models with inverse probability weighted (IPW) estimation techniques to address those challenges in a single model. METHODS: Generalized random effect models were used with weights that are calculated as inverse of probability being uncensored. The Gaussian family and log link function were chosen. We applied a test to see if possible censoring bias was present and also calculated the deviation from the consistent variance of standard pooled ordinary least squares were used. RESULTS: A total of 4,824 observations were used in the analysis. We obtained Medicare claim files for the two years following a lung-cancer diagnosis. Costs had high kurtosis and skewness. Moreover, 30% of the cases were censored. Therefore, their annual costs were not observed. The total cost of all care is $60K for the two years following a lung-cancer diagnosis and $55K for incomplete cases. Results significantly diverged from the standard regression model (p = 0.001). CONCLUSIONS: This paper applies the inverse probability weighted estimation bias in a retrospective cohort of patients newly diagnosed with lung cancer. Our findings suggest that the standard regression model yields an inconsistent estimator due to censoring bias. IPW least square estimation method removes that bias.

A REVIEW OF THE EVOLUTION OF HEALTH ECONOMIC MODELS OF SMOKING CESSATION

Fisher MP, Ploutz DR, Knight CJ

Horizon Evidence Development Ltd, Luton, UK, 1 Heron Evidence Development Ltd, Stapleby, Luton, UK

OBJECTIVES: To review how health economic models of the cost-effectiveness of smoking cessation interventions have evolved over the last ten years. METHODS: The Medline database was searched via PubMed using the keywords 'cost-effectiveness models smoking cessation'. Articles not available or dating before 1998 were excluded. The abstracts from these results were examined, and relevant studies were selected for the review. A bibliographic search was implemented on the selected studies and additional relevant studies were included. An extraction grid was built to record key comparable attributes of each included model as reported. RESULTS: Thirteen studies were included and extracted, dating from 1999-2008. The most common model design was Markov model simulations of hypothetical patient cohorts given various interventions (6 studies). Reported cost perspectives have widened over the period; since 2004, societal perspectives have been the most common (7/8 studies) and previously direct perspectives were more common (4/5 studies). Prior to 2004, 3/5 studies were of US patients whereas then 4 were for the US, 3 Scandinavia. The interventions included in the models were initially Nicotine Replacement Therapy (NRT) and bupropion (Zyban), but this has widened recently to also include varenicline (Champix), which was approved by the FDA in May 2006. It was not possible to draw clear conclusions as to the values of cost, health outcome and cost-effectiveness made given the wide variety of study designs, perspectives and countries modelled and the small number of studies. CONCLUSIONS: There is wide variability in the nature of published health economic models of smoking cessation in terms of perspective, country and model structure. Some trends have been observed. Markov models were the most common model design to use for modelling smoking cessation cost-effectiveness in the articles we reviewed.

USE OF STABILIZED INVERSE PROPENSITY SCORES AS WEIGHTS TO DIRECTLY ESTIMATE RELATIVE RISK AND ITS CONFIDENCE INTERVALS

Xu S, Ross C, Rabbiel MA, Shetterly S, Blanchet CM, Smith DH

Kaiser Permanente Colorado, Denver, CO, USA, 1 Locality Respiratory Research Institute, Albuquerque, NM, USA, 2 Center for Health Research, Kaiser Permanente Northwest, Portland, OR, USA

