groups and online survey transcripts. Approximately 80 descriptions of mindfulness concepts were identified and subsequently aggregated into conceptual clusters by research team members via an online sorting task. Next, a series of 12 interviews with mindfulness experts (mindfulness researchers and secular and Vipassana meditation teachers) helped ensure comprehensiveness of content in the emerging model. Finally, items from existing instruments were reviewed for relevance and sorted into the clusters and sub-clusters.

**Results:** The initial literature searches yielded 3,871 unique abstracts which included 229 existing instruments with a cumulative total of 4,570 items assessing mindfulness and potentially related concepts. Items were either eliminated due to irrelevance or categorized into one of the conceptual clusters that had emerged from qualitative analysis of focus groups, surveys, and interviews. The conceptual domains included 5 clusters and 12 sub-clusters: Cluster 1a:) Awareness; 1b) Attention/Observation/Concentration, 2a) Acceptance/Allowance/ Letting be, 2b) Nonreactivity, and 2c) Flexibility/Perspective-taking, 3a) Being a better person/Skillful action/Ethics/Values and 3b) Open-heartedness, 4a) Curiosity, 4b) Wise/Clear-mindfulness, and 4c) Impermanence/Direct experience, and 5a) Universality/Interconnectedness and 5b) Spaciousness. A second round of binning (sorting) and winnowing (removing) items reduced the pool of items further. Ongoing tasks include revision of items and writing new ones, cognitive interviews to insure content clarity, and calibration of items

using item response theory (IRT) and classical test theory methods in a sample of internet respondents (n = 4,200) and identified meditators (n = 500).

Conclusion: Item pools developed through this process reflect a broad and readily testable conceptualization of mindfulness and related constructs. In employing the rigorous instrument development applications based on PROMIS<sup>®</sup>, COMMENCE will create a comprehensive and standardized measurement system for mindfulness and related constructs that will further advance mindfulness research.

Anna Fishbein

Ann & Robert H. Lurie Children's Hospital of Chicago

afishbein@luriechildrens.org

F.J. Penedo (University of Miami Health System, Miami, FL, USA) E. Furgis (Ann & Robert H. Lurie Children's Hospital Department of Allergy, Division of Pediatric Allergy & Immunology, Chicago IL, USA)

C.B. Forrest (Department of Pediatrics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA)

A.S. Paller (Departments of Dermatology and Pediatric Dermatology, Northwestern University Feinberg School of Medicine, Chicago, IL, USA)

### **PROMIS Pediatric Sleep-Wake Function Measures Correlate With Objective Sleep Disturbance in Atopic Dermatitis**

Most children with atopic dermatitis (AD) suffer from sleep-wake disturbance. Guidelines recommend assessment of sleep in AD, yet methodology is lacking. We performed a cross-sectional study of the new PROMIS Pediatric sleep measure (sleep disturbance/SDI and sleep related impairment/SRI) in child/parent dyads. Participants ages 5-17 years with AD had assessments via self (if  $\geq 8$ y) or parent-proxy for disease severity (Patient Oriented Eczema Measure (POEM)) and health-related quality of life measures. Data were correlated with Wake after Sleep Onset (WASO) on actigraphy (Actiwatch Spectrum). AD severity ( $n=41$ , 44% male, mean  $11.4y \pm 4.0$  yrs ( $\pm$ SD)) was mild 29%, moderate 44%, and severe 27%. Self-reported race was 51% White, 20% Black, 10% Asian, 17% Other and ethnicity was 24% Latino. Subjects had  $72.4 \pm 28.8$  mins of WASO,  $\sim 20$  mins more than in age-matched published controls. Child-report of SDI and parent-proxy SRI correlated best with WASO ( $r=0.46$ ,  $p=0.01$ ,  $n=31$  and  $r=0.42$ ,  $p<0.01$ ,  $n=41$ ). The Children's Dermatology Life Quality Index sleep screening question correlated with PROMIS measures (child: SDI  $r=0.55$ ; SRI  $r=0.63$  and parent: SDI  $r=0.60$ ; SRI  $r=0.63$ ,  $p<0.01$ ). 68% of children reported to sometimes/almost always/always "toss and turn at night" and 54% had "difficulty falling asleep." The SRI item most strongly associated with WASO was "It was hard to have fun because I was sleepy" (child  $r=0.66$  and parent  $r=0.54$ ,  $p<0.01$ ). Itch and severity correlated with WASO ( $r=0.37$ ,  $p=0.02$  and  $r=0.32$ ,  $p=0.05$ ). WASO was associated with parent-proxy report of pain,  $r=0.43$ ,  $p<0.01$ , but not other tested quality of life measures. PROMIS Pediatric sleep-wake function items are an efficient tool to screen/assess AD-induced sleep disturbance and deserve testing in larger AD cohorts.

Brocha Z. Stern, MOT, OTR, CHT  
New York University  
brocha.stern@nyu.edu  
Tsu-Hsin Howe, PhD, OTR, FAOTA; New York University

### **PROMIS(e) of Self-Efficacy for Managing after Distal Radius Fracture: Opportunities and Challenges Beyond Chronic Conditions**

This presentation will highlight the potential of using the PROMIS Self-Efficacy for Managing Chronic Conditions measures for individuals recovering from acute musculoskeletal trauma or surgery. In an era of patient empowerment, self-management is an important cross-cutting construct to increase the value of care, including maximizing outcomes while minimizing costs. Employing an ongoing study as an example, the speaker will discuss opportunities and challenges for expanding administration of the Self-Efficacy for Managing measures to musculoskeletal populations beyond chronic conditions.

The objective of the in-progress doctoral dissertation is to describe and understand self-management in adults ages 45-74 after distal radius fracture (DRF) using a convergent mixed methods design. A conceptual framework of self-management is used to describe associations between non-physiological factors and perceived health. It complements biomedical models in explaining variation in health outcomes after DRF. A key variable of the self-management process is self-efficacy, defined as an individual's level of confidence to engage in specific health-promoting behaviors. One aim of the ongoing research is to explore associations between perceived health and self-efficacy for managing the sequelae of DRF. Self-efficacy has been discussed as theoretically relevant after DRF but has rarely been empirically examined in this clinical population. A limitation to investigation has been a lack of appropriate instrumentation. The current study incorporates three domains of the PROMIS Self-Efficacy for Managing Chronic Conditions item banks: Managing Symptoms, Managing Emotions, and Managing Daily Activities. Other PROMIS measures used in the study to capture elements of physical, mental, and social health include Pain Interference, Depression, Anxiety, and Ability to Participate in Social Roles and Activities. All PROMIS measures are administered as short forms. Use of measures designed for chronic conditions after an acute injury is innovative and has its merits.

The PROMIS Self-Efficacy for Managing Chronic Conditions measures are useful but require minor adaptation for studying recovery after DRF. Many of the health behaviors specified in the domains are relevant after DRF. Despite inclusion of the word "chronic" in the measures' title, chronicity is not included in the definition of the overarching construct. In addition, there are no timeframes associated with specific items that mandate chronicity to respond. Compared to other self-efficacy measures, the availability of multiple domains in the PROMIS Self-Efficacy measures is more consistent with the domain-specific nature of self-efficacy embedded in social cognitive theory. Finally, the ability to construct custom short forms allows customization of a measure to different populations, including those that differ from the original target population. However, despite the potential for use after DRF, implementation challenges include the name of the measure, the reference population, and the inability to use more precise computer-adaptive testing because of the need to exclude irrelevant items for the study population. In addition, the large number of medication-related items within Self-Efficacy for Managing Medications and Treatments precludes use of that domain despite relevance of managing treatments after DRF. These existing challenges provide opportunities for further research to expand the

promise of the PROMIS Self-Efficacy for Managing measures to non-chronic musculoskeletal populations.

Devin Peipert

Department of Medical Social Sciences, Northwestern University

john.peipert@northwestern.edu

Robert Chapman (Department of Medical Social Sciences, Northwestern University ), Michelle Langer (Department of Medical Social Sciences, Northwestern University), David Cella (Department of Medical Social Sciences, Northwestern University), Jin-Shei Lai (Department of Medical Social Sciences, Northwestern University)

### **Identifying Symptoms Clusters among Pediatric Chronic Kidney Disease Patients Using PROMIS® Computer Adaptive Tests**

**BACKGROUND:** Identifying symptoms clusters (SCs) helps characterize patients' health and enhances treatment planning. The Patient Reported Outcomes Measurement Information System (PROMIS®) captures SCs in very brief computer adaptive tests (CATs) based on validated item banks more reliably than standard, one-item-per-symptom assessments.

