Abstracts from the 26th Annual Meeting of the Society for Medical Decision Making

ABSTRACTS ORAL PRESENTATIONS

OPENING PLENARY SESSION
REACHING CONSENSUS ON INTERNATIONAL PATIENT DECISION AID STANDARDS (IPDAS) FOR THEIR DEVELOPMENT AND EVALUATION

O'Connor A1, Elwyn G1, Stacey D1, Barratt A1, Barry M1, Coulter A1, Holmes-Rovner M1, Llewellyn-Thomas H1, Moondjil N1, Thomson R1, Whelan T1 and Collaboration 1

1University of Ottawa, Ottawa, ON, Canada; 2University of Wales Swansea, Swansea, United Kingdom; 3Ottawa Health Research Institute, Ottawa, ON, Canada; 4University of Sydney, Sydney, Australia; 5Harvard University, Boston, MA; 6Picker Institute Europe, Oxford, United Kingdom; 7Michigan State University, East Lansing, MI; 8Dartmouth Medical School, Hanover, NH; 9GRESAC - UMR 5823 CNRS, Lyon, France; 10University of Newcastle upon Tyne, Newcastle upon Tyne, United Kingdom; 11Hamilton Regional Cancer Centre, Hamilton, ON, Canada

Purpose: The number of decision aids is expanding exponentially from 16 identified in 1999 to over 500 in 2004. However, there is considerable variability in their elements and quality. Our objective was to establish an internationally approved set of quality criteria for their development and evaluation.

Methods: In September 2003, a group of 56 patient decision aid researchers, representing nine countries (CA, FR, US, UK, NL, AU, FI, DE, NO), identified an initial set of criteria and established a steering committee and sub-committees focused on evidence, methods, and stakeholder identification. The Evidence sub-committee summarized the theoretical and empirical links between each criterion and its potential effects on decision quality and drafted voting items. A key evidence source was a Cochrane systematic review of 35 randomized trials and an inventory of 500+ decision aids.

Results: There were 12 broad classes of quality criteria focused on using a systematic development process; providing information on options; presenting probabilities; clarifying and communicating values; guiding/coaching in deliberation and communication; describing others’ experiences with decision making; disclosing conflicts of interest; delivering decision aids on the Internet; basing information on scientific evidence; and establishing its effectiveness. The evidence supporting some criteria (e.g. providing information, presenting probabilities, clarifying values) was stronger than for others (e.g. guiding/coaching, describing others’ experiences, Internet-based). The Delphi voting process is expected to be complete by September 2004.

Conclusions: The standards derived from this process will assist developers in improving their patient decision aids and assist users (patients, practitioners) and payers in judging the quality of patient decision aids. The evidence summarises provide the basis for developing an agenda to address gaps in research.

OPENING PLENARY SESSION
PREDICTING THE IMPACT OF A PARTIALLY EFFECTIVE HIV VACCINE AND SUBSEQUENT RISK BEHAVIOR CHANGE ON THE HIV EPIDEMIC IN DEVELOPING COUNTRIES: A SOUTH AFRICAN EXAMPLE

Mesesan K1, Owens D2 and Paltiel A1

1Yale University School of Medicine, New Haven, CT; 2VA Palo Alto Health Care System, Palo Alto, CA

Purpose: To assess the impact of partially effective HIV vaccines, and subsequent changes in behavior, in a population at high risk for heterosexually transmitted HIV.

Methods: We developed a dynamic, compartmental epidemic model to simulate the impact of various low-efficacy preventive HIV vaccination scenarios in Soweto, South Africa. We adapted a previously published model (Owens, 1998 & Edwards, 1998) of HIV vaccines in a homossexual male population in San Francisco. Differential equations were estimated to govern transitions between population subgroups defined by sex, disease stage, and vaccination status. We explored changes in vaccine efficacy and post-vaccination condom use. Assumptions included (1) initial HIV prevalence of 31%, (2) an anti-retroviral-naive population, (3) exclusive male negotiation of condom use, and (4) contact rates of 0.3 partners/yr and post-partner infectivity rates of 0.050-0.228, depending upon both disease stage and male-female vs. female-male transmission.