OBJECTIVES: Inverse propensity scores have been used in observational studies to reduce selection bias. To obtain estimates of the main effects, a pseudo data set is created with weights copies of observations for each subject and analyzed with conventional regression models. Currently variance estimation requires specific formulas depending on type of outcomes. Our goal is to demonstrate a statistical approach to directly obtain the correct estimates of variance of the main effects (usually group difference in outcomes) in regression models. METHODS: We carried out theoretical and simulation studies to show that the variance of the main effects estimated directly from regressions is underestimated, and that the type I error rate is higher due to the inflated sample size in the pseudo data. We propose to use the stabilized weights to directly estimate both the main effect and its variance from conventional regression models. RESULTS: We applied the approach to a study examining the effectiveness of serum potassium monitoring in reducing hyperkalemia-associated adverse events among 27,362 diabetics patients newly-prescribed a renin-angiotensin aldosterone system (RAAS) inhibitor from three HMO Research Network sites. The relative risk (with monitoring versus without monitoring) and confidence intervals were 0.52 (0.39, 0.70) using stabilized weights compared to 0.52 (0.43, 0.64) using typical inverse propensity scores. The sample size in the pseudo data using the stabilized weights was 27,312 compared to 54,632 using inverse propensity scores. The sample size in the pseudo data using the stabilized weights is slightly smaller than the original 27,362 largely due to missing covariates in the propensity score model. CONCLUSIONS: We demonstrated via theoretical, simulation results, and real data example demonstrate that the use of the stabilized weights in the pseudo data preserves the sample size of original data, produces correct estimation of the variance of main effect, and maintains an appropriate type I error rate.

“UNNATURAL” HISTORY: MODELING DISEASE PROGRESSION USING OBSERVATIONAL DATA

Najmi K, Bajorina A, Chipper AR, Schuld W, Hteza UK, Holloway K, Dick A

University of Rochester School of Nursing, University of Rochester: New York, NY, USA, 1 The RAND Corporation, Pittsburgh, PA, USA

OBJECTIVES: Cost-effectiveness analysis requires comparison of outcomes in treated and untreated populations. Data from randomized controlled trials (RCT) do not provide progression, therefore, their analogue of the general population, while treatment effects in observational data may be biased due to non-randomization. We developed a novel approach for estimating untreated progression rates (controls) by using data from a nationally representative patient cohort, as well as RCT estimates. METHODS: We used data from the 2000-2005 SEER-5 to estimate a Markov model progression was characterized by disability-based disease states and relapses. We modeled transitions of disease state changes using a first-order annual Markov model that adjusted for age, gender, disease duration, recent relapse rates, prior states, and the specific disease-modifying therapy (DMT). We developed an iterative multinomial logistic regression algorithm, constraining the effects of DMT to match those reported by RCTs. RESULTS: After correcting for the DMT treatment effects and other observable risk factors, the probability of disability progression was greater for estimates based on all MS patients compared to the estimates based on untreated individuals only. The 95% confidence intervals for the entire cohort (including treated and untreated individuals) were narrower than the intervals based on the subsample of untreated patients. CONCLUSIONS: Our results indicate that the untreated patients in our study had lower estimates of disease progression than the treated and that untreated participants would have a lower disability progression rate. This suggests that patients who forgo treatment are likely to have milder, slower progressing forms of MS. Correcting for treatment effects in a more inclusive group of patients likely provides a more realistic estimate of disease progression than simply characterizing progression in an untreated cohort. The use of a broader cohort also improves the precision of disease progression estimates.

A FACTORIAL SIMULATION OF THE GAINS FROM AN EMPIRICAL BAYES APPROACH OVER CLASSICAL METHODS

Murray JJ, Fryback DG

AI Lilly and Company, Indianapolis, IN, USA, 1 University of Wisconsin-Madison, Madison WI, USA

OBJECTIVES: Classical methods assume that any and all information exists only in the data; Empirical Bayes (EB) methods assume, under certain conditions, that an informative prior can be calculated from the data. We conducted a factorial simulation to examine the gains achieved by EB over classical methods. METHODS: We assumed that exchangeable units of analysis are available from which to calculate prior information. A non-parametric prior was estimated as a mixture of normal distributions. The simulation used a two-level, four-factor design to generate one-hundred sixty scenarios. The factors varied were: (1) the number of units of analysis [25, 50], (2) the sample size within each unit of analysis [10 or 25], (3) the number of distributions within the mixture [2 or 4] and (4) the average effect sizes between the distributions within the prior mixture [0.2 and 0.8 standard deviations]. A prior was estimated for each scenario using the EM algorithm across all units of analysis. This prior was combined with the individual data to estimate posterior means; bootstrap methods were used to estimate Bayesian confidence intervals. The Bayesian and Classical results were compared within scenarios on measures of “goodness”; (1) Absolute difference percentage, “true” mean and the classical and EB estimates, (2) frequency of the EB mean being closer to the “true” mean, (3) average widths of the EB and classical intervals, and the actual probability coverage of the estimated intervals. RESULTS: EB methods were consistently closer in absolute difference and were more frequently closer to the “true” mean. EB confidence intervals were consistently narrower giving less uncertainty about the location of the true mean while maintaining the expected probability coverage. CONCLUSIONS: Using prior information based on certain assumptions, Empirical Bayes methods provide better estimates of means and confidence intervals compared to classical methods.