**METHODS:** We used data from 384 pediatric chronic kidney disease (CKD) patients (42% female, mean age = 13 years) on PROMIS pediatric mobility (MOB), upper extremity functioning (UE), depressive symptoms (DEP), anxiety (ANX), and fatigue (FAT) CATs. PROMIS CAT scores are reported with a T-score metric (mean = 50, SD = 10), and higher scores indicate more of the measured construct (e.g., higher DEP scores indicate more depression). We modeled SCs at the domain level using two statistical approaches: bifactor exploratory analysis (EFA, oriented toward correlations among symptoms) and latent profile analysis (LPA, oriented towards identifying profiles of patients in which symptoms co-occur). Each PROMIS CAT T-score was entered into each model.

**RESULTS:** Mean CAT T-scores were: MOB = 51.5; UE = 50.1; DEP = 45.7; ANX = 46.3; FAT = 47.3. The bifactor EFA showed that a general factor representing overall symptom burden accounted for most of the variance (67%), suggesting that the PROMIS CATs tapped a similar construct (omega reliability = 0.88). The LPA suggested 3 SCs mapping onto symptom severity: High Burden/Low Function (MOB T-score: 44.4; UE: 44.5; DEP: 58.6; ANX: 58.0; FAT: 53.2), Average Function/No Burden (MOB T-score: 53.7; UE: 51.6; DEP: 43.0; ANX: 42.4; FAT: 49.2), and High Function/No Burden (MOB T-score: 56.0; UE: 54.7; DEP: 36.4; ANX: 36.0; FAT: 30.6).

**CONCLUSIONS:** Among pediatric CKD patients, symptoms as measured by PROMIS were clustered, and patients were characterized to different severity-based profiles. Sourcing PROMIS CATs for symptom identification is a reliable, clinically-feasible method for determining whether pediatric CKD patients experience high vs. low symptom burden.

Judy Frain

Goldfarb School of Nursing at Barnes-Jewish College

jfrain@bjc.org

Melissa Clettenberg undergraduate student, Goldfarb School of Nursing at Barnes-Jewish College

### **Exploring cognitive dysfunction in older adults living with HIV using the NIH Toolbox Cognition Battery**

There is evidence that adults living with HIV suffer from cognitive problems earlier, and more severely than persons without HIV. Functional MRI scans have shown brains of adults living with HIV can appear 10 years older than those of persons without HIV. The importance of cognitive function in this population cannot be overstated. Research has shown that in adults living with HIV 50 years of age and older, cognitive dysfunction is a primary cause of poor adherence to antiretroviral medications, which are essential to successfully managing HIV.

This pilot study examines cognitive function using the NIH toolbox cognition battery. Additional symptomology, including fatigue, depression, and sleep, areas identified as problematic for older adults living with HIV, are also measured using PROMIS measures, in addition to separate objective measures.

Forty three of 46 study participants have enrolled to date. Sixty-seven percent identified as male, 33% as female. The mean age was 57 (SD 5.83), ranging from 50 to 72 years. Education ranged from 6 years to 19 years, with a mean of 13.2 years (SD 2.58). The majority, 55%, lived alone, and the other 45% lived with others (partners, spouses, children, roommates). All participants had well-controlled HIV disease, were taking antiretroviral medications, and were in care for their disease.

The NIH Toolbox Cognition Battery was chosen as the optimal tool to evaluate multiple aspects of cognitive function, effectively compare this population to similarly situated persons without HIV, and would not be overly-burdensome on study participants. Fully-corrected T scores were chosen to measure Fluid Cognition, Crystallized Cognition, and Composite Cognition scores to a national sample with similar demographic characteristics.

Fully-corrected T scores for the normative sample had a mean of 50 (SD = 10). In our study 74.4% of participants scored below 50, with 39.5% scoring at least one SD below the norm, and 16% scoring at least 2 SD below the norm. Fluid cognition was deemed the best measure for our study because fluid abilities have less to do with prior learning and past experiences, and are more influenced by biological processes, and are sensitive to neurobiological integrity, including changes in brain functioning related to aging and disorders that alter brain structure and function. Nearly 40% of study participants scored at least one SD below the norm, suggesting the possibility of health-related, acquired cognitive impairment.

Fully-corrected T scores for Crystallized cognition composite found 58% scored below the 50 sample norm, with only 14% one SD or more below the norm. Further evidence that the cognitive dysfunction exhibited could be disease-related in nature.

Fully-corrected cognitive function composite scores, an average of fluid and crystallized scores, found 60.5% scoring below the sample norm, with 25.6% scoring at least 1 SD below the norm, and 9.3% scoring 2 or more SD below the norm.

In conclusion, this pilot study provides strong evidence that further research into cognitive dysfunction in this population could be of benefit, and could lead to development of interventions that could enhance fluid cognition, improving health and quality of life

Therese A. Nelson  
Northwestern University  
therese.nelson@northwestern.edu

Shirley V. Burton, University of Illinois at Chicago; Annette L. Valenta, University of Illinois at Chicago;  
Justin Starrin, Northwestern University

### **Sociotechnical Evaluation of PROMIS Patient-Reported Outcome (PRO) Measures in the Northwestern Memorial Hospital Urology Clinic**

**Objective:** Evaluation is an important element of implementation, enabling examination of the impact of new technology, such as PROMIS, in complex environments. Many evaluation factors revolve around sociotechnical “fit”—the human-computer interactions that are critical to successful adoption and use of systems. Our objective will be to develop and publish a sociotechnical evaluation tool that clinics using PROMIS can employ early in an implementation to identify benefits and to correct problems expeditiously.

**Methods:** To improve evaluation in health informatics, several theoretical frameworks and reporting protocols have been proposed to help standardize and create transparency, with a focus on users, technology, the organization and net benefits. Our team will build practical sociotechnical tools on these existing frameworks in order to turn insights from the literature into actionable clinical evaluation methods.

Northwestern Hospital’s Urology Clinic recently implemented PROs embedded in its Epic medical record. During February-May 2019, we will create evaluation tools informed by sociotechnical constructs and conduct a mixed-method evaluation. We will survey and interview the interdisciplinary team, including physicians, nurses, medical assistants and administrative personnel, focusing on clinician workflow and usage within the Urology Clinic.

To address the sociotechnical aspects of the implementation, we have chosen to utilize the Human-Organization-Technology Fit (HOT-fit) framework, which examines the relationships between the technological, human and environmental components of a health information system (Yusof et al., 2008). The technological dimension focuses on system, information and service quality; the human component encompasses system use and user satisfaction; and the organizational component concentrates on structure and environment (Erlirianto et al., 2015; Yusof et al., 2008). In addition, we will use two publications, Guidelines for Good Evaluation Practice in Health Informatics (GEP-HI) and the Agency for Healthcare Research and Quality (AHRQ) Health information Technology Evaluation Toolkit to guide the planning and performing of our scientific evaluation (Nykänen et al., 2011; Cusack et al., 2009).

**Results:** We will emerge from this study with both an evaluation of the Urology Clinic and with a tool that can be used for formative sociotechnical evaluation of clinical PROMIS implementations. Such a tool will allow for expeditious evaluation of project endpoints to help improve implementations in their early stages. The evaluation will delineate successes, challenges, and outcomes within the clinic, and promote early problem solving. It will focus on changes encountered in workflow and on time management, quality of care, patient/physician burden, and engagement/satisfaction. Our project will result in both a

full-featured evaluation framework and in reusable tools to help organizations examine the impact of PRO collection in the clinical milieu.

Conclusions: By constructing survey and interview tools informed by established theoretical frameworks, then testing those tools in a real-world setting, we will turn research insights into practical methods of evaluating the implementation of PROMIS in clinical settings. Concentrating on sociotechnical fit, organizations will be able to quickly evaluate and problem solve in the formative stages of new PROMIS clinical implementations. This work will lay the groundwork for future testing of tools in additional real-world settings.

Maggie Morris  
Bravado Health  
mmorris@bravadohealth.com

## **Patient Engagement and PROMIS**

### Problem

Hospital readmissions in the United States cost the healthcare industry \$15-20 billion annually. Patient non-compliance contributes significantly to the problem and leads to sub-optimal outcomes. It is estimated that more than half of patients don't follow their discharge plan. One-third of patients never even fill their prescriptions. Many of these readmitted patients are suffering from one or more chronic conditions. It is estimated that 59 percent of all Americans have at least one chronic condition, costing the healthcare industry \$2.7 trillion annually.

