**OBJECTIVES:** Hemophilia is a chronic disease typically diagnosed in infancy, characterized by bleeding requiring regular or episodic infusion of clotting factor. Taking care of a child with hemophilia (CWH) may cause burden for caregivers. We aimed to develop a first "Hemophilia associated Caregiver Burden Scale" (HEMOCABTM) assessing the burden of hemophilia for caregivers of CWH. **METHODS:** Questionnaire development included: 1) item generation (two semi-structured focus groups with 11 caregivers, evaluation of existing caregiver burden scales for relevance by 16 HCPs); 2) feasibility testing (cognitive interviews with 12 caregivers) and 3) pilot-testing (psychometric analysis of HEMOCABTM). **RESULTS:** Item generation resulted in a preliminary version of HEMOCABTM consisting of 109 items. During feasibility testing, mean completion time was 19.3±6.3 minutes, some problems were revealed concerning item formulation, applicability of items for young children and missing items. The revised HEMOCABTM contained 108 questions grouped in 13 domains. HEMOCABTM was pilot-tested in 40 caregivers of CWH with a mean age of 39.32±8.9. The majority of CWH had hemophilia A (95%), were severely affected by hemophilia (77.5%) and 15% had inhibitors. Reliability estimation showed high internal consistency of total score with Cronbach's α=.97, and for 2 summary scores 'FREQUENCY' with α=.95 and 'BURDEN' α=.92; Cronbach's α for the sub domains ranged from α=.77 to .93. HEMOCABTM revealed good convergent validity with Impact on Family Scale (r=-.867 for total score). Known groups validity showed significant differences in all domains of HEMOCABTM, except for 'school' among caregivers of CWH with inhibitors vs. without. Type of treatment and disease severity showed some differences between groups. Based on item and scale analysis 49 items were deleted and the final HEMOCABTM consists of 59 items. **CONCLUSIONS:** HEMOCABTM is the first hemophilia-specific instrument for the assessment of caregiver burden and revealed good psychometric characteristics in terms of reliability and validity.

PRM81

**USING THE CLINICAL SUMMARY SCORE FROM THE KANSAS CITY CARDIOMYOPATHY QUESTIONNAIRE AS AN ENDPOINT IN CLINICAL TRIALS: PSYCHOMETRIC SUPPORT**

Gwaltney C[1], Tiplady B[1], Deschaseaux C[2]
[1]*ERT, Pittsburgh, PA, USA,* [2]*Novartis Pharma AG, Basel, Switzerland*

**OBJECTIVES:** Symptoms and physical limitations can have an important impact on the day to day lives of heart failure patients. The Clinical Summary Score (CSS) of the Kansas City Cardiomyopathy Questionnaire (KCCQ), a patient-reported outcome instrument, provides a measure of symptoms and physical limitations associated with heart failure. The primary goal of this study was to evaluate the psychometric properties of the CSS and its utility as an endpoint in clinical trials. **METHODS:** Data from 3 randomized, controlled clinical trials with heart failure patients were included in analysis (ALOFT, PARAMOUNT, and PARADIGM trials). Studies were examined independently; within each study, data were collapsed across treatment groups. Study measures included the KCCQ, physician-rated New York Heart Association (NYHA) classification, patient global impression of change (PGIC), and/or NT-proBNP assay. **RESULTS:** Findings were similar across the 3 trials. Mean CSS scores at baseline ranged from 63-76 (on a 0-100 scale, with higher scores indicating better symptoms and physical functioning). Dimensionality assessments highlighted the complex nature of the scale, with evidence for first-, second-, and third-order factors. The CSS consistently discriminated between all four NYHA classifications (all pairwise comparisons p<.05). Correlations with BNP and NT-proBNP levels were statistically significant, but relatively small (-.12 and -.17 respectively). The CSS was sensitive to changes in patient status over time, as indexed by changes in the NYHA classification and PGIC. The responder definition – the amount of change within an individual patient that would be considered clinically meaningful - was in the range of 6 to 10 points, which is higher than what has been seen in studies of the KCCQ overall score. **CONCLUSIONS:** Data from multiple heart failure clinical trials confirm the psychometric characteristics of the CSS. This evidence supports the use of the CSS as an endpoint in clinical trials examining heart failure treatments.

