in type 2 diabetes mellitus. Comprehension of the checklist was assessed based on qualitative interviews of each analyst. A measure of the inter-analysts agreement was estimated to ensure the reliability of the checklist. **RESULTS:** Checklists identified from the literature included the checklist developed by the NICE Decision Support Unit and the one developed by the ISPOR task force. These two checklists were developed for analysts who conduct NMAs as well as analysts who critically review NMAs. However, they seem to lack clarity for non-statisticians. We developed a new checklist, which included the following items: definition of the study question (list of comparators, study population), methods (study selection, data extraction, statistical model, selection of fixed versus random effects model, assumptions for the base case, heterogeneity and inconsistency assessment and sensitivity analyses), reporting of results (network and source data, median or mean and 95% credibility interval) and interpretation of results. **CONCLUSIONS:** Our checklist can be used by analysts not trained in statistics to prepare or review NMAs to be submitted to NICE and/or to populate cost-effectiveness models.

PRM208

METHODOLOGY FOR SELECTING EXPERT GROUPS FOR THE PURPOSE OF DECISION-MAKING TASKS

Ivlev I[1], Bartak M[2], Kneppo P[1]

[1]*Czech Technical University in Prague, Kladno, Czech Republic,* [2]*Jan Evangelista Purkyne University, Usti nad Labem, Czech Republic*

OBJECTIVES: This work aims to develop a methodology for determining the qualitative composition of an expert group for the purpose of participation in decision-making in health care technology. Its goal is also to evaluate the methodology based on an example of the selection of large medical equipment. **METHODS:** The complex weighting factor is a comprehensive evaluation of an expert. It is based on the expert's overall work experience, experience in solving tasks, level of education and scientific record, interest in solving the particular task, current position, and awareness of how to solve the task. Also taken into account are the relevance of the expert's knowledge and the overall self-evaluation concerning his or her total competence in solving the task. For the purpose of validating the methodology, 96 potential experts were interviewed. These subjects included managers from relevant departments in hospitals and hospital staff members who were from 72 health facilities in the Czech Republic. **RESULTS:** Unlike the other models, the calculation model that was selected is able to eliminate errors in estimating the proportionality of extreme values and to reduce the impact of uncertainty in the experts' overall self-evaluations concerning their total competence to the combined ratio. Based on this model, a methodology for selecting experts was developed. A statistically significant correlation was found between the complex weighting factor and the following characteristics: the expert's experience in dealing with similar tasks ($r=0.512$, $p<0.001$), the expert's theoretical background (awareness) and the relevance of the expert's knowledge ($r=0.44$, $p<0.001$), the expert´s current position ($r=0.319$, $p=0.002$), and the level of his or her education and scientific record ($r=0.28$, $p=0.007$). **CONCLUSIONS:** This methodology will be especially useful in scientific and technological forecasting, medical and managerial decision-making, quality assessment, and operational research.

PRM210

MODELLING LONGITUDINAL TRAJECTORIES OF PATIENT-REPORTED OUTCOMES TO EVALUATE TREATMENT EFFECT

Jen MH

Eli Lilly and Company, Surrey, UK

OBJECTIVES: To evaluate treatment effect on longitudinal patient-reported outcomes using appropriate analytical strategy. **METHODS:** This was an ad-hoc analysis of longitudinal patient-reported outcomes using a two stages simulated data in which the true model is known, to explore and to evaluate the capability of the group-based trajectory method to identify the distinctive features of a highly irregular but still continuous population distribution of trajectories. Firstly, we created six different types of underlying trajectory in which the true model is known and added in level-one between occasion random noises. Then we added a level-two, between- patients variation (of the random intercept form) with differing variability to each of the six distinctive trends. This simulation allows us to examine how the software implementation identifies different group trajectories as well as their level-one and -two variances. It was recognized that a priori assignment of distinct longitudinal trajectories may not be appropriate and that no ability to calibrate the precision of individual classifications exists if ex-ante rules are used. Thus, latent group-based trajectory model, a method to map the developmental course of symptoms and assess heterogeneity in response to clinical interventions, was used to identify patient groups with varied response. **RESULTS:** The fitted trajectories closely approximate the true shapes and there is also a close correspondence for the percentage of places attributed to each group. Even the size of the level 1 random term is correctly estimated. The semi-parametric group-based trajectory method has demonstrated unequivocally its capability to capture the unobserved subgroups in the presence of considerable level-1 random variation. **CONCLUSIONS:** Patients in many disease areas experience changes in QoL in different ways. Identification of those groups is essential for appropriate evaluation of therapy treatment effects and identification of factors contributing to those groupings.

PRM211

THE RANDOMIZED BLIND START TRIAL: EVALUATION OF A NEW STUDY DESIGN FOR ASSESSING CLINICAL OUTCOMES IN RARE AND HETEROGENEOUS PATIENT POPULATIONS

Signorovitch J[1], Ayyagari R[2], Kakkis E[3]

[1]*Analysis Group, Inc., Boston, MA, USA,* [2]*Analysis Group Inc., Boston, MA, USA,* [3]*Ultragenyx Pharmaceutical Inc., Novato, CA, USA*