### Proposed solution

Ayva Engage is the AI-powered virtual assistant from Bravado Health designed to help patients achieve optimal wellness and to focus attention of the care team where it is needed most. Ayva guides patients through their discharge plan, raising alerts whenever plans are in danger of derailment, keeping the care team and friends and family informed and patients on track toward self-care.

Fully integrated with all leading EHRs, Ayva pulls in the patient record, Surescripts data and geo-demographics to assemble and operationalize a personalized plan for each patient. Plans are assembled from selected elements of the Engagement Plan Library, which contains the core content of our solution.

The Engagement Plan Library is a set of evidence-based care plans designed by board-certified physicians. Plan types include: condition-specific plans targeting the top ten most frequent chronic conditions, and cross-cutting plans addressing general care measures such as medication adherence, referral compliance and wound care. Point-in-time assessments are included as cross-cutting plans with biometric assessment and monitoring as indicated for each patient.

Integration of PROMIS measures as point-in-time assessment tools is envisioned as a powerful and unique enhancement to Ayva Engage. By leveraging objective, scored, validated survey instruments in the context of a friendly conversation, Ayva will provide a mechanism for iterative refinement of alert thresholds, focusing the attention of the care team on patients most in need of help while avoiding "alert fatigue" on the part of caregivers already overwhelmed with their caseloads.

Specific areas where we see the PROMIS measures to be applicable in the Ayva solution are:

- 1) Self-efficacy for Medications and Treatments as a component of the Medication Adherence plan
- 2) Anxiety
- 3) Depression
- 4) Daily dyspnea check

By embedding these measures and monitoring patient engagement and outcomes, we can more accurately detect when a patient needs help and alert the care team so that readmission is potentially averted.

Anna Fishbein

Ann & Robert H. Lurie Children's Hospital of Chicago

afishbein@luriechildrens.org

A.B. Fishbein<sup>1</sup>, F.J. Penedo<sup>2</sup>, E. Furgis<sup>1</sup>, C.B. Forrest<sup>3</sup>, A.S. Paller<sup>4</sup> ?<sup>1</sup>Ann & Robert H. Lurie Children's

Hospital Department of Allergy, Division of Pediatric Allergy & Immunology, Chicago IL, USA,

<sup>2</sup>Department of Oncology, Miami University Health System, Miami FL, USA <sup>3</sup>Department of Pediatrics,

The Children's Hospital of Philadelphia, Philadelphia, PA, USA, <sup>4</sup>Departments of Dermatology and

Pediatric Dermatology, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

### **PROMIS Pediatric Sleep-Wake Function Measures Correlate With Objective Sleep Disturbance in Atopic Dermatitis**

Most children with atopic dermatitis (AD) suffer from sleep-wake disturbance. Guidelines recommend assessment of sleep in AD, yet methodology is lacking. We performed a cross-sectional study of the new PROMIS Pediatric sleep measure (sleep disturbance/SDI and sleep related impairment/SRI) in child/parent dyads. Participants ages 5-17 years with AD had assessments via self (if  $\geq 8$ y) or parent-proxy for disease severity (Patient Oriented Eczema Measure (POEM)) and health-related quality of life measures. Data were correlated with Wake after Sleep Onset (WASO) on actigraphy (Actiwatch Spectrum). AD severity (n=41, 44% male, mean  $11.4y \pm 4.0$  yrs ( $\pm$ SD)) was mild 29%, moderate 44%, and severe 27%. Self-reported race was 51% White, 20% Black, 10% Asian, 17% Other and ethnicity was 24% Latino. Subjects had  $72.4 \pm 28.8$  mins of WASO,  $\sim 20$  mins more than in age-matched published controls. Child-report of SDI and parent-proxy SRI correlated best with WASO ( $r=0.46$ ,  $p=0.01$ ,  $n=31$  and  $r=0.42$ ,  $p<0.01$ ,  $n=41$ ). The Children's Dermatology Life Quality Index sleep screening question correlated with PROMIS measures (child: SDI  $r=0.55$ ; SRI  $r=0.63$  and parent: SDI  $r=0.60$ ; SRI  $r=0.63$ ,  $p<0.01$ ). 68% of children reported to sometimes/almost always/always "toss and turn at night" and 54% had "difficulty falling asleep." The SRI item most strongly associated with WASO was "It was hard to have fun because I was sleepy" (child  $r=0.66$  and parent  $r=0.54$ ,  $p<0.01$ ). Itch and severity correlated with WASO ( $r=0.37$ ,  $p=0.02$  and  $r=0.32$ ,  $p=0.05$ ). WASO was associated with parent-proxy report of pain,  $r=0.43$ ,  $p<0.01$ , but not other tested quality of life measures. PROMIS Pediatric sleep-wake function items are an efficient tool to screen/assess AD-induced sleep disturbance and deserve testing in larger AD cohorts.

Carly Strohbach

Ann & Robert H. Lurie Children's Hospital

cstrohbach@luriechildrens.org

Sina Malekian (Lurie Children's); Jamie Burgess, PhD (Lurie Children's & Northwestern); Cynthia LaBella (Lurie Children's & Northwestern)

### **Effect of one season of tackle football on young athletes' cognitive function: A novel use of NIH Toolbox Cognition Battery**

#### **Background:**

There is growing public concern about the risks associated with playing tackle football on young athletes' short and long-term cognitive health due to the effects of concussions and/or repeated subconcussive head impacts sustained during routine practice and games. However, there is limited data regarding the immediate effects of playing a season of tackle football on the various subdomains of cognitive function in young athletes.

#### **Purpose:**

The objective of the study was to evaluate the effects of a season of tackle football on young athletes' cognitive function, using the NIH Toolbox Cognition Battery (NIHTB-CB) in a community setting.

#### **Methods:**

Approximately 200 youth football players aged 7-14 years from a community football league were invited to participate, of which 74 (37%) enrolled. Exclusion criteria included diagnosed hearing loss, epilepsy, or developmental disability. Testing occurred one week before and one week after the 2018 football season in an urban park-district multipurpose field house. The NIHTB-CB was administered by 3 to 5 research personnel using the NIH Toolbox iPad App. Age-corrected total, fluid, and crystallized composite scores were derived from the 7 NIHTB-CB cognition tests measuring executive function and attention, episodic memory, working memory, language, and processing speed. Paired T-Tests were performed to examine the relationship between pre- and post-season scores with a significance level set at 0.05.

#### **Results:**

Forty-four of the 74 enrolled athletes (22%) completed both pre- and post-season testing and were included in statistical analyses. Mean age was  $11.04 \pm 1.7$  years and 100% were male. No head injuries were reported. There was improvement in the age-adjusted fluid composite scores from pre- to post-season ( $97.9 \pm 14.7$  vs.  $102.4 \pm 18.1$ ,  $p=0.023$ ; respectively), but no change in age-adjusted crystallized composite scores ( $110.8 \pm 14.4$  vs.  $112.3 \pm 16.1$ ,  $p=0.472$ ; respectively). Age-adjusted total scores improved from pre- to post-season ( $104.9 \pm 13.5$  vs.  $108.6 \pm 15.0$ ,  $p=0.027$ ; respectively).

#### **Conclusion:**

Similar to Rebchuk et al. (2018) and Akshoomoff et al. (2013), practice effects in fluid composite scores were expected. Our results showed no change in crystallized composite scores and an increase in fluid scores from pre- to post-season testing. Improvements may be attributed to participants being more

familiar with the test stimuli and tasks at post-season re-testing, rather than “real” improvements in cognitive functioning. The time between test administrations was dependent on the football season schedule, therefore practice effects were not methodologically controlled for. This abstract describes a novel application of the NIH Toolbox Cognition Battery in a community setting to evaluate the effects of a season of football on cognitive function.

