disease areas and therapeutic classes, but treatments for musculoskeletal and rheumatic disease are evaluated most frequently.

PRM190

EXTENSION OF THE HOSMER-LEMESHOW GOODNESS OF FIT STATISTIC TO LINEAR MODELS WITH REPEATED MEASUREMENTS

Li R[1], Su Z[2], Mendelsohn A[2], Gemmen E[1]

[1]Quintiles Outcome, Rockville, MD, USA, [2]Quintiles Outcome, Cambridge, MA, USA

OBJECTIVES: Studies of health outcomes commonly involve binary measures, assessed multiple times. Although generalized linear mixed (GLIMMIX) models are well suited for analyzing these data, there does not exist a formal statistic to assess the goodness of fit (GOF) for GLIMMIX models. We developed an extension of the Hosmer-Lemeshow GOF test used for logistic regression that can be applied to GLIMMIX models. **METHODS:** The correlation among repeated measurements of the binary outcome variable was accounted for by a random effect in the GLIMMIX model. The principles of Hosmer-Lemeshow method were followed. The linear unbiased estimate of dependent variables were transformed to the original probability, sorted from least to largest, and divided into deciles. A Chi-square statistic and corresponding p-value with eight degrees of freedom, was calculated based upon the expected and observed numbers among deciles. The proposed GOF test was validated by a simulation study with 1000 runs generated from logistic regression models with and without random effects. The results were compared with the Hosmer-Lemeshow GOF test in situations where the latter is appropriate. The proposed method was used in the analysis of a comparative effectiveness (CE) study of ophthalmologic treatments for open-angle glaucoma patients. **RESULTS:** When there was no random effect, the proposed GOF test results from the GLIMMIX procedure were almost identical to those of Hosmer-Lemeshow GOF test from the logistic procedure. With a random effect built in a correctly specified model, the goodness of fit rejection rate was 5.1%, which is close to the nominal level 5%. The proposed test did not indicate lack of fit for the regression models in the CE study. **CONCLUSIONS:** The proposed GOF test provides an assessment of model fit for models with binary outcomes and repeated measurements for predictor variables.

PRM191

ANNUAL HEALTH CARE CHARGES AND UTILIZATION IN ATRIAL FIBRILLATION (AF) PATIENTS ON DRONEDARONE COMPARED TO AMIODARONE

Dahal A, Marrouche N, Ghate SR

University of Utah, Salt Lake City, UT, USA

OBJECTIVES: Amiodarone, a frequently used drug for AF patients, has been associated with severe adverse events. A relatively new drug, dronedarone, has been shown to reduce the risk of adverse events and duration of hospital stays in AF patients. The objective of this study was to assess the annual health care charges and utilization in AF patients on dronedarone compared to amiodarone. **METHODS:** Data from the University of Utah Enterprise Data Warehouse were analyzed for AF patients from October 2009 to October 2012. Eligible patients had a prescription for either amiodarone or dronedarone on index date; had 6 months pre-index and 12 months post-index follow-up activity. Annual total charges and annual inpatient and outpatient visits were assessed during the follow-up period. Generalized linear model (GLM) with gamma distribution and log link and negative binomial model (NBM) were used to examine the annual charges and usage between the two groups, respectively, controlling for demographics, insurance status, baseline comorbidities, and prior drug use. **RESULTS:** Of the 1003 patients analyzed, 134 (13.4%) patients were prescribed dronedarone and 869 (86.6%) were prescribed amiodarone. The age and gender distribution was not significantly different between the two groups ($p>0.05$). The mean unadjusted annual health care charges for dronedarone were significantly lower compared to amiodarone ($40,395 vs. $96,387, $p<0.05$). The mean annual outpatient visits for dronedarone were significantly higher compared to amiodarone (7.96 vs. 4.78, $p<0.05$). GLM results indicate that dronedarone patients had 71% lower annual health care charges compared to amiodarone patients (coeff. -0.711, $p<0.05$). NBM results show that dronedarone patients were 61% less likely to have inpatient visits (coeff. -0.61, $p<0.05$) and 39% more likely to have outpatient visits (coeff. 0.39, $p<0.05$) compared to amiodarone patients. **CONCLUSIONS:** The annual health care charges and inpatient visits were significantly lower but outpatient utilization was higher in dronedarone compared to amiodarone patients.

PRM192

USING PARAMETRIC SURVIVAL CURVES TO ESTIMATE PROGRESSION FREE SURVIVAL IN A NETOWRK META-ANALYIS OF TREATMENTS FOR CHRONIC LYMPHOCYTIC LEUKEMIA

Devine B[1], Cheng MM[2]

[1]University of Washington, Seattle, WA, USA, [2]Abbott Diabetes Care, Alamdea, CA, USA

OBJECTIVES: When comparing survival time among competing treatments, the assumption of constant hazards, necessary for use of semi-parametric modeling techniques, cannot always be met. Use of a fully parametric model provides a more flexible approach and a better estimate of treatment effects. Using the results of our already published network meta-analysis, our objective is to focus on the use of parametric survival curves to estimate progression free survival (PFS) in patients with chronic lymphocytic leukemia. **METHODS:** We tested parametric Weibull and log-logistic regression models with a two-parameter relative treatment effect (scale and shape), to indirectly compare PFS from multiple trials. We scanned survival curves from each included study, and used data from each consecutive interval to calculate model parameters. We estimated the number of deaths using the binomial likelihood distribution. We conducted the network meta-analysis using Bayesian statistical methods. We fit fixed and random effects models, and modeled scale and shape parameters on the log hazard scale. We evaluated goodness-of-fit by visually inspecting the linearity of diagnostic plots and comparing deviance information criteria (DIC). Using parameter estimates from the posterior summary we derived hazard rates, hazard ratios (HRs) and PFS survival curves for each treatment. To estimate the mean duration of PFS for each treatment, we calculated the area under each PFS curve. **RESULTS:** Seven randomized controlled trials of five treatments were included. The fixed effects Weibull model was the best fit for the data, with stronger linearity in the diagnostic plots and a lower DIC value. Hazard rates, HRs, PFS, survival, and median survival, with 95% credible intervals, were calculated for each treatment. Results suggest the hazard of disease progression for two treatments was constant, and increased over time for the other three. **CONCLUSIONS:** Parametric survival methods are useful in comparing PFS in the oncology setting.

