

Outcome Measures in Orthopedics

*Jonathan Dubin, MD
Associate Professor
University of Missouri-Kansas City
Truman Medical Center*

*David K. Galos, MD
Assistant Professor
Temple University Hospital*

Objectives

- **Understand the differences in outcome measures**
- **Learn how to assess and choose an outcome measure**
- **Become familiar with the future direction of outcomes measures**

Quality Measures in Orthopaedic Trauma

- Due to complexity of polytraumatized patient, multidisciplinary care teams, and heterogeneity of injuries, outcome measures difficult to obtain
- Donabedian models categorizes Quality Measures into 3 categories
 - **Structure:** Setting of care & resources for patient e.g. available psych counseling for PTSD
 - **Process:** Implementation & execution of care e.g. time to antibiotics for open fractures
 - **Outcome:** Health and function resulting from rendered care
- Majority of current quality measures look at process
- “Outcome measures, arguably the most important category, evaluate the health and function of a patient or population as a direct result of the delivered care (e.g. mortality rate for patients with geriatric hip fractures).”
- Only 16 quality measures specific to orthopaedic trauma patients identified
 - 7 were outcome measures
 - E.g. “In-hospital deaths per 1000 hospital discharges with hip fracture as a principal diagnosis for patients aged 65 years and older”

Quality Measures in Orthopaedic Trauma

- Due to complexity of polytraumatized patient, multidisciplinary care teams, and heterogeneity of injuries, outcome measures difficult to obtain
- Majority look at process - i.e. time to antibiotics for open fractures
- However outcome measures are also useful
- “Outcome measures, arguably the most important category, evaluate the health and function of a patient or population as a direct result of the delivered care (e.g. mortality rate for patients with geriatric hip fractures).”
- Only 2 quality measures specific to orthopaedic trauma patients
 - Endorse arthroplasty for intracapsular femoral neck fracture in patients over 65
 - Endorse cephalomedullary nailing for reverse obliquity and subtrochanteric femur fractures in patients over 65 as opposed to screw and side plate constructs

Background

- **Shift in how to measure treatment “success” or “failure”**
- **Outcomes relevant to a surgeon may NOT be so for the patient**
 - e.g. Anatomically reduced and healed plafond fracture does not necessarily mean an “Excellent” result from the patient’s perspective
 - Similarly, a nonunion does not necessarily equate to a poor outcome, such as sometimes seen in clavicle fractures (or other bones)
- **“Treat the patient, not the disease”**
 - How can we define success or failure, above and beyond our images?
 - How can we quantify outcomes that depend on the subjectivity of patient responses?
 - What are the relevant outcomes? Pain? Mobility? Satisfaction? Quality of Life?

Background: Challenges and Role in Orthopedic Trauma

- Injuries tend to be heterogeneous compared to osteoarthritis which is homogeneous in nature
- Urgency in majority of cases makes it difficult to get preinjury data
- Depending on injuries e.g. multiple fractures, obtaining data may be difficult for a single injury
- Important to understand impact of traumatic injuries on patients
 - Functionally
 - Psychologically
- Objectively a measure patient's recovery

Background: Wilson and Cleary Model

- **Traditionally, medicine has focused on biologic and physiologic causes of pathology**
 - Understand causation to guide treatment
 - Randomized-Controlled Trials are the gold-standard
 - Outcomes are often “hard” e.g. lab values, imaging studies, range of motion
- **Social Sciences focus on functioning, quality of life,**
 - Understand interactions of behaviors, emotions, feelings, etc
 - Tends to be more observational due to limitations in designing “controlled” experiments
- **Wilson and Cleary developed a model integrating these concepts**

Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life: A conceptual model of patient outcomes. J Am Med Assoc 1995; 273: 59–65.



Background: Wilson and Cleary Model

- Conceptual model organizing the variables of patient outcomes into 5 “levels”
 - Biological Function
 - Symptoms
 - Functional Status
 - General Health Perceptions
 - Overall Quality of Life
- ENVIRONMENTAL & INDIVIDUAL FACTORS influence these outcomes at every level



Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life: A conceptual model of patient outcomes. J Am Med Assoc 1995; 273: 59–65.