Results: With no change in risk behavior, a 40% effective HIV vaccine would avert 181,000 infections in Soweto over a span of 10 years and reduce future HIV prevalence from 32% to 23%. A 30% effective vaccine would avert 141,000 infections and reduce HIV prevalence to 25%. A 20% effective vaccine would avert 99,000 infections and reduce HIV prevalence to 27%. Changes in risk behavior would have a significant impact on the efficacy of these vaccination programs. With a 40% effective vaccine, a 25% increase in condom use among vaccinated individuals would instead avert 223,000 infections and reduce the HIV prevalence to 21%. However, if condom use decreased by 25%, a 40% effective HIV vaccine would only avert 136,000 infections and reduce the HIV prevalence to 25%. Some combinations of vaccine efficacy and risk behavior change could even worsen the epidemic. A 30% effective HIV vaccine with a resulting 75% decrease in condom use, for example, would cause an additional 29,000 infections and increase the HIV prevalence to 33% over the 10-year period.

Conclusions: Even modestly effective HIV vaccines can confer enormous benefits in terms of HIV infections averted and decreased HIV prevalence. However, these findings are sensitive to assumptions regarding the impact of vaccination on subsequent risk behavior. For South Africa and other developing countries with similar epidemic profiles, programs to reduce risk behavior may be important components of successful vaccination programs.

DOI: 10.1177/0272989X04274093
ABSTRACTS

OPENING PLENARY SESSION

EXTENDING THE QALY MODEL TO INCORPORATE GOALS THAT ARE NOT TIME MODULATED

Hazen G
Northwestern University, Evanston, IL

Purpose: The QALY model is today the primary method for incorporating patient preferences into clinical decisions and cost-effectiveness analyses. In the QALY model, quality of health is given weight proportional to health duration. Typical patient goals addressed by QALY models, such as eliminating pain or increasing mobility, are time-modulated in the sense that their importance decreases with life duration and disappears entirely if life duration is zero. Other possible patient goals, such as an author wanting to complete a book or a parent wanting a child graduate from college, have importance that is not time-modulated. Such goals, which we call extrinsic goals, retain their importance even if life duration is zero, and cannot be adequately captured by the QALY model. Here we explore extensions of the QALY model that can capture both time-modulated and extrinsic goals.

Methods and Results: We consider von Neumann-Morgenstern utility functions that account for life duration, times-modulated goals, and extrinsic goals. We extend the preference assumptions underlying the QALY model, such as the zero condition and standard gamble independence (Myagamat et al. 1998), to this more general setting, and show that under such assumptions, the extrinsic goal component of utility combines additively (rather than multiplicatively) with quality-adjusted life duration. We examine duration surrogates for extrinsic goals, situations in which an extrinsic goal is adequately represented by the desire to survive for a particular duration. We show that two previously reported empirical violations of the QALY model - maximum endurable time preference, and the unwillingness to trade off life duration for quality when duration is short - can be counted for when the utility function includes a duration surrogate for an extrinsic goal. Finally, we revisit a published analysis of the decision to undergo carotid endarterectomy and show that the inclusion of a duration surrogate for an extrinsic goal can alter optimal treatment choice due to short-term mortality risks.

Conclusions: Von Neumann-Morgenstern utility functions that include an extrinsic goal component can not only account for observed violations of the QALY model, but can do so prescriptively, thereby providing a coherent basis for including such goals in decision and cost-effectiveness analyses.

OPENING PLENARY SESSION

HIGHLIGHTING “ADDITIONAL RISK” YIELDS MORE CONSISTENT INTERPRETATIONS OF SIDE EFFECT RISK COMMUNICATIONS

Ezemenla-Fisher H, Roberts Y, Fagerlin A, Berry H and Ubel P
VA Ann Arbor Healthcare System, Ann Arbor, MI; University of Michigan, Ann Arbor, MI

Purpose: When providing risk information, researchers often use side-by-side presentations of total risk. For example, people see their baseline risk of an outcome (e.g., a 10% chance of an MI) and then the risk experienced when taking a medication (e.g. a 4% risk of an MI). This approach, however, forces people to do mental subtraction to tell how much the risk medication causes (here, a 4% increase). We tested whether presenting this risk difference explicitly would improve comprehension in graphical and textual risk communications.

Methods: We recruited women from a demographically balanced panel to participate in an online survey experiment about tamoxifen. Participants read about the use of tamoxifen to prevent breast cancer and then saw information describing how the risk of four types of side effects would increase with tamoxifen. Risks were shown either in numerical text format or as pictographs, and participants were randomized to either see a side-by-side display of total risk with and without tamoxifen or a sequential presentation of the additional risk generated by taking tamoxifen. We also randomly varied two secondary factors that have been shown to bias risk perceptions: the denomination of risk statistics (“cost of 100” versus “out of 1000”) and the presentation order of the side effects (very unlikely, but severe, side effects first versus common but less severe side effects first). Analyses focused on women’s ratings of how much they would worry about each side effect if they were to receive tamoxifen.