REGRESSION ESTIMATORS FOR QUALITY OF LIFE AND QUALITY-ADJUSTED LIFE YEARS (QALY)

Basu A, Miao A

University of Chicago, Chicago, IL, USA, 1 University of York, York, UK

OBJECTIVES: Unlike the literature on estimators for cost data, not much attention has been paid to the appropriate use of regression estimators for the analysis of health benefits in economic evaluation, i.e. EQSD and QALYs. We explored the use of beta regression models to address characterics - such as ceiling effect, truncated distribution, and heteroscedasticity - typical of quality of life data. METHODS: We developed a two-part beta regression model, and an alternative Bayesian, defined incremental and marginal effects of covariates on the mean EQSD, and proposed relevant estimators. Using the UK multicentre EVAlUATE trial as a motivating example, we compared results from OLS regression to those obtained from our beta-based estimators in terms of estimated mean treatment effect and variance for other covariates. We complemented the case study with a microsimulation exercise. RESULTS: The OLS regression fit the EVAlUATE trial EQSD data as well as any of the advanced estimators based on Beta distribution. The estimated treatment effects were 0.0099 (SE - 0.011; 90% CI -0.022, 0.038) under standard OLS regression.
and 0.0099 (SE = 0.0085; 90% CI: 0.00002, 0.028) under one of the Beta-based estimators, while other Beta-based estimators produced similar results. Our simulation results revealed substantial advantage of the Beta-based estimators over OLS when the true incremental effect is large, but this advantage dissipates as the incremental effect gets smaller. CONCLUSIONS: One and two-part Beta regression models can provide substantial benefits, both in terms of bias and efficiency, over traditional OLS regression in modeling quality of life data such as EQ5D only when the true incremental effects are large but not when they are small. Our case study confirms these conclusions. Further exploration is required to provide definite ranges of incremental effects where alternative methods are to be preferred.

CONCEPTUAL PAPERS & RESEARCH ON METHODS – Patient-Reported Outcomes Studies

PMCS1

TESTING THE ROLE OF TIME IN AFFECTING THE UNIDIMENSIONALITY IN EQ-5D USING THE 2002-2003 MEDICAL EXPENDITURE PANEL SURVEY (MEPS): A FACTORS MODEL APPROACH

Gu NY; Doctor JN

University of Southern California, Los Angeles, CA, USA

OBJECTIVES: To evaluate the measurement property of the EQ-5D and to test the role of time in affecting the unidimensionality in EQ-5D. METHODS: We extracted a sample from the two-year panel (2002-2003) in Medical Expenditure Panel Survey (MEPS). Respondents were included if they completed the EQ-5D, were ≥18 years of age and were diagnosed with any of the top ten most prevalent chronic conditions using ICD-9-CM (n = 1428). Data sets using point-in-time were constructed to examine the EQ-5D 1) cross-sectional two repeated measures for each respondents. For measures and, 2) longitudinal cross-sectional study, we used Rasch rating scale model. For longitudinal analysis, we applied FACTORS model by parameterizing time with the hypothesis that time plays a significant role in characterizing the unidimensionality of the EQ-5D. Unidimensionality was evaluated using the goodness-of-fit of the EQ-5D items to the measurement models. Time sensitivity effect was examined via item x time interactions to evaluate the measurement invariance property of the EQ-5D. RESULTS: Results from both point-in-time measures revealed that the items “anxiety/depression” constantly showed misfit in all chronic conditions (infit/oufit mean square >1.3). Results from longitudinal analysis demonstrated that including time rendered improved model fit as unidimensionality of the EQ-5D was achieved in eight out of ten chronic conditions. There was significant time sensitivity effects (p < 0.05) indicating that the EQ-5D items were endorsed differently over time. CONCLUSIONS: Most analysis of health measures suggest unidimensionality fails because mental health and physical health items tend to form different scales. Our findings suggest that time mediates unidimensionality and that physical health and mental health are linked through time. As time is essential to all chronic disease diagnoses, controlling for time in a longitudinal study of health measurement is important for proper assessment on the measurement properties of health instruments.