Background: Wilson and Cleary Model

- Represents linear progression going from simple to complex
 - Cellular and organ system level → individual → individual in societal roles → Overall quality of life
 - Moving rightward necessarily means more complex interactions with psychological, social, motivational factors, among others
- No level is “preferable” to the next
- Each level represents separate facets needing measurement to provide a broad assessment of outcome

- Biological Function
- Symptoms
- Functional Status
- General Health Perceptions
- Overall Quality of Life

Background: Wilson and Cleary Model

- **Biological and Physiological Variables:** Joint space narrowing, non-union on x-ray, DEXA score, etc.
- **Symptoms:** pain, stiffness, instability, etc.
 - While strongly correlated, the relationship between biological/physiological and symptoms is inconsistent i.e. greater joint space narrowing does not necessarily mean greater pain
- **Functional Status:** Inability to bathe, ambulate to the grocery, etc.
 - Symptoms affect functional status, but other factors e.g. motivation, personality, social situation, may play a role, too



Background: Wilson and Cleary Model

- **General Health Perceptions:** assimilates all previous levels
 - affected by multiple variables, such as expectations
- **Quality of Life:** Broadest category
 - Patients can adapt to impairments i.e. functional status may not correlate strongly with happiness



Outcomes in Orthopedics

Symposium

The Outcomes Movement in Orthopaedic Surgery: Where We Are and Where We Should Go

“Clinical outcomes remain important as distinct measures of the success of orthopaedic interventions, however they should be accompanied by an assessment of functional outcomes”

Swiontkowski MF, Buckwalter JA, Keller RB, Haralson R. The outcomes movement in orthopaedic surgery: where we are and where we should go. JBS Am. 1999;81:732-40.



Background: Outcomes in Orthopaedics

Outcome Instruments: Rationale for Their Use

By Rudolf W. Poolman, MD, PhD, Marc F. Swiontkowski, MD, Jeremy C.T. Fairbank, MD, FRCS,
Emil H. Schemitsch, MD, FRCSC, Sheila Sprague, MSc, and Henrica C.W. de Vet, PhD

“**Clinical outcomes** can be subject to **interrater disagreement** and they often do not provide definitive answers about whether an intervention is useful from a patient’s perspective..... Well-designed **patient-reported instruments** that have undergone rigorous testing and may be **better validated** and have **greater reproducibility** than the so-called objective or clinical outcomes.”

Outcome Measures

- **Clinical Outcomes: tend to be “hard” or “objective”**
 - **Surrogate Endpoints: X-rays, ROM, strength**
 - **May not reflect what is important to the patient**
 - **Variable interobserver agreement**
 - **Often Create Composite Endpoint with pitfalls**
 - **Combining outcomes which can vary independently (e.g. pain and range or motion)**
 - **Inappropriate weighting of components of the composite i.e. 1 point for ROM, 1 point for return-to-work, etc.**

Outcome Measures

- **System-Specific: Developed for a single body-region or joint**
- **Disease Specific: Developed for a disease (OA, Scoliosis)**
 - Often uses both system specific (hip) and disease (OA) to assess a single patient
- **General Health Related Quality of Life (HR-QOL)**
 - Encompasses physical, social, and mental aspects
 - General instruments
 - In Ortho trauma will include all aspect of patient condition, e.g. fracture + chest injury + spine injury

Outcome Measures: Patient Reported Outcomes

- **Patient Reported(PRO):** tend to be “soft” or “subjective”
 - Because of rigorous testing often more valid and reliable than “objective” measures
 - DASH, SMFA, HOOS,KOOS, PROMIS
 - Require screening patients for mental state, substance abuse, and **ACTIVATION** (engagement in their own care) to improve interpretation of results.
 - Patient Activation Measure -13 (PAM-13)
 - Alcohol Use Disorders Identification Tests (AUDIT) and AUDIT-C (3 questions)
 - **Implementation concerns:**
 - Vendors available
 - Computer Adaptive Testing (CAT) can reduce patient burden



Outcome Measures

- **Mixed Clinician-Based and Functional Outcome Instruments – ASES, AOFAS, Harris Hip Score**
 - **Combination of questions asked to patients and exam findings by physician**
 - **High risk of interobserver variability**
 - **Often not validated**