Results: 1789 women completed the online survey. Across all four side effect categories, women’s worry ratings were significantly higher for side-by-side total risk presentations than for presentations that highlighted the additional risk caused by tamoxifen (all p’s < 0.001). This effect was significantly larger with textual risk descriptions than with pictographs in 3 out of 4 cases. Furthermore, while side-by-side total risk presentations repeatedly showed significant unwanted denominator and probability order effects, additional risk presentations never did.

Conclusions: Additional risk descriptions lowered worry about side effect risk by highlighting the fact that the risk must exist at baseline. Moreover, the additional risk format was resistant to two cognitive biases: order effects and denominator effects. Presenting side effect risk in a pictograph format with an additional risk specifically highlighted, yields the most consistent pattern of risk perceptions.

OPENING PLENARY SESSION

COULD WE HAVE DONE BETTER? A RETROSPECTIVE COST-EFFECTIVENESS ANALYSIS OF ROUTINE SCREENING MAMMOGRAPHY

Stout N, Rosenberg M, Teutsch-Dietz A, Smith M, Robinson S and Frybark D
University of Wisconsin, Madison, WI

Purpose: Screening mammography is recommended every 1-2 years for women over age 40 and surveys indicate that over 70% of women now participate in routine screening. Few studies have examined the societal impact on cost and quality adjusted life years (QALYs) of screening practices over the past decade. To inform current and future practice we ask retrospectively if we could have done better.

Methods: Using a validated discrete-event simulation model of the epidemiology of breast cancer, the total costs and health effects of daily hypothetical screening scenarios implemented from 1998-2000 were estimated. This analysis considered the effects for all women age 40 and older in the state of Wisconsin for the 10 year study time period. Screening scenarios varied by starting and ending age as well as frequency of mammography. Screening was set as it actually occurred during this period, was also included as a scenario. Accounting for the effects associated with screening use and breast cancer treatment, we compared the QALYs accrued and the total costs generated by each screening program. Costs and QALYs were discounted at 3%.

Results: The estimated total cost of screening and treating breast cancer in the state of Wisconsin from 1998-2000 was $3.15 billion. We estimate the total QALYs accrued were 18.2 million. The actual screening use consisted of a mixture of different screening patterns in the population. Screening across the state was done better.

Conclusions: While the estimated total costs and QALYs represent the effects for the state of Wisconsin, the results can be extrapolated to the entire US population using our methods. Our population-based analysis is unique in that it accounts for screening use prior to 1998. By recommending less frequent mammograms and ensuring more women participate, we could have achieved more quality-adjusted life years at a lower cost compared with actual practice. This should lead to reconsideration of screening polices and implementation for the future.
ORAL CONCURRENT SESSION A - PUBLIC HEALTH I
MODELING COHORT-SPECIFIC TRENDS IN BODY MASS INDEX FOR THE US POPULATION

Wong T, Colditz G and Kotzka K
Harvard School of Public Health, Boston, MA; Brigham and Women’s Hospital and Harvard Medical School, Boston, MA

Purpose: Simulation models are increasingly being applied to characterize, evaluate and set goals for public health policies. These models seek to integrate secular changes in risk factors for specified birth cohorts to evaluate disease outcomes. Cohort-specific trends may be estimated using longitudinal data, but they are rarely generalizable to the US population. As part of the Cancer Intervention and Surveillance Modeling Network (CISNET), which seeks to decompose past trends in cancer incidence and mortality into underlying changes of lifestyle risk factors, screening, and treatment, we analyzed nationally-representative cross-sectional surveys to provide longitudinal estimates of body mass index (BMI) changes for US adults.


Results: We found significant heterogeneity in weight change patterns by race, sex and birth cohorts. Mean BMI values at age 20 are higher in more recent cohorts and increase more rapidly than earlier cohorts, especially in black women. Based on the model, adults in year 2010 are predicted to be 6-17 kg (14-38 lb) heavier than adults of the same age, race, and sex in 1970. If the trend continues, we predict 33% of white men, 35% of white women, 38% of black men and 51% of black women will be obese in 2010, contrasting the Healthy People 2010 goal of no more than 15% prevalence of obesity in the US. Using these predictions, the CSINET model estimates that approximately 19,000 (15%) colorectal cancer cases and 9,000 (6.6%) colorectal cancer deaths during the period of 1970-2004 were attributable to the secular trends of BMI.