Caroline Terwee

Amsterdam UMC, location VUmc

cb.terwee@vumc.nl

Haan EJA1,2,3, Lameijer CM4, van Bruggen SGJ4, Van Wier MF5, Willigenburg NW5, Van Deurzen DFP5, Pisters MF2, Kaat AJ6, V Stouten7, K Van der Elst7,8, Roorda LD3. 1Amsterdam UMC, Vrije Universiteit Amsterdam, Department of Epidemiology and Biostatistics, Amsterdam Public Health Research Institute, Boelelaan 1117, Amsterdam, The Netherlands; 2Physical Therapy Sciences, Program in Clinical Health Sciences, University Medical Center Utrecht, Utrecht University, The Netherlands; 3Amsterdam Rehabilitation Research Center | Reade, Amsterdam, The Netherlands; 4Amsterdam UMC, Vrije Universiteit Amsterdam, Department of Traumasurgery, Boelelaan 1117, Amsterdam, The Netherlands; 5Department of Orthopedic Surgery, Joint Research, OLVG, Amsterdam, The Netherlands; 6Department of Medical Social Sciences, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA; 7KU Leuven - University of Leuven Department of Development and Regeneration, Skeletal Biology and Engineering Research Center, Leuven, Belgium; 8University Hospitals Leuven, Department of Rheumatology, Leuven, Belgium

### **Translation and validation of the Dutch-Flemish PROMIS® v2.0 Upper Extremity item bank in patients with upper extremity disorders**

**Objective** The PROMIS® Upper Extremity (UE) item bank v2.0, consisting of 46 items, has recently been developed to replace the 16-item v1.2 version, to increase the range of upper extremity function that could be precisely measured, and to improve the psychometric properties. The v2.0 item bank showed good psychometric properties in US individuals with upper-extremity disorders. The aim of this study was to develop a Dutch-Flemish translation of the PROMIS® UE item bank v2.0, and to evaluate its psychometric properties in Dutch patients with musculoskeletal upper-extremity disorders.

**Methods** Of the 46 items, 42 are included in the v1.2 Physical Function item bank, which was already translated and validated in Dutch-Flemish. Four new items were translated according to state-of-the-art translation methodology. In this cross-sectional study, 521 patients with upper extremity disorders visiting the outpatient clinics of the department of orthopaedic surgery at the Onze Lieve Vrouwe Gasthuis (level II traumacenter, n=218) or the department of traumasurgery at the Amsterdam University Medical Center (level I traumacenter, n=303) were included. Assumptions of the IRT model (unidimensionality (confirmatory and exploratory factor analyses and bi-factor analysis), local independence (residual correlations), and monotonicity (non-parametric IRT response curves and scalability) were checked and a logistic Graded Response Model was fitted. Item fit (S-X2) and item parameters were estimated. The number of patients for whom a theta could reliably be estimated ( $SE < 0.32$ , which equals a reliability of 0.90) was calculated, based on the full item bank, as well as on the standard 7a Short Form. Differential Item Functioning (DIF) was examined for age, gender, duration of complaints, center, and country. For the latter analysis, our sample was compared to a sample of 246 US patients from an online panel, who endorsed having some difficulty due to upper extremity pain or function. Construct validity of the Dutch-Flemish PROMIS UE v2.0 item bank was evaluated by correlating UE T-scores to scores of the Disabilities of the Arm, Shoulder and Hand (DASH), the Michigan Hand Outcomes Questionnaire subscale activities of daily living (MHQ-ADL), the Functional Index for Hand Osteoarthritis (FIHOA), and the Patient-Rated Wrist Evaluation (PRWE) function subscale.

Results The assumptions for IRT were considered met (CFI: 0.93, TLI: 0.93, RMSEA: 0.10, SRMR: 0.09, ratio 1th/2nd factor: 10.7, Omega H 0.80, ECV 0.68) and all items fitted the GRM model. A theta could reliably be estimated for 97.1% of the population when using the full item bank and for 39.0% when using the 7a Short Form. None of the Dutch-Flemish PROMIS UE v2.0 items were flagged for DIF for age, gender, center or duration of complaints. Seven items were flagged for center DIF and twelve items for language DIF. Correlations with the legacy instruments were all >0.70, which supported construct validity.

Conclusion This is the first validation study of the PROMIS<sup>®</sup> UE item bank v2.0 outside of the US. The Dutch-Flemish PROMIS<sup>®</sup> UE item bank v2.0 showed good psychometric properties in patients with upper extremity disorders. The importance of the observed DIF should be further studied.

Devin Peipert

Department of Medical Social Sciences, Northwestern University

john.peipert@northwestern.edu

Richard Gershon (Department of Medical Social Sciences, Northwestern University), Jerry Slotkin (Center for Health Assessment Research and Translation, University of Delaware), Vitali Ustsinovich (Department of Medical Social Sciences, Northwestern University), Katy Bedjeti (Department of Medical Social Sciences, Northwestern University), Julie Hook (Department of Medical Social Sciences, Northwestern University), Cindy Nowinski (Department of Medical Social Sciences, Northwestern University), Sandra Weintraub (Departments of Psychiatry and Behavioral Sciences and Neurology, Northwestern University)

### **ARMADA: Advancing Reliable Measurement in Alzheimer's Disease and Cognitive Aging**

**Background:** Detecting cognitive decline in older individuals is currently among the issues of greatest concern for public health, particularly regarding dementia due to Alzheimer's disease (AD) and related disorders. AD is anticipated to increase in prevalence as the aging population grows worldwide. There is a need for brief, practical, automated measures that can be deployed in a widespread manner in diverse settings. The recently launched ARMADA (Advancing Reliable Measurement in Alzheimer's Disease and Cognitive Aging) project will help address this need by validating the NIH Toolbox for Assessment of Neurological and Behavioral function (NIHTB) in individuals over age 65 (including over age 85) who are cognitively healthy, or who have symptoms of mild cognitive impairment (MCI), or of early dementia of the Alzheimer type. In addition, to reflect the diversity of patients affected by cognitive decline, we established samples of African Americans in the relevant clinical groups and a sample of Spanish-speaking Hispanic individuals to validate the Spanish version of the NIHTB. We will compare baseline characteristics of the clinical groups. Two annual longitudinal follow ups are planned to study how the NIHTB predicts future cognitive decline.

**Method:** Samples of individuals over age 65, including a cohort over age 85, who are either cognitively normal or have cognitive impairment (mild cognitive impairment or early dementia of the Alzheimer type) were tested with Cognition, Emotion, Motor and Sensory modules of the NIHTB at baseline. The study has recruited participants with available AD biomarkers, including amyloid imaging, cerebrospinal fluid markers of amyloid and tau, and carrier status for the AD risk factor gene, ApoE4, in order to identify a behavioral phenotype of biomarker positivity.

**Result:** Baseline data collection will be completed by the 2019 HealthMeasures User's Conference, and preliminary results will be presented.

**Conclusions:** To date, the NIHTB has been well tolerated by older participants, including individuals with cognitive impairment.

Dodds, Nathan E  
University of Pittsburgh/UPMC/Northwestern University  
nathaniel.dodds@northwestern.edu

Elwy R, Gillman A, Wasan A, Johnston K, McFarland C, Kelly N, Kringle E, Kelsey MR, Schneider MJ, & Greco CM

### **Listening to the Stakeholders: Implementing HEAL contextual factor assessments into clinical settings**

**Purpose:** Treatment outcomes are influenced not only by the specific actions of a modality, but also by nonspecific or contextual factors such as the patient's views of the treatment and the provider. The Healing Encounters and Attitudes Lists (HEAL) are patient-reported measures of nonspecific or contextual factors in treatment that can influence outcomes. The HEAL were developed using PROMIS methodology and validated in over 400 patients with chronic pain receiving integrative medicine/complementary and alternative medicine (CAM) care or conventional medicine treatments (NIH R01 AT006453; PCORI ME-1402-10114). The HEAL include 5 item banks and 1 static short form: Treatment Expectancy (TEX), Patient-Provider Connection (PPC), Health Care Environment (HCE), Positive Outlook (POS), Spirituality (SPT); and a static, short form on Attitudes toward CAM. Although several patient and provider interviews endorsed the value of including HEAL measures in clinical treatment, building the measures into routine practice and using HEAL reports may raise challenges. The current study describes the early phase of HEAL implementation in seven Pain Medicine Clinics at the University of Pittsburgh Medical Center (UPMC).

**Methods:** We conducted 25 formative interviews with clinicians and staff at the seven UPMC Pain Medicine Clinics, and 13 interviews with patients receiving care at the clinics. The purpose was to obtain these stakeholders' perceptions about the use of HEAL measures 1) in UPMC Pain Clinic assessments, and 2) in conversations between providers and patients about pain management. We reviewed interviews based on concepts from the Theory of Diffusion of Innovations (Rogers, 2003) in order to better understand potential barriers and facilitators to implementing HEAL in UPMC Pain Clinics.