PRM82

**VALIDATION AND U.S. POPULATION NORMS OF HEALTH-RELATED PRODUCTIVITY QUESTIONNAIRE**

Tundia N[1], Hass S[2], Fuldeore M[3], Wang LL[4], Cavanaugh T[4], Boone J[4], Heaton P[4]
[1]*AbbVie, North Chicago, IL, USA,* [2]*Genzyme, Cambridge, IL, USA,* [3]*AbbVie Inc., North Chicago, IL, USA,* [4]*University of Cincinnati, Cincinnati, OH, USA*

**OBJECTIVES:** To validate the Health-Related Productivity Questionnaire (HRPQ), a new health-related productivity instrument, and estimate the US population norms by age and gender. **METHODS:** An online survey was developed that consisted of four components: the HRPQ; a screener to determine relevant disease conditions-related and health-state questions; validated instruments such as Work Productivity and Activity Impairment (WPAI) and EQ-5D; and socio-demographic questions. The survey was administered by a third-party company for a 6-week period. Weightings were calculated to allow extrapolation of results from the 10,000 respondents to achieve values representing the general United States population. Validation analysis included concurrent and criterion validity, construct validity with group difference, extreme group comparison and factor analysis. Mean, median, standard error, 25th and 75thpercentiles were calculated for absenteeism, presenteeism measures for employed and household work and stratified by age, gender, age and gender. **RESULTS:** The HRPQ showed strong concurrent validity with WPAI (Pearson's r≥0.6, p-value<0.05). Correlations of total productivity at work and home from HRPQ with EQ-5D scores, were small to moderate, r=0.3-0.5 (p-value<0.05) and aligned with direction of the hypothesis. Several group difference analyses showed positive results for HRPQ. Presenteeism items heavily loaded on one factor and absenteeism and scheduled hours items loaded on second factor. General population estimates for average percent lost productivity at work was 14%, (absenteeism=4%, presenteeism=10%) and for household activities was 28% (absenteeism=18%, presenteeism=10%). Furthermore, 17% of people reported any kind of loss of workforce participation due to illnesses/treatments. Average work loss productivity was significantly higher in females vs males (16% vs 13%, p<0.05) and decreased with increase in age (20%, 17%, 14%, 11% and 7% in 18-29, 30-39, 40-49, 50-64 and >64 years old, respectively, p-value <0.05). **CONCLUSIONS:** HRPQ has good construct and criterion validity. Presenteeism remains higher for paid work, while absenteeism remains higher for unpaid work.

PRM83

**MEASURING UPPER LIMB FUNCTION IN MULTIPLE SCLEROSIS: ENHANCING THE ABILHAND'S PERFORMANCE**

Cano S[1], Cleanthous S[2], Marquis P[1], Hobart J[3], Naoshy S[4], Mikol D[4], Petrillo J[4], Steiner D[4], Chen Y[4], Watson C[4]
[1]*Modus Outcomes LLC, Newton, MA, USA,* [2]*Modus Outcomes LLC, London, UK,* [3]*Plymouth University, Plymouth, Devon, UK,* [4]*Biogen Idec, Cambridge, MA, USA*

**OBJECTIVES:** ASCEND is a phase 3, randomized, double-blind, placebo-controlled trial assessing whether natalizumab slows disability progression in secondary progressive multiple sclerosis (SPMS) patients. The aim of this current analysis was to use Rasch Measurement Theory (RMT) methods to evaluate the ABILHAND (56 item version), a PRO assessing manual ability, in SPMS patients, and to explore an optimized scoring structure based on empirical post-hoc analyses. **METHODS:** Baseline blinded data from the 889 randomized patients in ASCEND were analyzed. In stage 1, RMT methods examined: scale-to-sample targeting, item fit, local dependency, and reliability. In stage 2, a post-hoc revision of the ABILHAND-56 scoring structure and conceptual grouping of items was conducted and re-evaluated using the same RMT methods. **RESULTS:** Stage 1 analyses showed adequate scale performance: minor item misfit (2/56); minimal dependency (4 pairs of items); good reliability (Person Separation Index = 0.93). However, there was ABILHAND-to-ASCEND sample mis-targeting (person location range [mean]: -7.31 to 1.83 [-3.62]; item location range [mean]: -5.75 to 4.27 [0.00]). In stage 2, all items were rescored on a dichotomous response scale (easy & difficult/impossible) in an attempt to improve targeting. Also, the 56 items were re-categorized into two conceptually clearer manual ability sub-scales: 'fine motor skills' and 'power'. These ABILHAND-56 revisions improved targeting: 'fine motor skills' (person range [mean]: -5.39 to 5.34 [-1.75]; item range [mean]: -4.15 to 3.54 [0.00]), 'power' (person range [mean]: -4.47 to 4.38 [-1.72]; item range [mean]: -4.15 to 3.54 [0.00]). Sample measurement and item fit were consistent with the original ABILHAND-56. **CONCLUSIONS:** The ABILHAND-56 revised scoring demonstrated improved psychometric performance and provides an initial evidence-base for the enhancement of ABILHAND-56's measurement performance in people with SPMS. Additional research will determine whether the two sub-scale structure of the revised ABILHAND-56 provides better interpretability of patient-reported manual ability.