PRM193

APPLICATION OF RECLASSIFICATION MEASURES IN COMPARING RISK ADJUSTMENT MODELS

Mehta HB, Mehta V, Girman CJ

Merck, Sharp & Dohme, North Wales, PA, USA

OBJECTIVES: A traditional statistical measure such as concordance(c) statistics has been used widely for comparison of risk adjustment models; however, c-statistic has been criticized for its insensitivity. To overcome the limitations of c-statistics, novel reclassification measures (reclassification calibration statistics, Net Reclassification Index (NRI) and Integrated Discrimination Index (IDI)) have recently been proposed. The objective was to compare Charlson Comorbidity Index (CCI), Chronic Disease Score (CDS) and CCI+CDS in predicting one year mortality in type-2 diabetes mellitus patients (T2DM) by applying novel reclassification measures. **METHODS:** The Clinical Practice Research Database, electronic medical record data from UK, was used for this retrospective longitudinal cohort study. Patients diagnosed with T2DM from January 1, 2006 to December 31, 2006 were included. Diagnosis and prescription information upto 1-year prior to the index date, i.e. first date of T2DM diagnosis, was used to create CCI and CDS, respectively. Patients were followed for 1 year from the index date to observe mortality. Descriptive statistics was used to describe the study cohort. In addition to traditional c-statistics from logistics regression, NRI and IDI were used to compare risk adjustment models. **RESULTS:** The cohort consisted of 26,191 patients with T2DM. The mean CCI and CDS were 0.24±0.67 and 1.58±1.06, respectively. The c-statistics values for CCI, CDS and CCI+CDS models were 0.791 (95%CI: 0.777-0.805), 0.788 (95%CI: 0.774-0.802) and 0.803 (95%CI: 0.789-0.817), respectively. The CDS and CCI+CDS reclassified 1.92% ($p=0.238$) and 6.50% ($p=<0.001$) patients into correct strata compared to the CCI. The IDI values for CDS and CCI+CDS were -0.64% ($p=<0.001$) and 0.43 ($p=<0.001$). This means that addition of CDS in CCI improved the prediction of mortality. **CONCLUSIONS:** Combined score (CCI + CDS) performed better than individual scores. In addition to c-statistics, reclassification measures such as NRI and IDI can be added to the armamentarium of risk adjustment model comparisons.

PRM194

MATCHING-ADJUSTED INDIRECT COMPARISONS: A SIMULATION STUDY OF STATISTICAL PERFORMANCE

Signorovitch J, Ayyagari R, Cheng D, Wu EQ

Analysis Group, Inc., Boston, MA, USA

OBJECTIVES: When indirectly comparing treatments across separate clinical trials, matching-adjusted indirect comparisons (MAICs) can help avoid bias due to cross-trial differences in baseline characteristics. The approach uses propensity scores to adjust individual patient data from trials of one treatment to match published baseline characteristics from trials of comparator treatments. We assessed the statistical properties of MAIC, including the accuracy of estimated treatment effects and their standard errors, in a simulation study. **METHODS:** Each simulation scenario included two randomized controlled trials with a common control arm and a dichotomous outcome. Sample sizes ranged from 125 to 1000 patients per arm. Cross-trial differences in baseline characteristics were simulated to generate low, moderate and high levels of potential bias. For each simulated dataset, MAIC was used to estimate the relative treatment effect using individual patient data from one trial and aggregate data from the other. Estimated treatment effects and standard errors were evaluated for accuracy across 1000 simulations. Indirect comparisons without matching adjustment were evaluated in parallel. **RESULTS:** By design, indirect comparisons without matching exhibited biases ranging from 10% to 200% of the true treatment effect across simulation scenarios. In contrast, the MAIC estimators exhibited negligible bias, falling within +/- 2% of the true treatment effect when all confounding variables were considered. The sandwich estimator closely approximated the true standard errors, and was slightly conservative, overestimating by as much as 8%, but usually less than 5%. These findings were consistent across the range of investigated sample sizes and levels of confounding. **CONCLUSIONS:** MAIC can remove bias due to observed cross-trial differences and provide reliable assessments of statistical uncertainty for indirect comparisons that combine individual patient data and aggregate data.

PRM195

BAYESIAN MODELS WITH A WEAKLY INFORMATIVE PRIOR: A USEFUL ALTERNATIVE FOR SOLVING SPARSE DATA PROBLEMS

Soliman AM[1], MacLehose RF[2], Carlson A[3]

[1]College of Pharmacy, University of Minnesota, Minneapolis, MN, USA, [2]University of Minnesota, Minneapolis, MN, USA, [3]Data Intelligence Consultants, LLC, Eden Prairie, MN, USA

OBJECTIVES: Separation problems (perfect prediction of a binary outcome by one or more covariates) are common in health outcomes research in high prevalence,