**Results:** We identified interview quotes that exemplified relevant key concepts from the Theory of Diffusion of Innovation. For example, whether implementing HEAL 1) has a perceived relative advantage over current practices, 2) is compatible with perceived needs, values and norms, 3) is low in complexity, and 4) is amenable to being tested out on a limited basis. Based on interviews, potential barriers such as lack of knowledge of how to address patients' treatment expectations and views of their provider were identified; thus, informing staff needs for communication skills training. Some perceived barriers informed decisions about which HEAL measures to include or not include in routine assessments. For example, staff and clinicians frequently stated that, although spirituality (SPT) and views of the Health Care Environment (HCE) are important, these are not areas that they viewed as addressable as part of pain treatment. Patients, staff, and clinicians provided input on when various HEAL assessments should take place in order to best facilitate their usefulness in clinic.

**Conclusion:** Implementation science procedures such as theory-based formative interviewing is highly relevant when considering how best to integrate research tools such as PROMIS into clinical settings.

Dr Llinos Haf Spencer and Beryl Cooleedge  
Bangor University  
l.spencer@bangor.ac.uk  
Alexander Smith, Aneurin Bevan University Health Board

## **Linguistic Validation of the Welsh Language Version of the PROMIS-10 Global Health Measure**

### Objective

Health questionnaires and assessment tools are used increasingly in the health service to help monitor patients' health status and health-related quality of life. In order to ensure validity and robust research, it is essential that health measures are offered in the participants first language. Furthermore, a rigorous approach to translation is required to ensure that health questionnaires are conceptually and functionally appropriate in the language of the participants. Simply translating does not capture the essence. It is vital that constructs have the same meaning. Different language versions need to have parity, and group comparisons should reflect true group differences rather than linguistic discrepancies or cultural bias in translation.

In the bilingual context of Wales, health questionnaires are required in Welsh as well as English. However, their availability is lacking and there is a shortage of data to describe their performance when administered to patient groups. LLAIS, based at N.WORTH, Bangor University is tasked with increasing the availability and accessibility of Welsh language versions of health measures.

### Methods

This study focuses exclusively on a step by step systematic approach to the linguistic validation of the Welsh language version of the PROMIS-10 global health questionnaire (see <http://www.micym.org/llais/static/translations.html> for more details on the translation guidelines including forward and back translation, consensus review and cognitive testing).

### Results

The cognitive interviewing stage of the step by step linguistic validation included testing out the Welsh measure with 12 service user representatives, lay participants, and healthcare professionals to ensure that the concepts and language was easily understood by native Welsh speakers.

### Conclusions

A Welsh language version of the PROMIS-10 is now available for use by clinicians and researchers from the micym.org website which hosts all available Welsh language patient reported outcome measures

<http://nworth-ctu.bangor.ac.uk/documents/PROMIS-10.pdf>.

The psychometric properties of these measures when administered to a wider patient group will be the focus of a future study.

Joel Alcantara, DC

The International Chiropractic Pediatric Association

dr\_jalcantara@yahoo.com

Whetten A, Research Consultant; Zabriskie C (Utah State University) and Jones S (Research Consultant).

## **Exploratory factor analysis of the PROMIS-29 V1.0 and the RAND SF-36: results from a practice-based research network.**

### Introduction

The PROMIS-29 is analogous to the SF-36 and similar to this legacy measure, continues in its popularity to be the most widely used patient reported outcome profile instrument to date. The PROMIS-29 has been used for quality of life measure in a plethora of patient populations. Of interest herein is our use of these instrument in chiropractic patients. To understand the underlying structure of the PROMIS-29 and the RAND SF-36 scales as it is applied to chiropractic patients, we performed an exploratory factor analysis (EFA) to uncover the implicit structure(s) within these instruments.

### Methods:

In addition to self-reported socio-demographic information and clinical covariates, we examined the quality of life of patients (≥18 years) attending care in a chiropractic practice-based research network using the PROMIS-29 V1.0, PROMIS Global Health and the RAND SF-36. We report here our EFA findings with PROMIS-29 and RAND SF-36.

### Results

The summary of factor loadings for the PROMIS-29 orthogonal 2-factor model found that anxiety, depression, and fatigue were strongly correlated with Factor 1. Pain Interference, Pain NRS, and Physical Function were strongly correlated with Factor 2. Sleep disturbance had a moderate correlation with Factor 1 while satisfaction with social role was moderately correlated with both Factor 1 and 2. We hypothesize from these results that Factor 1 and the domains Anxiety, Depression, Fatigue uncovers the mental health structure of patients. Similarly, Factor 2 uncovered that physical health structure of patients. An examination of the oblique factorization was also examined and a similar structure was identified.

Since satisfaction with social role and sleep disturbance were not highly correlated with either factor, a 3-factor EFA was constructed. For Factors 1 and 2, similar structures were identified in the orthogonal 2-factor model. Satisfaction with social role, however, was now highly correlated in the 3-factor model. Sleep disturbance remained as moderately or mildly correlated with all 3 factors.

The RAND SF-36 orthogonal 2-factor model revealed an apparent lack of stable structure. Only the physical functioning items demonstrated correlational relationships with Factor 2, and many of the domains lacked correlation with either of the factors (i.e. role limitations due emotional problems, general health, energy/fatigue). From oblique 3-factor model, we observed that Factor 1 is almost exclusively correlated with physical functioning. Factor 2 is strongly correlated with role limitations due emotional problems and physical limitations, and moderately correlated with pain. Factor 3 was strongly

correlated with role limitations due emotional problems, general health, and energy/fatigue. Social functioning was not highly correlated with any of the three factors.

#### Conclusion

From the PROMIS-29, fatigue strongly correlated more as a mental health domain while sleep disturbance was found to be part of but not exclusively a physical, mental, or social health domain. We found strong evidence that the RAND SF36 is not designed to assess mental and physical health as distinctly as the PROMIS-29.

John Parkhurst, PhD; Caroline Kerns, PhD; Anthony Vesco, PhD; John V. Lavigne, PhD  
Ann & Robert H. Lurie Children's Hospital & Northwestern University's Feinberg School of Medicine  
jtparkhurst@luriechildrens.org

### **Use of PROMIS anxiety and depression youth and parent proxy short forms in a clinical sample.**

Accurate mental health diagnosis is a prerequisite for treatment selection and is associated with better clinical outcomes (Jensen-Doss & Weisz, 2008). Evidence-based assessment is predicated on the use of reliable and validated screening and assessment measures to help identify mental health concerns across community and clinical populations. Psychometrically sound measures are also needed to understand the impact of treatment on outcomes (e.g., changes in depressive or anxiety symptoms) (APA Presidential Task Force on Evidence-Based Practice, 2006). The PROMIS Health Measures, specifically short form versions of youth self-report and parent proxy of anxiety (PROMIS-A) and depressive (PROMIS-D) symptoms, may be especially useful to both mental health and primary care providers. These forms are grounded in item response theory, are matched to the US census distribution (gender, age, race and education), and are practical (free, brief, and available in multiple languages) which make them optimal measures to use for both screening and progress monitoring.

The PROMIS-A and PROMIS-D are currently the most practical measures to be used across multiple settings considering the technical requirements of the CAT administration of PROMIS measures (e.g., having readily available electronic devices, obtaining clinic approval of data transmission, and needing to orient patients to the technology). To date, few studies have examined the reliability and validity of the PROMIS-A and PROMIS-D with a child and adolescent population. Studies have only examined the youth self-report PROMIS-A and PROMIS-D in a general pediatric outpatient and medical subspecialty clinic (not a psychiatric clinical sample) and did not collect parent proxy reports (Varni et al., 2014). The current study seeks to address this gap in the research.

The current study examined the psychometric properties of the PROMIS-A and PROMIS-D short-form youth self-report and parent proxy versions of the measures in a mental health treatment-seeking population. Convergent validity for the PROMIS-A measures will be assessed via comparisons to the youth- and parent-report Screen for Child Anxiety Related Emotional Disorders (SCARED; Birmaher, Khetarpal, Cully, Brent, & Mckenzie, 1995). Convergent validity for the PROMIS-D will be evaluated via comparisons to the youth- and parent-report Short Moods and Feelings Questionnaire (Angold & Costello, 1987). Test-retest reliability, internal consistency, and parent-child concordance will also be examined. The sample will be comprised of approximately 120 children and adolescents (aged 7-18) and their parents. Data has been collected from 80 from 30 participants thus far. Study results are expected to greatly inform the use of the PROMIS-A and PROMIS-D for both screening and progress monitoring for youth presenting for mental health treatment.