PRM84

**VALIDATION OF THE DEPRESSION AND FAMILY FUNCTIONING SCALE (DFFS)**

Francois C[1], Danchenko N[2], Williams V[3], Nelsen L[3], Williams NJ[3], Yarr S[4], DiBenedetti DB[3], Lancon C[5]
[1]*Lundbeck LLC, Deerfield, IL, USA,* [2]*Lundbeck S.A.S. Paris Fr, EU, Paris, France,* [3]*RTI Health Solutions, Research Triangle Park, NC, USA,* [4]*The Pavilion, Manchester, UK,* [5]*Hospital Sainte-Marguerite, Marseille, France*

**OBJECTIVES:** Patient-reported outcomes (PROs) are necessary to assess disease impacts from the patient's perspective. In line with the Food and Drug Administration's (FDA's) guidance on PROs, the Depression and Family Functioning Scale (DFFS) was developed to assess the impact of major depressive disorder (MDD) on family functioning. Psychometric analyses were conducted to establish the reliability, validity, and responsiveness of the DFFS according to the FDA PRO guidance. **METHODS:** Data from PERFORM, a longitudinal multicenter, prospective, 2-year observational study in the United Kingdom and Spain, were analyzed (NBaseline=478; NMonth2=433). The 15 DFFS items use a 5-point rating scale to assess partner and family interactions and quality of relationships; higher scores indicate greater (worse) impacts. Test-retest reliability (intraclass correlations), construct validity (correlations and factor analysis), discriminating ability (analyses of variance), and responsiveness (effect size estimates) were evaluated. **RESULTS:** Factor analyses resulted in a single factor, confirmed by highly satisfactory Cronbach's alphas (0.85 at baseline, 0.89 at month 2). The DFFS demonstrated satisfactory test-retest reliability (intraclass correlation=0.75). Hypothesized correlations with other measures provided evidence of convergent and divergent validity. For example, the correlation of the DFFS with SF-12 mental component scores was –0.35 (baseline) –0.49 (month 2), and with SF-12 physical component scores, –0.05 (baseline) and –0.31 (month 2). Hypothesis tests were generally in the predicted direction and many were statistically significant, substantiating the discriminating ability of the DFFS. Effect size estimates of responsiveness were 0.44–0.84, demonstrating that the items were capable of detecting change. **CONCLUSIONS:** The psychometric analyses strongly support the reliability, validity, and responsiveness of the DFFS and its usefulness for assessing the impacts of depression on family functioning. It has the potential to provide important information not traditionally captured in clinical practice or research and will facilitate a more comprehensive evaluation of treatments of MDD.

PRM85

**DOES DIFFERENTIAL FRAMING OF OPT-OUT ALTERNATIVES IN DISCRETE CHOICE EXPERIMENTS (DCES) MATTER? COMPARISON OF RANDOM UTILITY MAXIMIZATION (RUM) AND RANDOM REGRET MINIMIZATION (RRM) MODELS**

Chaugule S[1], Hay JW[1], Young G[2], Martin OA[3], Drabo EF[1]
[1]*University of Southern California, Los Angeles, CA, USA,* [2]*Children's Hospital Los Angeles, Los Angeles, CA, USA,* [3]*Barcelona GSE, Barcelona, Spain*

**OBJECTIVES:** We systematically investigate random utility maximization and random regret minimization modeling approaches to establish the impact of dif-