OBJECTIVES: Clinical development of therapies for rare diseases can benefit from improvements to conventional trial designs. This study evaluated a new trial design, the randomized Blind Start. **METHODS:** The Blind Start design randomizes patients to ≥ 3 groups which initiate double-blind active therapy at different times from baseline, preceded by 0,1, 2 or more intervals of placebo. Analytical and simulation-based investigations were conducted to compare the statistical power, required assumptions, practical considerations and economic features of the Blind Start to conventional single-arm and randomized designs. **RESULTS:** Given the same number of patients, the randomized Blind Start provides equivalent statistical power for detecting changes pre- vs. post-treatment compared to a conventional single-arm design. However, by concealing treatment initiation times, the Blind Start enables more objective assessments of outcomes that are effort-based, patient-reported or subjectively assessed by investigators. In addition, compared to a conventional 2-arm randomized trial, analysis of parallel treatment and placebo groups embedded within a Blind Start design provides greater power to detect treatment effects over any fixed time interval. For example, with N=16 and a treatment effect equal to 1 standard deviation of the outcome measure, a 4-arm Blind Start design provides 85% power in a pre-post analysis and 79% power in an analysis of embedded parallel groups. In contrast, a conventional 2-arm randomized trial provides 52% power in this scenario. Benefits of the Blind Start design come at the expense of 1) more patient-time in the trial and 2) lack of stringent control over patient status upon active treatment initiation. **CONCLUSIONS:** The randomized Blind Start design can improve precision for treatment effect estimation vs. parallel-group designs and reduce risk of bias vs. single-arm designs. Endpoint choice and statistical analysis strategies for the Blind Start design can maximize the assessment of treatment effects on multiple outcomes.

RESEARCH ON METHODS – Study Design

PRM212

THE QUALITY OF SEARCH METHODOLOGY AND SEARCH REPORTING IN PUBLISHED SYSTEMATIC REVIEWS OF ECONOMIC EVALUATIONS: SEARCH SOURCES

Wood H, Arber M, Glanville J

York Health Economics Consortium, York, UK

OBJECTIVES: The economic evaluation of health care interventions is now an accepted element of health care decision-making and priority-setting. As the number of published economic evaluations has grown, so has the number of systematic reviews of economic evaluations. However, the quality of search methodology used in recent reviews has not been widely investigated. This study sought to identify which search resources are being used to identify studies in recent, published systematic reviews of economic evaluations, and to investigate whether choice of resources reflects current recommendations for the conduct of such reviews. **METHODS:** A search to identify systematic reviews of economic evaluations published since January 2013 was undertaken in MEDLINE. Two reviewers extracted the following information from reviews which met the inclusion criteria: general medical literature databases searched, specialist economic databases searched, health technology assessment sources searched, supplementary search techniques used. Results were compared against the search resources recommended by NICE when searching for economic evidence for single technology appraisals, and the summary of current best evidence provided in Sure Info (http://vortal.htai.org/?q=node/336). **RESULTS:** Sixty-five systematic reviews met the inclusion criteria; 23 of these could not be accessed in full text, data was extracted from 42 reviews. Five reviews (12%) met or exceeded the search resources recommended by NICE (MEDLINE, Embase, NHS EED, EconLit). Nine reviews (21%) searched at least four of the six types of resource recommended by Sure Info (specialist economic databases, general databases, HTA databases, webpages of HTA agencies, grey literature, collections of utility studies). None of the reviews searched all six. Although all reviews explicitly described the resources searched, reporting frequently contained errors or lack of clarity in the names of databases and interfaces. **CONCLUSIONS:** The information resources used to identify evidence for the majority of recently published systematic reviews of economic evaluations do not conform to current recommendations.

PRM214

IDENTIFYING PSORIASIS AND PSORIATIC ARTHRITIS PATIENTS IN RETROSPECTIVE DATABASES WHEN DIAGNOSIS CODE IS NOT AVAILABLE: A VALIDATION STUDY COMPARING MEDICATION/PRESCRIBER VISIT BASED ALGORITHMS TO DIAGNOSIS CODES

Dobson-Belaire W[1], Borrelli R[1], Goodfield J[1], Liu FF[2], Khan ZM[3]

[1]*IMS Brogan, Mississauga, ON, Canada,* [2]*Celgene Inc., Mississauga, ON, Canada,* [3]*Celgene Corporation, Summit, NJ, USA*

OBJECTIVES: Retrospective database studies rely on the ability to accurately identify patient cohorts of interest within health care databases. Diagnosis code-based algorithms are the primary method of identifying patient cohorts; however, many databases lack reliable diagnosis code information. Our aim was to develop precise algorithms based on medication claims/prescriber visit (MC/PV) to identify psoriasis (PsO) patients or psoriatic patients with arthritic conditions (PsO-AC), a proxy for psoriatic arthritis, in databases lacking diagnosis codes. **METHODS:** Algorithms were developed using medications with narrow indication profiles in combination with prescriber specialty to define PsO and PsO-AC. For the study period of July 1, 2009 to June 30, 2013, algorithms were validated using the PharMetrics Plus™ (PharMetrics) database, which contains both adjudicated medication claims and diagnosis codes. Positive predictive value (PPV), negative predictive value (NPV), sensitivity and specificity of algorithms developed for PsO and PsO-AC were assessed using diagnosis code as the reference standard. **RESULTS:** In the PharMetrics database, 183,328 patients were identified by diagnosis code or medication claim for validation. The highest PPVs for PsO (85%) and PsO-AC (65%) occurred when a predictive algorithm of ≥2 MC/PVs was compared to the reference standard of ≥1 diagnosis code. The majority of PsO-AC false positives had a diagnosis of PsO and pain or joint symptoms. NPV and specificity were also high (99 – 100%), while sensitivity was low (≤30%). Reducing the number of MC/PVs or increasing diagnosis claims decreased