Conclusion: Population-based cross-sectional surveys such as NHANES can provide estimates of average annual changes in BMI as well as other risk factors for specific cohorts in demographic subgroups. These estimates can form the foundation for public health research on evaluating population-wide policies targeting at health behaviors.
ORAL CONCURRENT SESSION A - PUBLIC HEALTH 1
TENSIONS IN PUBLIC HEALTH POLICY: PREVENTION, CHOICE AND INEQUALITIES IN DECISION MAKING
Murtagh M, Thomson R and Khaw F
University of Newcastle upon Tyne, Newcastle upon Tyne, United Kingdom

Background: Health policy in the UK and internationally has recently emphasized population health improvement and reduction of health inequalities whilst also embracing the challenges of greater public and patient engagement, both in individual treatment decisions and in policy making and priority setting. We argue that there are considerable tensions between these major public health policy initiatives.

Methods: 1) Using social scientific methods of policy analysis we examine evidence-based health policy (exemplified in the UK by National Institute for Clinical Excellence (NICE) guidelines), policy promoting public and patient engagement (through strategies designed to promote patient information and choice and by requirements for health service agencies to engage patients and the public) and the policy goal of reducing health inequalities. 2) To illustrate the population impact of patient choice we model paternalistic and shared decision making in the management of hypertension in a population-based cohort: a random age and sex stratified community sample of residents of south east Northumberland, England aged 65+, from 27 contiguous general practices.

Results: 1) Evidence-based health policy and public and patient engagement. Engaging patients in decisions on their own treatment may not produce the same decisions as would be derived from a more paternalistic or prescriptive application of evidence-based guidelines. If health inequalities and public and patient engagement in decision making. Policy promoting patient choice facilitates the involvement of patients who are already more likely to engage in decision making—younger, well educated patients of higher social class—but may not improve communication with less articulate, less well educated patients. 3) Mathematical modelling of the uptake of anti-hypertensive treatment in paternalistic application of evidence based guidelines and in shared decision making demonstrates that engaging patients in treatment choices may lead to fewer patients taking anti-hypertensive therapy, with a subsequent impact on population incidence of CVD.

Conclusions: We argue that there may well be unintended effects on population health of incongruence in major components of public health policy. Greater public and patient engagement may militate against the disease prevention goals of evidence-based policy. Further, these policies may better engage sections of the population who already benefit from greater access and better health associated with social status and opportunity, serving to increase inequalities by further marginalising those already suffering from relative exclusion.

ORAL CONCURRENT SESSION B - HEALTH SERVICES RESEARCH
THE IMPACT OF PROGNOSIS ON THE TREATMENT OF EARLY STAGE PROSTATE CANCER PATIENTS
Howard D
Emory University, Atlanta, GA

Purpose: Aggressive cancer screening has increased the number of patients diagnosed with localized tumors. Many are elderly and have pre-existing comorbidities. Yet, treatment guidelines are surprisingly vague on how competing risks should factor into treatment choice. Traditionally, concerns about overtreatment in cancer have focused on patients diagnosed with late-stage tumors. The increase in early detection raises the issue in a different guise. The purpose of this study is to examine the degree to which prognosis (i.e., the risk of death from competing causes) influences treatment of patients diagnosed with localized prostate tumors.

Methods: Observational data were used for the study. The analysis has three steps. 1) Weibull model was used to estimate the impact of age and comorbidities on survival time among non-cancer Medicare enrollees in 1992 to allow for 10 years of follow-up (N = 44,800). 2) Parametric estimates from the model were used to predict life expectancy in the absence of cancer for prostate cancer patients diagnosed in 1997-1999 with localized tumors (N = 12,796). This imputed life expectancy measure captures “competing risks.” 3) The association between the imputed life expectancy measure and treatment (prostatectomy versus other) was assessed via logit regression.

Results: Prognosis is a strong predictor of treatment type. A one year increase from the mean value of life expectancy for a patient with a moderately differentiated tumor is associated with a one percentage point increase in the likelihood of surgery (p < 0.01). The relationship is highly non-linear, and it varies in the expected direction by tumor grade. The strength of the relationship does not vary by race or SES.