PMCS2

CULTURAL ADAPTATION AND VALIDATION OF A HEATH RELATED QUALITY OF LIFE QUESTIONNAIRE: FUNCTIONAL OUTCOMES OF SLEEP QUESTIONNAIRE

Mihali S, Ozkucu E, Tukavec F

Istanbul University, Ankara, Turkey; Ankara University, Ankara, Turkey

OBJECTIVES: The aim of this study was to determine the reliability and validity of the Turkish versions of Sleep Questionnaire (FOSQ) for the Turkish population. METHODS: A total of 200 university students participated. A total of 52% of them were men, 48% of them were women. A total of 43% of them lived in the dormitory and 57% of them at the apartments. Data were collected with a sociodemographic form, FOSQ, and SF-36. Cronbach's Alpha was used to assess reliability and factor analysis to assess dimensionality. The SF-36 was used for concurrent validity. RESULTS: The internal consistency coefficient (Cronbach's alpha) of FOSQ was 0.90. Factor analysis of the scale revealed that it was composed of three factors with Eigenvalues >2.0, accounting for 73.6% of the total variance. All items of the Turkish FOSQ had a factor load ranging from 0.413 to 0.731. There was a strong relationship between Functional Outcomes of Sleep Questionnaire, and SF-36. CONCLUSIONS: The culturally adapted Functional Outcomes of Sleep Questionnaire had good validity and reliability, making it a potentially useful outcome measure in the evaluation of insomnia patients in Turkey.

PMCS3

SHOULD AN SF-10 REPLACE THE SF-12? – Comparisons are provided

University of Southern California, Los Angeles, CA, USA

OBJECTIVES: To evaluate the item measurement property of the SF-12v2 and the merits of the physical and mental health component scales (PCS and MCS) using Rasch model. METHODS: We extracted a sample from the 2005 Medical Expenditure Panel Survey (MEPS). We included respondents who had completed the SF-12v2, were ≥18 years of age and were diagnosed with any of the top ten most prevalent chronic conditions using the primary ICD-9-CM (n = 5151). Rasch analyses were applied to subgroups on all SF-12v2 items as a 12-item scale as well as on PCS and MCS items separately as two 6-item scales. Items were evaluated using the Rasch model fit statistics to examine 1) whether or not the 12 items measure a unidimensional construct and 2) whether or not PCS items measure a physical health construct and MCS items measure a mental health construct. RESULTS: In the 12-item scale, the two mental health (MH) items “Have you felt calm and peaceful?” and “Could you feel downhearted and depressed?” consistently demonstrating misfit in all subgroups (unfit/oufit mean square >1.3). By subdividing the PCS and MCS items into two separate 6-item scales, all six MCS items had good model fit including the two MH items. On the other hand, two PCS items concerning general health and bodily pain showed misfit to the Rasch model. CONCLUSIONS: The SF-12v2 items do not form a unidimensional scale as the two MH items did not fit with other items. However, there are tradeoffs of treating PCS and MCS as two distinctive scales as PCS items had a reduced fit. A generic SF-10 for adult population may be a unidimensional component of the SF-12v2 and could be modeled as a brief health survey instrument that improves upon the SF-12v2.

