0.40), and 0.72 (0.50-0.90), respectively, but were stable over time (re-test). For data available in 2009, incidence of PML (cases per 100,000 patients) over two years was estimated at 0.22 (95% CI 0.00078–3.13) for patients initiated on efalizumab and 1.21 (0.0052–14.81) for patients remaining on treatment. ORs versus placebo were 1.7 (1.4–2.2) for AEs and 1.5 (0.6–5.1) for SAEs. Combination of normalized weights and scores yielded a safety estimate of efalizumab versus placebo in 2009 of -0.65 (stdev 0.28) (on a scale of -1 to +1). The same approach was applied to compare efalizumab with three active comparators. **CONCLUSIONS:** The method supports comparative safety assessment and captures uncertainty in judgments and in data, allowing the inclusion of rare, disparate and newly emergent safety data in quantitative assessments.

PRM17

LONGITUDINAL AND CROSS SECTIONAL ASSESSMENTS OF HEALTH UTILITY IN ADULTS WITH HIV/AIDS: A SYSTEMATIC REVIEW AND META-ANALYSIS

Tran B
Hanoi Medical University, Hanoi, Vietnam

OBJECTIVES: Utility estimates are important health outcomes for economic evaluation of care and treatment interventions for patients with HIV/AIDS. We conducted a systematic review and meta-analysis of utility measurements to examine the performance of preference-based instruments, estimate health utility of patients with HIV/AIDS by disease stages, and investigate changes in their health utility over the course of antiretroviral treatment. **METHODS:** We searched PubMed/Medline, Cochrane Database of Systematic Review, NHS Economic Evaluation Database and Web of Science for English-language peer-reviewed papers published during 2000-2013. We selected 49 studies that used 3 direct and 6 indirect preference based instruments to make a total of 218 utility measurements. Random effect models with robust estimation of standard errors and multivariate fractional polynomial regression were used to obtain the pooled estimates of utility and model their trends. **RESULTS:** Reliability of direct-preference measures tended to be lower than other types of measures. Utility elicited by two of the indirect preference measures - SF-6D (0.171) and EQ-5D (0.114), and that of Time-Trade off (TTO) (0.151) was significantly different than utility elicited by Standard Gamble (SG). Compared to asymptomatic HIV patients, symptomatic and AIDS patients reported a decrement of 0.025 (p=0.40) and 0.176 (p=0.001) in utility scores, adjusting for method of assessment. In longitudinal studies, the pooled health utility of HIV/AIDS patients significantly decreased in the first 3 months of treatment, and rapidly increased afterwards. Magnitude of change varied depending on the method of assessment and length of antiretroviral treatment. **CONCLUSIONS:** The study provides an accumulation of evidence on measurement properties of health utility estimates that can help inform the selection of instruments for future studies. The pooled estimates of health utilities and their trends are useful in economic evaluation and policy modelling of HIV/AIDS treatment strategies.

PRM18

IMPROVING PARAMETER ESTIMATION FOR A DECISION-ANALYTIC MARKOV MODEL TO EVALUATE THE USE OF NOVEL BIOMARKER LED STRATEGIES FOR PREVETION OF CARDIOVASCULAR DISEASE

Lamrock F[1], Cairns K[1], Kee F[1], Conrads-Frank A[2], Schnell-Inderst P[2], Salomaa V[3], Blankenberg S[4], Zeller T[4], Saarela O[5], Kuulasmaa K[3], Kühne F[2], Siebert U[2]
[1]*Queen's University Belfast, Belfast, UK,* [2]*UMIT - University for Health Sciences, Medical Informatics and Technology, Hall i.T., Austria,* [3]*THL - National Institute for Health and Welfare, Helsinki, Finland,* [4]*University Health Center Hamburg, Hamburg, Germany,* [5]*McGill University, Montreal, QC, Canada*