Juliet Monet Torres

Ann & Robert H. Lurie Children's Hospital of Chicago

torresju@grinnell.edu

S.R. Fair (Ann & Robert H. Lurie Children's Hospital of Chicago and Stanley Manne Children's Research Institute) , A. Zhou (Ann & Robert H. Lurie Children's Hospital of Chicago and Stanley Manne Children's Research Institute, Northwestern University) , L. Garcia (Ann & Robert H. Lurie Children's Hospital of Chicago and Stanley Manne Children's Research Institute) , R. Patel (Ann & Robert H. Lurie Children's Hospital of Chicago and Stanley Manne Children's Research Institute), F.A. Zelko (Ann & Robert H. Lurie Children's Hospital of Chicago and Stanley Manne Children's Research Institute), C.M. Rand (Ann & Robert H. Lurie Children's Hospital of Chicago and Stanley Manne Children's Research Institute), A. Bradley (Ann & Robert H. Lurie Children's Hospital of Chicago and Stanley Manne Children's Research Institute) , K. Bijawat (Ann & Robert H. Lurie Children's Hospital of Chicago and Stanley Manne Children's Research Institute, Northwestern University), S. Tsao (Ann & Robert H. Lurie Children's Hospital of Chicago and Stanley Manne Children's Research Institute, Northwestern University), M.L. Chen (Seattle Children's Hospital) , I.A. Perez(Children's Hospital of Los Angeles) , & D.E. Weese-Mayer (Ann & Robert H. Lurie Children's Hospital of Chicago and Stanley Manne Children's Research Institute, Northwestern University)

### **Utilization of NIH Toolbox Cognition Battery in a Rare Disease Conference Setting**

Background: Congenital Central Hypoventilation Syndrome (CCHS), an extremely rare (<2,000 cases reported worldwide) neurocristopathy, typically presents in the newborn period and is characterized by autonomic dysregulation and severe hypoventilation. Clinical neurocognitive assessments have indicated reduced neurocognition in CCHS cohorts of limited size, but also demonstrate a wide range of individual function, from severe neurocognitive impairment to scores in the above-average range. Further studies in larger cohorts are necessary to better understand this variability and determine intrinsic and extrinsic disease factors that may contribute to these varying neurocognitive outcomes. However, the extreme rarity of CCHS makes such studies difficult and protracted in the traditional clinical setting. The NIH Toolbox Cognition Battery (NTCB) can be quickly and conveniently administered with an iPad by a trained tester across a wide range of ages and environments. This flexibility and convenience make NTCB a possible solution for capturing robust neurocognitive data in non-clinical settings. NTCB has been utilized in CCHS patients in the clinical setting since January 2016 at Lurie Children's Hospital of Chicago and May 2017 at Seattle Children's Hospital, two leading referral Centers for CCHS. This resulted in NTCB assessment of 35 CCHS patients (4-47 years). In 2018, the CCHS Network hosted an International CCHS Conference, geared toward families and researchers, offering a singular opportunity to expand this cohort. However, the Conference was held in a setting without access to traditional clinical testing environments and with limited time available for research procedures.

Hypothesis: We hypothesized that NTCB could be successfully utilized to assess CCHS neurocognition in this setting without affecting results.

Methods: Participants were informed of the IRB-approved NTCB testing at conference sign-in. Research team members consented interested families and scheduled testing during an 8-hour window made available for research during the conference. NTCB was performed in up to six participants

simultaneously (average duration 45-minutes). Participants, wearing noise-canceling headphones, were seated at individual tables in two connecting conference rooms. Participants were compensated for participation (\$10 gift card).

Results: NTCB was performed in 32 (31 previously untested) CCHS subjects (8-38 years). No significant differences were identified in age-corrected composite scores (total, crystallized, fluid, and early-childhood cognition) between the clinically-collected and the conference-collected cohorts. Conference-collected data also replicated findings from the clinically-collected cohort demonstrating that NTCB fluid cognition scores, but not crystallized cognition scores, are significantly below population norms in CCHS.

Discussion: NTCB utilization in a conference setting allowed us to dramatically improve the robustness of available neurocognition data in CCHS, nearly doubling our cohort size (collected clinically over ~3 years) in a single day. Further, this method allowed replication of our previous findings in a cohort of patients not seen at our Centers. While testing space/conditions were not ideal, including lack of private testing spaces for each participant, comparison of composite scores between this cohort and the cohort obtained clinically (individual testing in a quiet, private room) did not reveal significant differences. These results support the potential of utilizing NTCB in similar, non-traditional settings to empower capture of neurocognitive data in rare disease populations.

Noelle E Carlozzi  
University of Michigan  
carlozzi@med.umich.edu

Angelle M. Sander (Baylor College of Medicine and TITT Memorial Hermann), Nicholas R. Boileau (University of Michigan), Michael A. Kallen (Northwestern University), Rael T. Lange (Walter Reed National Military Medical Center/Defense and Veterans Brain Injury Center), Louis French (Walter Reed National Military Medical Center/Defense and Veterans Brain Injury Center), Jill Massengale (James A Haley VA), Risa Nakase-Richardson (James A Haley VA), Robin Hanks (Wayne State University/Rehabilitation Institute of Michigan), Elizabeth A. Hahn (Northwestern University), Anna L. Kratz (University of Michigan), David S. Tulsy (University of Delaware), Jennifer A. Miner (University of Michigan), Phillip Ianni (University of Michigan), & Tracey Brickell (Walter Reed National Military Medical Center/Defense and Veterans Brain Injury Center)

### **Caregiver Vigilance and Emotional Suppression: New measures for caregivers of individuals with traumatic brain injury**

**Objective.** Caregivers of individuals with traumatic brain injury (TBI) frequently experience anxiety related to the caregiver role they are filling. Much of this caregiver anxiety is due to their feeling the need to avoid people and situations that might upset or “trigger” their care recipient. In addition, caregivers often feel pressure to maintain appearances that things are “okay”, even when they are not, when interacting with other people; this is especially true for caregivers living within the culture of the military. There are currently no measures that capture these caregiver-specific feelings; thus, this work describes the development of and preliminary validation efforts for the TBI-CareQOL Caregiver Vigilance and TBI-CareQOL Emotional Suppression item banks.

**Design.** 534 caregivers of civilians ( $n = 218$ ) or service members/veterans (SMVs,  $n = 316$ ) with TBI completed 32 caregiver vigilance items, 43 emotional suppression items, several other measures of health-related quality of life (Rand-12, PROMIS Depression, PROMIS Social Isolation, Caregiver Appraisal Scale), and the Mayo-Portland Adaptability Inventory-4 (for determining perceived functional capacity in the person with TBI).

**Results.** For Emotional Suppression, the initial item pool had 43 items. Exploratory and confirmatory factor analyses supported the retention of 25 items, and graded response model (GRM) fit analyses and differential item functioning (DIF) studies supported the retention of 21 items for the final item bank. For Caregiver Vigilance, the initial item pool had 32 items. Exploratory and confirmatory factor analyses supported the retention of 20 items; GRM and DIF fit analyses supported retaining 18 items in the final item bank. Expert review and GRM calibration data informed the selection of two 6-item static short forms (SFs) and programming of computer adaptive tests (CAT) for both item banks. Internal consistency reliabilities for the different administration formats (CAT, SF) for both measures were excellent ( $\alpha \geq .90$ ), and three-week test-retest reliability was supported across the combined sample and for the individual subgroups, for both CAT and SF administrations (Caregiver Vigilance: all  $\alpha \geq .78$ ; Emotional Suppression: all  $\alpha \geq .87$ ). The patterns of correlations among Caregiver Vigilance, Emotional Suppression, and other self-report measures supported convergent and discriminant validity. Caregivers of low-functioning individuals had significantly higher Caregiver Vigilance and Emotional Suppression scores than those

caring for high-functioning individuals, supporting known-groups validity for both administration formats (all  $ps \leq .002$ ).

Conclusions. The new TBI-CareQOL Caregiver Vigilance CAT, TBI-CareQOL Emotional Suppression CAT, and corresponding 6-item short forms were developed using established rigorous measurement development standards. These measures are part of the TBI-CareQOL measurement system, a comprehensive measurement system designed to capture HRQOL that is specific to caregivers of civilians and SMVs with TBI. This item bank development work indicates that these measures show evidence of having strong psychometric properties.