PMCS4

ARE YOU CHILDREN Aged 6 TO MAKE TRADE-OFFS? A QUALITATIVE STUDY OF CHILDREN Aged 6 TO 12 YEARS AND THEIR PERCEPTIONS OF INFLUENZA VACCINATION

Hood E, Rousseau MD, Mattess M, Toback V, Divine V, Mahadeva P, Beatson KM

Oxford Outcomes, Bethesda, MD, USA; Medimmune, Inc; Gallow, MD, USA; Oxford Outcomes, Goughern, MD, USA

OBJECTIVES: This study explored children’s perceptions of influenza and their ability to make trade-offs among the vaccine attributes. METHODS: A qualitative study was performed to inform the development of a conceptual model on children’s perceptions of influenza vaccination. Interviews were conducted by a trained interviewer using an interview guide. A thematic analysis was conducted to understand perceptions of influenza and vaccination including influenza vaccine attributes such as efficacy, side effects, and mode of administration. Study subjects were children aged 6 through 12 years from the Washington, DC area. RESULTS: A total of 28 children (two males and two females per year of age) participated. Parents reported that 75% of the children had previously been vaccinated against the influenza virus. Knowledge of influenza varied among children; some were unable to describe influenza illness, while others described it as a virus and were able to list specific symptoms. Once influenza was described by the interviewer, all children demonstrated a basic understanding of the illness. Efficacy, risk of side effects and mode of administration were found to be children’s decision to be vaccinated. Children as young as 8 years old were able to differentiate low, medium and high risk of side effects and showed the ability to make rational and consistent tradeoffs between vaccines with differing levels of efficacy, chance of side effects and/or mode of administration. Under certain vaccine scenarios, each child was willing to be vaccinated. CONCLUSIONS: Children may be able to make decisions pertaining to influenza vaccination. Young children appear to understand the concept of risk and are able to make trade-offs among influenza vaccine attributes.

PMCS5

IMPACT OF THE FDA DRAFT GUIDANCE ON PATIENT REPORTED OUTCOMES (PRO) LABEL CLAIMS FOR APPROVED DRUG PRODUCTS IN THE US: HAS IT MADE A DIFFERENCE?

Mordin H, Clark M, Siersma C, Copley-Marman K, Gnosackson A

Wells Health Solutions, Ann Arbor, MI, USA; AstraZeneca Pharmaceuticals Corporation, East Hanover, NJ, USA

OBJECTIVES: To determine the influence of the FDA Draft PRO Guidance on obtaining PRO label claims, for drug products in the US, since its release in February 2006. METHODS: Packages inserts (indication, clinical trials sections) and medical review sections from publicly available summary basis of approvals (SBA) for FDA approved drug products named by the MAI PRO Labels database as having a US-based PRO label claim between February 2006 and August 2008 were reviewed. RESULTS: Of 33 products reviewed, 44 PRO claims were granted. Signs and symptoms (SS) represented the majority (n = 32; 73.7%) of the claims. Within this category, the greatest number of claims were pain-related (primarily based on a VAS or NRS rating scale) and more than half of all SS claims were based on a patient diary. After SBA review for the 24 drug products identified with an available medical review, 3 products were found to have collected additional PRO within the context of the registration trial (reported in the medical review) that did not result in a PRO claim; of these, 6 had PROs reported to have statistically significant results and 2 of these (Voltaire and Fontora) reported clinically meaningful results. Both products were reviewed by the same FDA Division of Anesthesia, Analgesia, & Rheumatology (DAAR) and are indicated for pain. Of the 33 products reviewed, the Pulmonary and Allergy Products review division reviewed the most claims (eleven) in the label. Of the 33 products reviewed, SEALD was involved in 4 product SBA reviews. CONCLUSIONS: Evidence suggests that since the release of the Draft PRO Guidance, many PRO claims continue to be approved by FDA reviewing divisions; however, the reviewing divisions are not always adhering to the current standards when assessing PRO data for a claim.