OBJECTIVES: This study sought to compare and improve the methods for estimating input parameters to populate a decision-analytic Markov model for use in the health economic assessment of novel biomarkers for primary prevention of Cardiovascular Disease. **METHODS:** An initial five-state Markov model is built using the FINRISK97 cohort and follow-up for coronary and stroke events. Two different approaches to estimate the transition rates between health states are used. The first approach involves a separate analysis for each of the transitions between different health states each using a separate Cox Proportional Hazards model (censoring any movements outside the health states of interest). In the second approach, all the transitions are examined in one process using the R package 'msm' to maximize the overall likelihood expression. **RESULTS:** The first approach requires considerable data manipulation to account for the censoring in each separate analysis; however convergence of results is less problematic. While msm considers the overall likelihood expression, it has difficulty fitting a five-state model due to the number of parameters which need estimated. Two recommendations from this research for using msm include building the model iteratively, adding covariates one step at a time. For example to build the model with age plus one covariate, the baseline transition matrix and regression coefficients from the age only model are used as initial guesses. The second recommendation would be to use the regression coefficients obtained in the first approach for initial "guesses" of the coefficients in the second approach. **CONCLUSIONS:** Parameter estimation for the Markov process within decision-analytic Markov models can be obtained from separate sources, but we recommend fitting Markov models to one source of longitudinal data. The output for the parameter estimates varied between the approaches used, and while msm is a useful tool for parameter estimation, it requires reasonable initial approximations.

PRM19

DO DIFFERENT MODELING TECHNIQUES CHANGE RANKINGS OF HOSPITAL PERFORMANCE? : MULTILEVEL MODELING VS. STANDARD LOGISTIC REGRESSION

Yucel A, Ferries EA, Sharma M, Johnson M, Aparasu R
University of Houston, Houston, TX, USA

OBJECTIVES: There are conflicting results about whether using multilevel modeling (MLM) produces different rankings in hospital performance compared to using traditional statistical techniques. We would like to compare hospital rankings obtained from standard logistic regression (LR) relative to MLM modeling of risk-adjusted hospital mortality rates for stroke among hospitals in the south. **METHODS:** The 2012 U.S. Nationwide Inpatient Sample (NIS) was used to identify patients with a primary diagnosis for ischemic stroke, using ICD-9 diagnosis criteria. Stepwise backward selection technique with logistic regression and multilevel modeling was performed to examine the variation among hospitals in the south by adjusting for patient-level and hospital-level risk factors. The analysis included hospitals with more than 30 stroke cases in the southern region of the U.S. The hospital IDs have been masked in order to comply with the NIS data user agreement. **RESULTS:** There were 19,071 stroke hospitalizations in 320 hospitals. The same rankings were observed for top performers in hospitals' observed/expected ratios (O/E) by LR and MLM. However, high O/E outlier status differed between the two statistical methods, yielding different rankings for worst performance. Hospital K, L, M were ranked as the worst 3 performers in LR whereas K, N, L were ranked the same in MLM. O/Es of Hospital K were 4.49 (CI=1.62-9.84) and 3.942 (CI=1.42-8.64) in LR and MLM, respectively. O/E's of Hospital L were 4.24 (CI=1.90-8.13) and 3.33 (CI=0.70-9.54) in LR and MLM, respectively. LR identified Hospital M's O/E ratio as 3.58 (1.47-7.26). MLM identified Hospital N's O/E as 3.33 (CI=0.70-9.54). **CONCLUSIONS:** Although both logistic regression and multilevel modeling produced similar rankings for top performers, the MLM approach was more conservative in its O/E ratio estimates which were shrunken towards the overall mean. The MLM method is recommended to better statistically adjust and avoid false positive identification of outliers.

