labeled “new drug,” “old drug,” or “no drug” - non-fatal stroke, non-fatal myocardial infarction (MI), cardiovascular death, minor bleeding, major bleeding, and death, and need for therapeutic monitoring. We estimated preference weights and maximum acceptable risks. RESULTS: A total of 341 patients and 352 individuals from the general population completed the questionnaire. On average, patients perceived a 1% increased risk of a fatal bleeding equivalent to a 1% increase in non-fatal MI, a 3% increase in non-fatal stroke, a 3% increase in cardiovascular death, a 6% increase in major bleeding, or a 16% increase in minor bleeding. As compared to the patients, the general population had similar preferences except that they perceived a 3% increase in non-fatal MI or a 13% increase in minor bleeding equivalent to a 1% increase in risk of bleeding death. Patients were less likely to choose “no drug” (odds ratio, 0.72; 95% confidence interval, 0.61-0.84) or “old drug” (odds ratio, 0.86; 95% confidence interval, 0.81-0.93) than “new drug.” The general population sample was not different from the drug labels. CONCLUSIONS: Patients and the general population had similar relative preferences for anticoagulant treatment outcomes but were more likely to choose “new drug,” irrespective of its relative benefits and risks.

PP4
MEASURING TREATMENT PREFERENCES OF PATIENTS DIAGNOSED WITH IDIOPATHIC PULMONARY FIBROSIS USING BEST-WORST SCALING
Camponeschi G1, Fitzgerald J2, Danoff SK1, Richardson B1
1Johns Hopkins University, Baltimore, MD, USA, 2Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA
OBJECTIVES: Idiopathic pulmonary fibrosis (IPF) is a rare, degenerative disease. While recently approved therapies provide hope, research is needed to assess the value of treatment benefits and risks. This study sought to develop and test a patient reported outcome (PRO) instrument to understand patient preferences for therapies.
METHODS: Using patient and stakeholder engagement, researchers developed a survey instrument for measuring the treatment preferences of IPF patients. This led to a best-worst scaling index to assess six treatment attributes, each defined across three levels, including lung function, shortness of breath, persistent cough, gastrointestinal problems, skin problems and risk of liver toxicity. The survey was completed in person at the annual IPF patient meeting profiles, created through a main-effect orthogonal experimental design, and asked to identify the best and worst aspect of each treatment. Preference weights were estimated using a simple score consisting of the number of times a level was chosen. The mean number of times it was chosen as worst and divided by the total number of times the level was shown. Conditional on the level chosen in the experiment, attribute importance was estimated by comparing the range of scores across each attribute, relative to all such deviations. RESULTS: Thirty-five participants completed the survey. The most important attribute preferred was effect on lung function (35%), followed by risk of gastrointestinal problems (23%), risk of liver toxicity (21%) and impact on persistent cough (11%). Patients estimated the level of liver toxicity to be risk of the measured attributes to be of equal importance (50%) and impact on shortness of breath (9%). CONCLUSIONS: This research demonstrates the merits of a community-centered approach to survey instrument development to measure preferences and illustrates the value in quantifying preferences. Further research is needed to assess the generalizability of these findings and the implications for decision making.