Conclusion: Findings increase confidence that physicians consider prognosis when prescribing treatment and that aggressive screening is not leading to substantial overtreatment.
ORAL CONCURRENT SESSION A - PUBLIC HEALTH I
THE COST EFFECTIVENESS OF ALTERNATIVE STRATEGIES FOR STOCKPILING AND DISPENSING MEDICAL AND PHARMACEUTICAL SUPPLIES FOR A RESPONSE TO ANTHRAX BIOTERRORISM
Bravo D, Zaric G, Holty F, Brandow M, Wilhelm E, McDonald K and Owens D
Stanford University, Stanford, CA; University of Western Ontario, London, ON, Canada; VA Palo Alto Healthcare System, Stanford, CA

Purpose: There is an ongoing national debate regarding appropriate strategies for regionalization of medical and pharmaceutical supplies for bioterrorism responses. We developed a simulation model to evaluate the health benefits, costs, and cost effectiveness for changes in the levels of local inventories, delays in the arrival of regional inventories, and changes in the local capacity to dispense inventories.

Methods: We simulated an aerosol release of anthrax spores in a US metropolitan area, varying the exposed population from 50,000 to 250,000 and the number seeking prophylaxis from 100,000 to 5 million. Our disease progression model, based on 60 cases of inhalational anthrax, included three disease stages: asymptomatic incubation, prodromal, and fulminant. Based on the literature, the base case assumed attack detection within 48 hours, local inventories containing 64,500 person-days of prophylaxis and 781 person-days of treatment, a regional inventory containing 2,450,000 person-days of prophylaxis and 4,000 person-days of treatment would be available for dispensing 12 hours after attack detection, the local capacity to dispense prophylactic antibiotics was 1,400 individuals/hour, and 64% of individuals were fully adherent with prophylaxis regimens.

Results: We found that mortality was highly dependent on the number of individuals requiring prophylaxis, dispensing capacity, adherence with prophylactic antibiotics, and delays in attack detection. For an attack that exposed 250,000 people and required the prophylaxis of 5 million, the expected mortality fell from 240,575 to 157,738 as the dispensing capacity increased from 1,400 to 22,500 individuals per hour. At low dispensing capacities (base case), nearly all exposed individuals died, regardless of the rate of adherence with prophylaxis. At higher dispensing capacities, the expected mortality was dependent on the rate of adherence with prophylactic antibiotics. There was no benefit to doubling the local inventory at low dispensing capacities; however, at higher dispensing capacities, the cost effectiveness of doubling local inventories fell from $43,628 to $373/QALY as the dispensing capacity increased.

Discussion: The CSDM is a fully-staffed office suite located in the DHMC’s main atrium. Depending on patients’ presenting characteristics, 4 different levels of intensity of decision support are provided. Level 3 involves walk-in referred patients for whom decision support is provided using the Ottawa Personal Decision Guide and condition-specific decision aids (e.g., PSA screening, elec- tronic back surgery, breast or prostate cancer, etc.). These Level 3 patients complete evaluative questions about this decision support service.

Results: To date, over 1,500 patients have received Level 3 support. Age and sex distributions are consistent with those of the underlying patient populations. The majority have high school education. After viewing their condition-specific DA, they report the following patterns. Uncertainty: Fewer patients are unsure of their treatment preferences (21%), compared with before (38%). Confidence: Positive evaluations range from 88% (vs. option’s risks) to 92% (vs. overall comparisons of the options). Values: Positive evaluations range from 74% (vs. options’ risks) to 78% (vs. options’ benefit). Making a Choice: 86-87% report enough support/Free of social pressure; 88% have enough advice; 57% are sure about what to choose, while 32% remain uncertain. Decision Confidence: For the majority, the DAs helped them to organize thoughts (98%); consider pros and cons (92%); iden- tify questions to ask (99%); consider their own involvement in decision making (91%); prepare to make a better decision (88%); Effects on MD-Patient Communication: For the majority, the DAs helped them to know what to expect at their next visit (72%) improve their use of clinic time (69%); make visits smoother (73%); and communicate with their MD (91%), while not negatively affecting the relationship with their MD (89%).