Noelle E Carlozzi  
University of Michigan  
carlozzi@med.umich.edu

Rael T. Lange, Nicholas R. Boileau, Michael A. Kallen, Louis French, Jill Massengale, Risa Nakase-Richardson, Angelle M. Sander, Robin Hanks, Elizabeth A. Hahn, Anna L. Kratz, David S. Tulsky, Jennifer A. Miner, Tracey Brickell

### **Health Care Frustration in Caregivers of Service Members/Veterans with Traumatic Brain Injury**

**Objective.** Caregivers of service members/veterans (SMVs) with traumatic brain injury (TBI) encounter a number of barriers when navigating the military health care system. This, can lead to their anger and frustration with the health care services. The purpose of this study was to develop, using established PROMIS methodology, a military-specific computer adaptive test (CAT) and associated short form (SF) as patient-reported outcome measures as related to caregivers' anger and frustration with health care services and benefits.

**Design.** PROMIS methodology was used for this study. 317 caregivers of SMVs with TBI completed 68 health care anger and frustration items, along with other measures of health-related quality of life (PROMIS Anger, Caregiver Appraisal Scale, and Rand-12), and the Mayo-Portland Adaptability Inventory-4, a measure designed to determine perceived functional capacity in a person with TBI.

**Results.** Exploratory and confirmatory factor analyses supported the retention of 58 items. Graded response model (GRM) fit analyses and differential item functioning investigations supported the retention of 38 items in the final measure. Expert review and GRM calibration data informed the selection of two 6-item short forms (one that measures frustration with services provided for caregivers, and another to measure frustration with services provided for SMVs) and programming of a computer adaptive test (CAT). Internal consistency reliabilities for the different measure administration formats (CAT, SF) were excellent ( $\alpha \geq .95$ ). Both the CAT and the two SF measures were devoid of excessive floor and ceiling effects. Strong correlations between Health Care Frustration, PROMIS Anger ( $r = 0.53$ ), and caregiver appraisals (i.e., CAS Caregiver Burden,  $r = -0.60$ ) provided good evidence of convergent validity, while moderate correlations between Health Care Frustration and mental HRQOL (i.e., Rand-12 Mental Health Composite,  $r = -0.38$ ) also supported convergent validity. Weak correlations between Health Care Frustration and physical HRQOL (i.e., Rand-12 Physical Health Composite,  $r = -0.15$ ) and between Health Care Frustration and positive aspects of caregiving (i.e., CAS Caregiver Relationship Satisfaction, Caregiving Ideology, and Caregiving Mastery,  $r = -0.26, 0.10, -0.40$ , respectively) were interpreted as evidence of discriminant validity. Caregivers of low-functioning individuals also had significantly higher Health Care Frustration scores than those caring for high-functioning individuals, supporting known-groups validity.

**Conclusions.** The new TBI-CareQOL Health Care Frustration CAT and two corresponding 6-item short forms, were developed using established rigorous measurement development standards. Health Care Frustration provide the first self-report measures developed to evaluate caregiver feelings of anger and frustration with DoD/VA services. Given the many efforts underway within the military health care system to address the needs of SMVs with TBI, and their families, these new measures have the potential to help identify caregivers (and SMVs) in need of additional, expedited, or standard health care services.

Rina S. Fox  
Northwestern  
rina.fox@northwestern.edu  
Julie N. Hook, PhD, MBA;  
Maria J. Marquine, PhD;  
Jennifer J. Manly, PhD;  
Helena Correia, BA;  
Jerry Slotkin, PhD;  
Sandra Weintraub, PhD;  
Richard C. Gershon, PhD;

### **Reliability and Validity of the Spanish Language Measures of the NIH Toolbox Cognition Battery**

**Objective:** The NIH Toolbox for Assessment of Neurological and Behavioral Function (NIHTB) was developed under contract from the National Institutes of Health to create a set of easy-to-administer neuropsychological measures for use across the lifespan (ages 3-85). The NIHTB Cognition Battery (NIHTB-CB) includes two language measures that were developed, calibrated, and normed separately in English and Spanish. This analysis presents the test-retest reliability and construct validity of the Spanish-language picture vocabulary test (S-PVT) and oral reading recognition test (S-ORRT) among adults. **Participants and Methods:** Participants were adults age 18-85 who took part in the NIHTB norming study in Spanish (N=408, Age: M=44.1, SD=16.7; Education: M=10.7, SD=4.3; 65.0% female). Of these, 48 repeated the battery 1 week later. Both the S-PVT and the S-ORRT were administered using computer adaptive testing and scored using item response theory. Spearman's correlations were used to evaluate test-retest reliability. Convergent validity was evaluated by correlating S-PVT scores with scores on the Bateria-III Woodcock-Muñoz Vocabulario Sobre Dibujos, and by correlating S-ORRT scores with scores on a 48-item version of the Word Accentuation Test. Adjusted Spearman's correlations and general linear models related scores to age, education, and sex. **Results:** Both the S-PVT ( $\rho=0.87$ ,  $p<.001$ ) and the S-ORRT ( $\rho=0.88$ ,  $p<.001$ ) demonstrated good test-retest reliability. Good convergent validity was found for both the S-PVT ( $\rho=0.76$ ,  $p<.001$ ) and the S-ORRT ( $\rho=0.65$ ,  $p<.001$ ). Scores on the S-PVT were positively related to education ( $\rho=0.38$ ,  $p<.001$ ), and scores on the S-ORRT were negatively related to age ( $\rho=-0.18$ ,  $p<.01$ ) and positively related to education ( $\rho=0.30$ ,  $p<.001$ ). **Conclusions:** The Spanish language measures of the NIHTB-CB demonstrated acceptable reliability and validity, suggesting they can be used to measure language ability among Spanish-speaking adults in the United States.

Sina Malekian

Ann & Robert H. Lurie Children's Hospital

smalekian@luriechildrens.org

Jamie Burgess, PhD (Lurie Children's & Northwestern); Cynthia LaBella, MD (Lurie Children's & Northwestern)

### **Using PROMIS instruments to evaluate emotional symptoms in children and adolescents with concussions**

#### **Purpose:**

We aimed to describe emotional symptoms in children and adolescents after sustaining a concussion using pediatric PROMIS instruments. Our objectives were two-fold: 1) compare post-concussion PROMIS anxiety (PROMIS-A) and depressive symptoms (PROMIS-DS) between sexes and across age groups, and 2) identify predictors of increased PROMIS-A and PROMIS-DS scores.

#### **Methods:**

Patients aged 8-17 years presenting to a tertiary-care pediatric concussion clinic from August 2014 to February 2018 were invited to participate. Non-English speaking patients were excluded. Participants completed PROMIS surveys for anxiety and depressive symptoms domains at their initial clinic visit. Based on general guidelines for interpretation of PROMIS T-scores, scores were dichotomized as <55 (normal-to-mild) and ≥55 (moderate-to-severe). Exposure variables in the multivariable logistic regression model were collected from the EMR, including age group (elementary (ES), middle (MS), or high school (HS)), sex, time from injury to initial evaluation, history of anxiety/depression, previous head injury, injury mechanism (non-sport or sport), and abnormal performance on Vestibular/Ocular Motor Screen (VOMS) at initial clinic visit.

#### **Results:**

Of the 614 patients approached, 114 (18.6%) declined. Of the 500 enrolled, 14 (2.8%) had incomplete data, leaving 486 (97.2%) participants for analyses (43.8% male). There were 57 (11.8%) ES students (grade 1-5, 10.1±1.1 years), 164 (33.7%) MS students (grade 6-8, 12.9±1.0 years), and 265 (54.5%) HS students (grade 9-12, 15.9±1.1 years). PROMIS-A scores were similar for females and males among ES (48; 95% CI 42-53 vs. 45; 95% CI 42-49) and HS patients (50; 95% CI 48-51 vs. 48; 95% CI 46-51). By contrast, among MS patients, females scored significantly higher than males in PROMIS-A (49; 95% CI 47-52 vs. 44; 95% CI 41-46). PROMIS-DS scores were similar for female and male ES patients (44; 95% CI 39-49 vs. 42; 95% CI 39-46). Among MS and HS patients, females reported significantly greater PROMIS-DS scores than males (46; 95% CI 43-49 vs. 41; 95% CI 39-43 and 48; 95% CI 46-50 vs. 45; 95% CI 43-48, respectively). Factors associated with increased odds of scoring in the moderate-to-severe range for PROMIS-A included female sex (OR=1.58; 95% CI 1.00-2.49), history of anxiety (OR=2.75; 95% CI 1.60-4.74), duration ≥1 month from injury to initial clinic visit (OR=2.38; 95% CI 1.38-4.08), and abnormal VOMS (OR=1.72; 95% CI 1.10-2.69). Factors associated with increased odds of scoring in the moderate-to-severe range for PROMIS-DS included female sex (OR=1.68; 95% CI 1.03-2.73), history of anxiety (OR=2.20; 95% CI 1.24-3.93), history of depression (OR=4.82; 95% CI 2.20-10.57), and both a duration >2 weeks (OR=1.81; 95% CI 1.04-3.15) and ≥1 month from injury to initial clinic visit (OR=2.07; 95% CI 1.17-3.67).