Patient Reported Outcome Measurements

Examples

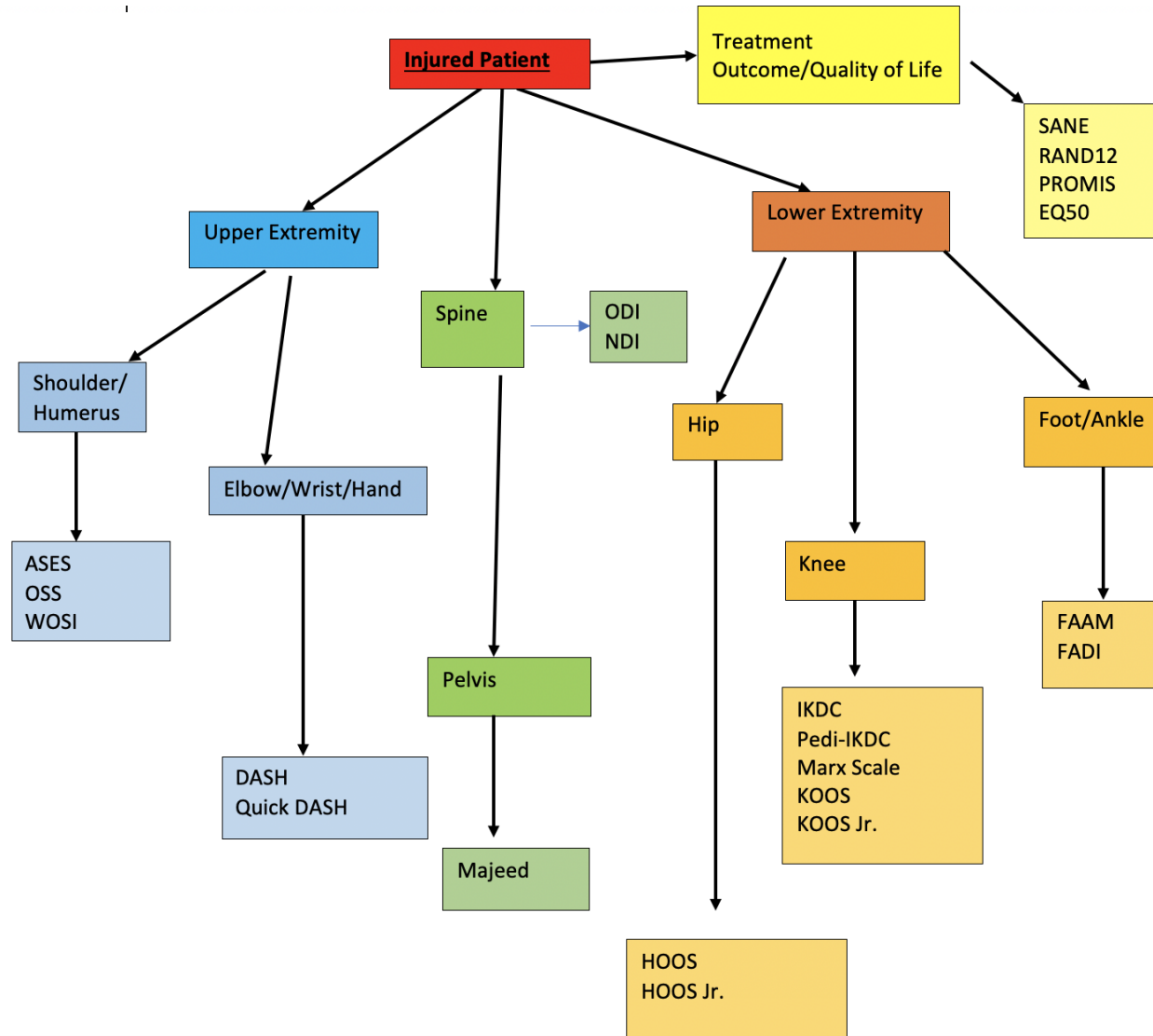
- AAOS has a list of PROMs
 - <https://www5.aaos.org/CustomTemplates/landingPage.aspx?id=4294968282&ssopc=1>
- System-specific Outcome measures
 - Oxford Shoulder Score (OSS)
 - American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form (ASES)
 - Foot and Ankle Ability Measure (FAAM)
 - Disabilities of the Arm, Shoulder and Hand Score (DASH)

Patient Reported Outcome Measurements

Examples

- **Disease-specific Outcome measures**
 - Knee Injury and Osteoarthritis Outcome Scores (KOOS)
 - Hip Disability and Osteoarthritis Survey (HOOS)
 - May have excessive variability between measures
- **General Quality of Life Measure**
 - Veterans RAND12
 - PROMIS
 - EQ-5D most common used internationally

AAOS Algorithm for PROMs



Validation of an Outcome Instrument

- **Good outcome measure should be Valid + Reliable + Responsive**
 - **Valid = does it measure what it intends to measure?**
 - **Reliable = does it measure repeatedly in the same way?**
 - **Responsive = does it show change when the patient's status changes?**