PRM20

COMPARING THE PREDICTIVE PERFORMANCE OF TWO VARIANTS OF THE ELIXHAUSER COMORBIDITY MEASURES FOR ALL-CAUSE IN-HOSPITAL MORTALITY IN A LARGE MULTI-PAYER U.S. ADMINISTRATIVE DATABASE

Fortin Y[1], Crispo J[1], Emons MF[2], Mattison DR[3], Krewski D[1]
[1]*University of Ottawa, Ottawa, ON, Canada,* [2]*Cerner Corporation, Culver City, CA, USA,* [3]*Risk Sciences International, Ottawa, ON, Canada*

OBJECTIVES: Tools used to address confounding, such as measures of patient comorbidity, should preferably be validated prior to conducting health services and outcomes research using novel sources of administrative health data. Our objective was to validate and compare the predictive performance of two variants of the Elixhauser comorbidity measures for predicting all-cause, in-hospital mortality using Cerner Health Facts®, a large nationwide U.S. multi-payer electronic health record database. **METHODS:** Quan et al. and the AHRQ version 3.7 adaptations of the Elixhauser comorbidity measures were selected to estimate the comorbidity burden of Health Facts® inpatients, outpatients, and emergency care recipients aged 18 to 89 years from Jan. 2000 to Dec. 2012. Diagnoses identified one and two years preceding an index encounter were assessed in a group of unique patients. C-statistics obtained from forward stepwise multiple logistic regression fitted models and ROC curve differences were derived to compare the predictive performance of each variant/lookback period combination. **RESULTS:** Of 8,128,713 patients with a mean age of 51.4 years and a male composition of 41.6%, 76,235 (0.9%) were deceased at 1-year. Age, sex and all comorbidities were significantly correlated with mortality in the two Elixhauser variants irrespective of lookback period. Excellent predictive performance (c-statistic) was found for the Quan (c1yr=0.874, c2yrs=0.873) and AHRQ (c1yr=0.871, c2yrs=0.870) variants and both exceeded the baseline model (c=0.803, P<.0001) limited to age and sex. The Quan variant performed slightly better than the AHRQ variant in both lookback periods. However, increasing the lookback period from 1 to 2 years did not improve predictive performance. **CONCLUSIONS:** Variants of the Elixhauser comorbidity measures were validated in Cerner Health Facts® and demonstrated excellent predictive performance for all-cause in-hospital mortality at one year. These results support the utility of Health Facts® as a novel and attractive source of data for administrative database research.

PRM21

STATISTICAL IDENTIFICATION OF PATIENT SELECTION BIAS IN RETROSPECTIVE CHART REVIEWS

Signorovitch JE[1], Reichmann WM[1], Li N[1], Liu Z[2], Hao Y[2], Perez JR[2]
[1]*Analysis Group, Inc., Boston, MA, USA,* [2]*Novartis Pharmaceuticals Corporation, East Hanover, NJ, USA*

OBJECTIVES: Retrospective chart reviews are often used to generate real-world evidence of patient outcomes. Physicians participating in these studies may be asked to randomly select patient charts for inclusion among all charts meeting eligibility criteria. Random selection is important for obtaining representative samples and unbiased estimates of patient outcomes. This study examined a statistical method for detecting non-random selection of patient charts with an application to real data. **METHODS:** Example data were drawn from a recent retrospective oncology chart review study that estimated overall survival (OS) and progression-free survival (PFS) following the start of treatment. In the study, participating physicians were asked to provide a random sample of their eligible patients. Latent class analysis (LCA) was used to test whether the distribution of PFS and OS was consistent with random selection for all charts (a one class model) versus a mixture in which some physicians selected random charts and others selected non-random, convenient charts (models with > 1 class). The best fitting model was identified using information criteria. Physician characteristics, PFS and OS were compared across latent classes. **RESULTS:** In overall sample of charts, median durations of OS and PFS were significantly longer than published benchmarks drawn from randomized trials. LCA identified evidence for two classes of physicians. One class was consistent with biased selection for recently seen patients who were still alive at the time of chart review, resulting in bias towards longer OS and PFS. The other class was consistent with random selection, and exhibited OS and PFS that were significantly closer to published benchmarks. **CONCLUSIONS:** In this example using data from a retrospective oncology chart review, LCA was helpful in identifying physicians who may have selected convenient but non-random patient charts. LCA warrants