## Conclusion:

This is the first large study using PROMIS to describe differences in post-concussion emotional symptoms across age groups and sex in a pediatric population. Females scored higher than males in MS (anxiety and depressive symptoms) and HS (depressive symptoms) groups, but no sex differences were seen in ES group. Predictors of greater emotional symptoms on PROMIS were similar to those found in prior studies that used other tools to measure anxiety and depressive symptoms.

Victoria D. Kolbuck, MSW

The Potocsnak Family Division of Adolescent and Young Adult Medicine, Ann & Robert H. Lurie Children's Hospital of Chicago  
vkolbuck@luriechildrens.org

Afiya Sajwani, BA,<sup>1</sup> Timothy Johnson, PhD,<sup>2-3</sup> Frank Zelko, PhD,<sup>4-5</sup> Robert Garofalo, MD, MPH,<sup>1, 6</sup> & Diane Chen, PhD<sup>1, 5-7</sup> <sup>1</sup>The Potocsnak Family Division of Adolescent and Young Adult Medicine, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL USA; <sup>2</sup>Department of Public Administration, University of Illinois in Chicago; <sup>3</sup>Survey Research Laboratory, University of Illinois in Chicago; <sup>4</sup>Department of Pediatric Neuropsychology, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL USA; <sup>5</sup>Department of Psychiatry & Behavioral Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL USA; <sup>6</sup>Department of Pediatrics, Northwestern University Feinberg School of Medicine, Chicago, IL USA; <sup>7</sup>Department of Child and Adolescent Psychiatry, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL USA

### **Using the NIH Toolbox with Transgender and Non-binary Youth**

**Background:** To our knowledge, there are no normative standards based on transgender and non-binary reference groups for measures of emotional functioning. There is no consensus on how to use validated assessment measures with gender-specific reference norms. Some clinician-researchers have used norms based on respondents' assigned sex at birth (Edwards-Leeper et al., 2017), whereas others have advocated scoring measures using both assigned sex and affirmed gender norms or advised against using gendered norms altogether when possible (Keo-Meier & Fitzgerald, 2017). Preliminary research suggests measure interpretation may differ based on whether assigned sex versus affirmed gender norms are used (Keo-Meier et al., 2015).

The Trans Youth Decision-making Project (TYDP) is a National Institutes of Health (NIH)-funded study that examines how cognitive and emotional functioning impacts decisional capacity in the context of initiating pubertal suppression treatment. This study uses the NIH Toolbox Emotion battery. Because the automated scoring program uses gender-referenced norms, study investigators had to intentionally decide which norms to use (assigned sex versus affirmed gender). This study examines the impact of norm selection on scores for Emotion battery measures.

**Method:** Between October 2017 and December 2018, data were collected from youth who self-identified as transgender, non-binary, or gender-fluid, were in early puberty, and expressed interest in or had begun pubertal suppression treatment. Potential participants were identified via chart review of Lurie Children's Gender Development Clinic patients. Youth were also recruited from the community via parent support groups and flyers provided to community-based clinicians. Study visits included structured clinical interviews and measures of neurocognitive and emotional functioning, including the NIH Toolbox Emotion battery (Gershon et al., 2013).

**Results:** Preliminary analysis included 34 young adolescents ages 9-14 ( $M=11.79$ ,  $SD=1.26$ ; 56% birth-assigned females; 47% racial/ethnic minorities). Paired-samples t-tests were used to determine whether there were statistically significant mean differences between NIH Toolbox Emotion battery measures scored based on female and male norms.

For birth-assigned males (n=15), significant mean differences were found in Psychological Well-Being (PW),  $t(14)=3.6$ ,  $p=.003$  (female norms:  $m=45.3$ ,  $SD=10.3$ ; male norms:  $m=43.5$ ,  $SD=11.3$ ), Negative Affect (NA),  $t(14)=-9.4$ ,  $p<.001$  (female norms:  $m=51.7$ ,  $SD=12.2$ ; male norms:  $m=54.1$ ,  $SD=13.2$ ), and Social Satisfaction (SS),  $t(14)=4.4$ ,  $p=.001$  (female norms:  $m=47.5$ ,  $SD=12.0$ ; male norms:  $m=46.3$ ,  $SD=12.8$ ).

For birth-assigned females (n=19), significant mean differences were found in PW,  $t(18)=3.6$ ,  $p=.002$  (female norms:  $m=47.4$ ,  $SD=13.5$ ; male norms:  $m=46.0$ ,  $SD=14.2$ ), NA,  $t(18)=-9.8$ ,  $p<.001$  (female norms:  $m=51.4$ ,  $SD=9.1$ ; male norms:  $m=53.6$ ,  $SD=9.8$ ), SS,  $t(18)=4.9$ ,  $p<.001$  (female norms:  $m=46.3$ ,  $SD=10.3$ ; male norms:  $m=45.3$ ,  $SD=10.9$ ), and Negative Social Perception (NSP),  $t(18)=-4.0$ ,  $p=.001$  (female norms:  $m=47.8$ ,  $SD=6.8$ ; male norms:  $m=48.3$ ,  $SD=7.1$ ).

Conclusion: Psychological well-being, negative affect, social satisfaction, and negative social perception T-scores differed significantly based on whether assigned sex versus affirmed gender scoring norms were used. Interestingly, among transgirls (birth-assigned males), using affirmed gender norms (female) consistently resulted in a more positive picture of emotional functioning (higher PW, lower NA, and higher SS), whereas among transboys (birth-assigned females), using affirmed gender norms (male) consistently resulted in a more negative picture of emotional functioning (lower PW, higher NA, lower SS, and higher NSP).

Heather E. Brown

The Permanente Medical Group, Kaiser Permanente

Heather E. Brown 1, Michael A. Kallen 2, Joeffrey R. Hatton 1, William A. Doyle 1, Ryan Murphy 1, Ryan Elliott 1, Ann T. Tran 1, Mark A. Gutierrez 1, John D. Litten 1, Richard C. Gershon 2, Vincent X. Liu 1; 1

The Permanente Medical Group, Kaiser Permanente; 2 Department of Medical Social Sciences, Northwestern University

### **Clinician-reported Inpatient Physical Function: A new measure derived from PROMIS Adult Physical Function**

**Background:** Our aim was to develop a precise, score-level targeted clinician-reported inpatient physical function (PF) measure, with item content derived from existing PROMIS PF (Adult) item bank content, and with obtained scores reported on the PROMIS PF metric.

**Methods:** An expert team of psychometricians and clinicians reviewed the PROMIS PF item bank to identify items for measuring lower-level PF status (T-scores 10-50) that, collectively, offered high score-level reliability ( $\geq 0.90$ ), suggested by established PROMIS PF item performance precision estimates. Selected items were edited for clinician reporting, reviewed by clinicians external to the project, and field tested. Item response data were assessed as to meeting PROMIS measure development standards, and items were then calibrated on the PROMIS PF metric via a single-group design linking study, employing patient-reported responses to the original PROMIS PF items identified for conversion to clinician reporting. A subset of the newly calibrated items was chosen for short-form (SF) application; initial validation evidence was obtained with an independent sample.

**Results:** Nine PROMIS PF items were candidates for clinician reporting of inpatient PF status; three new items were written to meet clinician expectations for inpatient content coverage. An inpatient sample (N = 515; 55.1% female; mean age = 66.2 years) was assessed by physical therapists using the new inpatient PF items. Response data analyses indicated the item set met expected measure development standards (e.g., eigenvalue 1 = 83.8% of variance; Cronbach's alpha = 0.97). The 12 inpatient PF items were successfully calibrated on the PROMIS PF metric, using patient responses to the nine original PROMIS PF items (raw score  $r = 0.73$ ) as anchoring data. The project's expert team constructed a 5-item SF; in an initial validation study (N = 481), the 5-item SF demonstrated very good to excellent discrimination for patient discharge home without home health (46.8% of patients; c-statistic 0.78) and home including home health (76.1% of patients; c-statistic 0.87), compared with any other discharge disposition.

**Conclusions:** A precise, score-level targeted item set was developed for clinician reporting of inpatient PF status; an accompanying 5-item SF renders this measure an effective, efficient means of assessing inpatient PF.