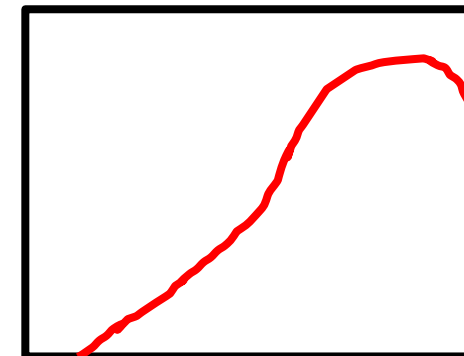
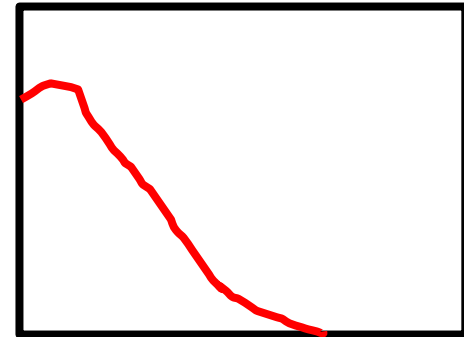
Outcome Measure: Validity

Typically has 3 components

- Content
- Construct
- Criterion

Also 2 problems: the floor and ceiling effect

- A floor effect occurs when a measure possesses a distinct lower limit for potential responses and a large concentration of participants score at or near this limit
- A ceiling effect refers to the point at which an independent variable no longer has an effect on a dependent variable, when a kind of saturation has been reached



Content Validity

How thoroughly does the measure test the domain of interest?

- Often use panel of experts
- Analyze “floor” and “ceiling” effects where differentiation at the ends of the spectrum are problematic
- Example: Musculoskeletal Functional Assessment (MFA)
 - 100-item self reported instrument developed in 90’s
 - “Content validity also was demonstrated, based on a review of item selection procedures, expert opinion, and the distribution of scores on the instrument”

Development of a musculoskeletal extremity health status instrument: The musculoskeletal function assessment instrument

Diane P. Martin, Ruth Engelberg, Julie Agel, Deborah Snapp, Marc F. Swiontkowski

Journal of Orthopaedic Research | Volume 14, Issue 2

First published: 01 March 1996



Construct Validity

- How well does the measure quantify some unobservable construct or hypothesis?
- Examples:
 - MFA – Hypotheses that patients with worse clinical presentation would score higher on MFA (correlations and ANOVA reported)
 - Convergent and discriminant validity against other health status measures were assessed (i.e.: SF-36, WOMAC)

Development of a musculoskeletal extremity health status instrument: The musculoskeletal function assessment instrument

Diane P. Martin, Ruth Engelberg, Julie Agel, Deborah Snapp, Marc F. Swiontkowski

Journal of Orthopaedic Research | Volume 14, Issue 2

First published: 01 March 1996



Criterion Validity

- **How well does the instrument correlate with a gold standard or well established measure of the characteristic?**
- **Example: MFA – Instrument was tested against physicians' ratings (11-point scale of dysfunction) and clinical measures (grip strength, walking speed, etc.). Adequate correlations were reported (Spearman' s rho >0.4 and p<0.001)**

Development of a musculoskeletal extremity health status instrument: The musculoskeletal function assessment instrument

Diane P. Martin, Ruth Engelberg, Julie Agel, Deborah Snapp, Marc F. Swiontkowski

Journal of Orthopaedic Research | Volume 14, Issue 2

First published: 01 March 1996



Reliability

- **Is the measure consistent or stable across time, patients or observers?**
- **Three main types of reliability testing:**
 - **Internal consistency**
 - **Test-retest reliability**
 - **Interobserver reliability**

Reliability: Internal Consistency

- **Extent that items in a scale are intercorrelated, measuring the same idea**
- **Instruments may ask questions phrased differently, but aimed at the same idea to truly capture the entirety of the respondent's outcome**
 - **Responses to these should correlate with each other**
 - **Cronbach's Alpha- a statistical measure of this correlation**
 - **Low - poor correlation**
 - **High - may indicate redundancy of items**
 - **Marginal Reliability- test often used in Item Response Theory instruments (discussed later)**

Reliability: Test- Retest

- Does the same test given at different time points yield similar results?
- Example: MFA
 - A sample of patients repeated the same instrument 5-8 days after initial administration. Percentage agreement, Spearman's rho and intra-class correlation were calculated

Reliability: Interobserver reliability

- Ability of instrument to produce the same result when assessed by different observers i.e. do the observers agree?
- Typically use kappa statistics