RESEARCH POSTER PRESENTATIONS - SESSION I
RESEARCH ON METHODS STUDIES
RESEARCH ON METHODS - Clinical Outcomes Methods
PRM1
DEVELOPMENT AND VALIDATION OF A U.S. ADMINISTRATIVE CLAIMS-BASED ALGORITHM TO CLASSIFY PATIENTS WITH TYPE 2 DIABETES MELLITUS INTO RENAL IMPAIRMENT STAGES
Johnson G1, Sheehan J2, Koehle E1, Cappelli KA1, Varker H1, Anzalone D1, Ghannam A1, Kaiser K1
1Tufts Health Analytics, Bethesda, MD, USA, 2AstraZeneca, Fort Washington, PA, USA, 3Tu Jenna Health Analytics, Ann Arbor, MI, USA, 4AstraZeneca, Wilmington, DE, USA
OBJECTIVES: The validity of diagnosis/procedure coding for determining the severity of renal impairment is unknown. This retrospective, observational study developed administrative claims-based algorithms, modified ROME III diagnostic criteria, and patient-reported physician diagnoses to identify diabetes mellitus (T2DM) into renal impairment stages using estimated glomerular filtration rate (eGFR). RESULTS: A total of 341 patients and 352 individuals completed a cross-sectional survey that included questions pertaining to IBS-C/CC symptoms based on modified ROME III criteria and patient self-report of IBS-C and CC physician diagnoses to confirm claims-based diagnoses. RESULTS: Among 236 claims-based IBS-C patients, 22% met ROME III IBS-C criteria and 43% reported being told by a physician they had IBS-C. In addition, 33% of claims-based IBS-C patients reported being told by a physician they had CC. Among 456 claims-based CC patients, 27% met ROME III CC criteria and 39% reported being told by a physician they had CC. However, 38% of claims-based CC patients met ROME III criteria for IBS-C and 18% reported being told they had IBS-C. Patients who did and did not meet ROME III criteria had similar demographic and clinical characteristics. CONCLUSIONS: A majority of patients identified as having IBS-C or CC met ROME III criteria. The lack of agreement between claims-based criteria and patient-reported physician diagnoses than ROME III criteria. Our findings suggest that patients identified through claims may have been asymptomatic at the time of the survey, and those identified as CC patients may be IBS-C patients who never received an IBS claim.
PRM3
PSYCHOMETRIC VALIDATION OF PERFORMANCE OUTCOMES (PERFOS) FOR USE WITH HIP FRACTURE (HF) POPULATIONS
Gentile B1, Ballinger BS1, DoHill A2, Bush EN1
1UCON, San Francisco, CA, USA, 2UCON, Oxford, UK, 3Eli Lilly, Indianapolis, IN, USA
OBJECTIVES: We evaluated the reliability and validity of three Perfos (Timed Up-and-Go (TUG), 4-Step Stair Climb (ASC) and Repeated Chair Stands (RCS, in two versions with arms folded, RCS-A, or arm rests, RCS-B) were evaluated in hip fracture (HF) surgery patients. METHODS: Patients were recruited from 15 clinical sites across 6 US states. Participants visited sites at designated time points after HF surgery when patient reported outcome (PRO) measures, patient- and clinician-reported global concept items (CCI), and Perfos were administered. Perfos were scored as time (seconds) to complete each test. Perfo measurement properties evaluated included: reliability (inter-rater, test-retest), construct validity (known groups, convergent/divergent), ability to detect change, minimal important difference (MID) and responder definitions (BCP). RESULTS: Data were collected from 75 patients (76 years; 68.0% female) at baseline; from 68 and 66 at visits 2 and 3. Inter-rater (CCCs: 0.87 to 0.97) and test-retest (CCCs: 0.91-0.95) reliability was excellent across the three Perfos. Known groups validity: Those without an assistant device had complete mean times for all Perfos but the RCS-A. In addition, TUG times were shorter for patients with high versus low SF-12 physical component summary (PCS) scores (p = .009). Convergent/divergent validity: the TUG, RCS-B, and 4SC demonstrated moderate correlations with SF-12 PCS (r ranged -0.227 to -0.449), and stronger correlations with the individual physical dimensions than the mental component (MCS) and dimension scores. Ability to detect change: patients demonstrated significant changes in Perfos from baseline to Visit 2 for the RCS-B (p = 0.030) and 4SC (p = 0.034). MDIs ranging from 1.5s (4SC) to 6s (TUG) were found. Based on Best Cut Points (BCP) of one-point change in clinician GCIs and values of minimal detectable change (MDC90), responder definitions between 2.0s (4SC) and 3.5s (TUG) are recommended. CONCLUSIONS: Overall, the three Perfos demonstrated adequate psychometric properties.
PRM4
GETTING THE FULL PICTURE: THE IMPORTANCE OF EXTRAPOLATING BEYOND THE DATA
Al-Majed A1, Rittenhouse B2
1MCPRU University, Boston, MA, USA
OBJECTIVES: Clinical assessments with a limited time horizon for data collection are typical. One example is a (now dated) 10-year study of patient therapy by Pharrak & Hollander (2011). This study suggests we underestimate the effect of the intervention by limiting its time horizon. We explore this suggestion as a methodological point. METHODS: Using life table methods we simulated rounds of follow up of patients that could result from two treatments (X and Y). Assuming a fully persistent treatment effect over lifetimes, this increased to 2.35 in the extrapolation, clearly differently differ-