Responsiveness

- **Ability to detect clinically significant changes i.e. “sensitivity to change”**

Example: Putting it Together

Longitudinal Validation of the PROMIS Physical Function Item Bank in Upper Extremity Trauma

Aaron J. Kaat, PhD, Nan E. Rothrock, PhD,* Mark S. Vrahas, MD,† Robert V. O'Toole, MD,‡
Sarah K. Buono, BA,* Timothy Zerhusen, Jr, BA,‡ and Richard C. Gershon, PhD**

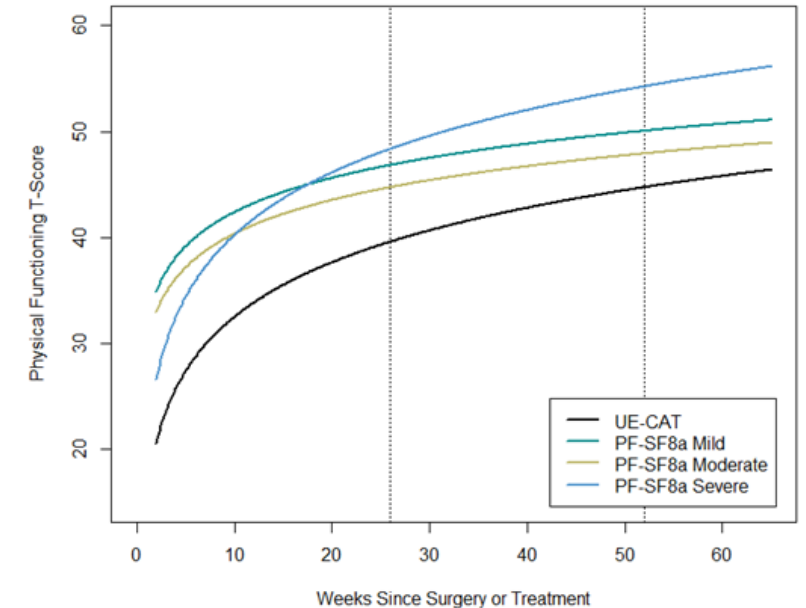
“Objective: To evaluate the reliability, validity, and responsiveness of the Patient-Reported Outcomes Measurement Information System (PROMIS) Upper Extremity Computer Adaptive Test (UE-CAT) and the 8-item Physical Function short form (PF-SF8a) for monitoring outcomes after musculoskeletal injuries in upper extremity trauma patients.”

J Orthop Trauma 2017; 31: e321-e326

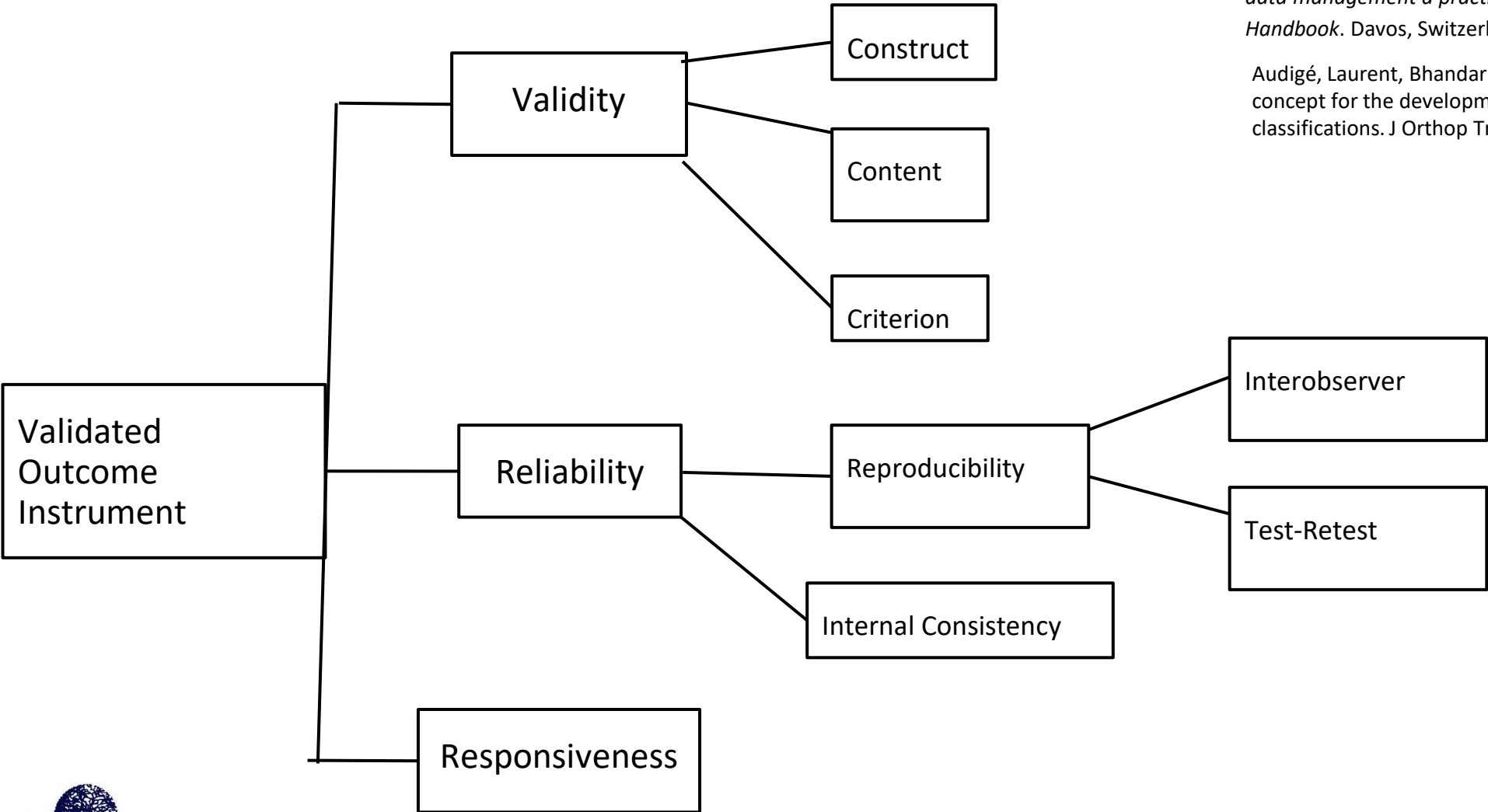
Example: Putting it Together

- **Brief Study Protocol: enrolled 424 patients at 2 Level-1 trauma centers**
 - **Administered PROMIS measures and “legacy” measures at 2 time points**
 - **Time 1= within 1 yr of Tx / Time 2 = 6 months after 1st assessment**
 - **Results:**
 - **Reliable:**
 - **Internal Consistency:**
 - **Marginal reliability: UE-CAT = 0.9, PF-SF8a=0.96**
 - **Cronbach’s alpha: SMFA = 0.96, Quick-DASH = 0.93**
 - **Validity: PROMIS Convergent with legacy measures**
 - **Responsive: Effect Sizes (ES) and Standard Response Means (SRM) comparing time 1 to 2 indicate response to improvement over time**

Figure, Supplemental Digital Content 2
PROMIS Longitudinal Trajectories



Summary of Validation



Bhandari, M., Hanson, B. P., & Stengel, D. (2009). pg 63. *Statistics and data management a practical guide for orthopaedic surgeons: Handbook*. Davos, Switzerland: AO Publishing.

Audigé, Laurent, Bhandari, M., Hanson, Beate, Kellam, James. A concept for the development and validation of fracture classifications. *J Orthop Trauma* 19: 404 – 409; 2005



Outcome Instruments: Other Considerations

- **Psychometric properties are important, but an instrument must also be...**
 - **User-Friendly for the patient**
 - **User-Friendly for the assessor**
 - **Cost-effective**
 - **Interpretable**
 - **Comparable across studies**
- **Outcome instruments have been shown to be excellent clinical tools, in addition to for research purposes**
 - **Time, administrative burden, EMR integration**

Administration and Collection of Data

- **Multidisciplinary team utilized to improve patient outcomes**
 - **Surgeon**
 - **Nursing practitioners**
 - **Dietician**
 - **Pharmacist**
 - **Social Workers**
 - **Physical Therapists**
 - **Pain Management**
 - **Alternative practitioners**

Administration and Collection of Data

- **Barriers to collection of patient reported outcomes include:**
 - **Time**
 - **Effort**
 - **Resources**
 - **Patient Factors**

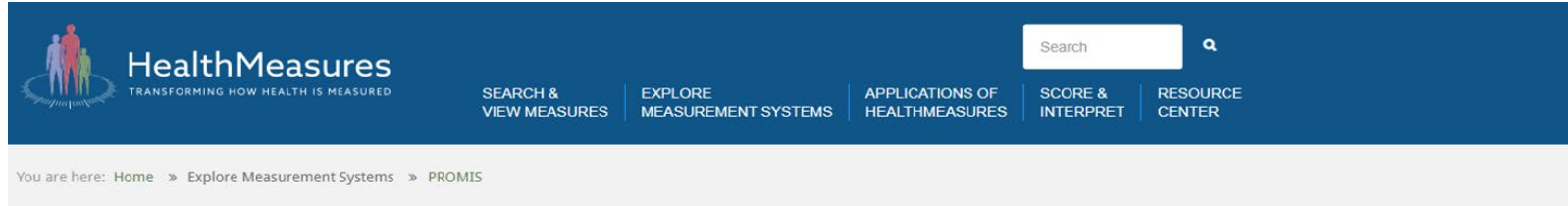
Administration and Collection of Data

- **Pencil and Paper vs Electronic**
 - Computer/Tablet/Cell Phone
 - Cost
 - Time to physicians and patients
- **Given prior to visit**
 - Saves time
 - Allows physician to anticipate any patient issues prior to visit
- **Given during the visit**
 - Easier to implement

PROMIS: Patient Reported Outcome Measures Information System

- **Started from 2004 NIH initiative**
 - **Goal to give clinicians and researchers a standardized set of PROMs**
 - **Design to improve reporting of patient symptoms, function, HR-QOL**
 - **Publicly available, precise, efficient, flexible**
 - **T-Score output:**
 - **Normalized to general population w/ mean score = 50 and SD = 10**
 - **Easier to interpret**

PROMIS



Intro to PROMIS®

PROMIS includes over 300 measures of physical, mental, and social health for use with the general population and with individuals living with chronic conditions.

- Overview
- PROMIS
 - **Intro to PROMIS**
 - List of Adult Measures
 - List of Pediatric Measures
 - Available Translations
 - Obtain & Administer Measures
 - Measure Development & Research
- Neuro-QoL
- ASCQ-Me
- NIH Toolbox

Adult Assessment

- PROMIS developed self-report measures for adults for functions, symptoms, behaviors, and feelings. See the [List of Adult Measures>>](#)
- Measures are available in Spanish and many other languages. See the [List of Available Translations>>](#)
- PROMIS measures are available for free and can be administered in multiple ways. Learn how to [obtain and administer PROMIS measures>>](#)



PROMIS® Adult Self-Reported Health

PROMIS Profile Domains

Physical Health

Fatigue
Pain Intensity
Pain Interference
Physical Function
Sleep Disturbance

Dyspnea
Gastrointestinal Symptoms

Mental Health

Anxiety
Depression

Alcohol
Anger
Cognitive Function

Global Health

Social Health

Ability to Participate in Social Roles & Activities

Companionship
Satisfaction with Social Roles & Activities



<https://www.healthmeasures.net/explore-measurement-systems/promis>

PROMIS

- **Significantly decreases time burden for patients**
- **Uses Item-Response Theory (IRT) and Computer Adaptive Testing (CAT)**
 - **Selects subsequent questions based on previous answers**
 - **Eliminates unrelated questions**
 - **Improves precision with few questions**
- **Found to be as accurate and precise as SMFA**
- **Significantly quicker than SMFA (44seconds vs 599seconds!!!)**
- **CAT and IRT may be pivotal developments in outcomes measurements**

Orthopaedic Advances

Item Response Theory and
Computerized Adaptive Testing
for Orthopaedic Outcomes
Measures

Brodke, Dane Jensen BA; Hung, Man PhD, MEd, MStat; Bozic, Kevin J. MD, MBA
Journal of the American Academy of Orthopaedic Surgeons: Volume 24 -
Issue 11 - p 750-754



PROMIS Reference Scoring

Establishing “Normal” Patient-Reported Outcomes Measurement Information System Physical Function and Pain Interference Scores

A True Reference Score According to Adults Free of Joint Pain and Disability

Sreten Franovic, MS, BS, Caleb M. Gulledge, BS, Noah A. Kuhlmann, MS, BS, Tyler H. Williford, MS, BS, Chaoyang Chen, MD, PhD, and Eric C. Makhni, MD, MBA

- **PROMIS score uses 50 as a mean score for the population, ie a reference however, this may not always be true.**
- **For individuals <40 years old PROMIS-PF and UE scores were >50 and were <50 for PROMIS-PI.**
- **For individuals >40, the reference range was accurate.**
- **Important to better understand efficacy of treatment.**

JB JS Open Access. 2019 Oct-Dec; 4(4): e0019.

Published online 2019 Dec 10. doi: 10.2106/JBJS.OA.19.00019



PROMIS vs Other PROMs

PROMIS Physical Function Computer Adaptive Test Compared With Other Upper Extremity Outcome Measures in the Evaluation of Proximal Humerus Fractures in Patients Older Than 60 Years

Jordan H. Morgan, BS, Michael A. Kallen, PhD, MPH,† Kanu Okike, MD, MPH,‡
Olivia C. Lee, MD,§ and Mark S. Vrahas, MD**

- **Found to have similar results with shorter assessment time compared to DASH and SMFA for upper extremity PF**

Journal of Orthopaedic Trauma: June 2015 - Volume 29 - Issue 6 - p 257-263

PROMIS

Floor and Ceiling Effects, Time to Completion, and Question Burden of PROMIS CAT Domains Among Shoulder and Knee Patients Undergoing Nonoperative and Operative Treatment

Caleb M. Gullette, BS, D. Grace Smith, BS, Alexander Ziedas, BS, Stephanie J. Muh, MD, Vasilios Moutzouros, MD, and Eric C. Makhni, MD, MBA

- **PROMIS shown to have favorable floor and ceiling effects for physical function (PF) and Upper Extremity (UE) questionnaires.**
- **Pain interference questionnaire may have a moderate floor effect but this may be improved in updated versions**
- **Depression questionnaire was found to have moderate to significant floor effects**
- **Conclusion: PROMIS UE and PF is has been shown to be reliable in certain setting while more needs to be learned before widespread utilization for PI and Depression questionnaires in orthopaedic practices**

Effect of Insurance on PROMs

Research Article

The Effect of Health Insurance Coverage on Orthopaedic Patient-reported Outcome Measures

- Retrospective review of over 10,000 foot and ankle patients with varying insurances
- PROM assessed
 - FAAM
 - PROMIS Global Physical
 - PROMIS Global Mental
 - PROMIS PF
- Scores compared with insurances
- Type of insurance was found to affect PROMs.
 - PROMS significantly reduced in workers compensation/motor vehicle insurance patients and Medicaid patients compared to those with commercial insurance carries
- In orthopaedic trauma, we see a large volume of workers compensation/motor vehicle insurance and Medicaid which may correlate with poorer reported PROMS.

Big Picture

- **PROMIS gives general health info but may be affected by multiple disease states or physiologic conditions**
- **Therefore, best to pair general health and disease or anatomy specific measures together to get fuller picture of outcomes**

Value-based Outcomes

- **Value=Outcome/Cost(dollar spent)**
- **Payers using outcome measures to assess value.**
 - **Quality of life measures and PROMs utilized**
 - **PROMIS, SF-36**
- **Institutions incentivized to provide higher value care by improving outcome**
 - **Institutions assessed by submitting outcome scores**
- **Due to complexity and heterogeneity of injuries, a generalized health measure like PROMIS may be best suited for this**



Value-based Outcomes

- **Centers for Medicare and Medicaid Services (CMS) monitors quality and efficacy of care, and financially incentivizes performance.**
- **CMS also governs a Merit-based Incentive Payment System under the Quality Payment Program**
 - **Pay adjustment based on quality of care**
- **Pay for Performance being adopted by private insurers as well.**

Bundled Payments

- **Push for improved value has led to adoption of bundled payments**
- **2016 CMS announced bundled payment model for hip and femur fractures**
- **CMS also created voluntary episode payment model to evaluate bundled payment model over 30 episodes of care**

Limitations/Future Directions in Trauma

- **Unable to get accurate baseline pre-injury assessment**
 - Patient reported preoperative status shown to be strongly predictive of outcomes after surgery in several conditions in orthopaedics
 - Lumbar stenosis, Carpal Tunnel, Hip and knee OA
 - New data suggesting benefit in use of proxies
 - Self-reported preinjury functional also may be beneficial
- **Use of PROs to get clinically meaningful assessment**
 - Establishing a minimum clinically important difference
- **More trauma specific outcomes PROs may be helpful especially in current value-based environment**
- **Predictive modeling with the use of large registries**

Conclusion

- **Outcome measures are how we as surgeons can assess the efficacy of our treatment.**
- **Many factors associated with outcomes in orthopaedic trauma patients**
- **Paucity of orthopaedic trauma specific measures**
- **General measures are becoming more useful and user friendly**
- **Outcome measures are becoming increasingly important in a value-based health care system and may drive reimbursement**