Conclusions: It is possible to conduct quality assurance assessments of a formal decision support service for patients in a busy clinical setting. These assessments indicate a) that the service generally encourages effective decision making in close-call situations in which there is no single “best” option; and b) that there are sub-groups of patients who may benefit from more intensive levels of decision support (e.g., the 32% who are still “uncertain”).
ORAL CONCURRENT SESSION A - PATIENT AND PHYSICIAN BEHAVIOR/ PREFERENCES

1 CAN WE BRIEFCURTHER? EVIDENCE OF LIMITED EXTERNAL VALIDITY OF MODELS MAPPING SF-36 TO HEALTH UTILITY
Ya A, Vu Y and Nichol M
University of Southern California, Los Angeles, CA
Purpose: Many studies have attempted to develop a model mapping health status to utility score based on a sample of population where both measures were obtained. However, evidence have shown different models yielded dramatically different estimated utilities. Can this disagreement be partially explained by differences in characteristics of populations where the models were generated? External validity (generalizability) of the model can be limited when subscale coefficients vary with patient characteristics. The purpose of this study is to examine whether a regression model that bridges from health status to utility is sensitive to patient characteristics.

Methods: Data included 6,923 Southern California Kaiser Permanente members who filled both SF-36 and HUI3 in year 1994-1995. Missing item responses were imputed by MEMC and propensity score method with a missing at random assumption. The linear regression framework mapping from SF-36 subscales to HUI3 utility was used as a base model, which also adjusted for age and gender. Interaction effect between subscales and each of the following ten patient factors was examined: ethnicity, work status, marital status, income level, education, chronic disease score (CDS), diabetes, depression, COPD/asthma and cardiovascular diseases (CVD). To evaluate the individual characterisitic effect, these variables were examined separately. Analysis of variance (F-test) was used to test the significance of all interactions.

Results: Every patient factor examined in this study had significant interaction effect with at least one SF-36 subscale. The effects of some subscales on utility were more likely to be influenced by patient factors. The coefficient of body pain on utility were significantly associated with marital status, education, income and COPD/asthma, role limitation (emotional) with marital status, education, CDS and CVD, physical functioning with work status, CDS, diabetes and depression; mental health with income and ethnicity, role limitation (physical) with COPD/asthma. When all the patient factors were presented in a model allowing full interaction, interactions were highly significant (F = 1.57, P < 0.0001).

Conclusion: This study cast doubts on the existence of a reliable mapping model that can be universally applied to any patient sample. In this study, the mapping model appeared to be strongly influenced by respondent's demographic attributes, which could limit external validity of the model. Therefore, such regression models may not be appropriate to estimate utility in different populations without further adjustment.

ORAL CONCURRENT SESSION B - CLINICAL EFFECTIVENESS AND QUALITY OF LIFE
CLINICAL EFFECTIVENESS AND COST-EFFECTIVENESS OF DISEASE MANAGEMENT PROGRAMS FOR PATIENTS WITH HEART FAILURE
Gohler A1, Dietz R2, Osterziel K1 and Siebert U3
1 Charité Campus Virchow Klinikum, Humboldt Medical School, Berlin, Germany; 2Charité Campus Virchow Klinikum, Humboldt Medical School, Berlin, Germany; 3Massachusetts General Hospital, Harvard Medical School, Boston, MA
Purpose: Congestive heart failure (CHF) is a major cause of morbidity and mortality in the population. Currently causing 48% of hospitalization of the elderly and 1-2% of the annual health care expenses it is likely to escalate in the next decades. In a meta-analysis of 16 randomized controlled trials investigating disease management programs (DMP) in the treatment of CHF, we could show a statistically significant reduction in mortality and rehospitalization, but cost-effectiveness of DMPs remains uncertain. Therefore, we sought to evaluate life expectancy and long life medical costs for DMPs.

Methods: Design: Cost and cost-effectiveness analysis using a 6 state Markov Model represent- ing the number of prior hospitalizations (1 in 8-4 in 8 and death). Data sources: Pooled efficacy data from our meta-analysis of randomised clinical trials. SOMED registry data for age-dependent hospi- talizations and mortality rates adjusted for additional benefit from beta-blocker therapy and reim- bursement costs in the Australian health care system. Target population: All patients who have been admitted with severe heart failure. Time horizon: Lifetime. Perspective: Societal. Intervention: Conventional therapy and DMP. Outcome measures: Life years gained and lifetime direct medical costs.

Results: For a population aged 73 at onset of CHF (27% female, 13% on betablocker), our model yielded, on average, a remaining life expectancy of 3.24 years for conventional treatment and 3.38 years for DMP. Mean undiscounted lifetime costs per patient were estimated at EUR 11,600 and EUR 12,700 respectively. The discounted incremental cost-effectiveness ratio (ICER) of DMP vs. conven- tional care was EUR 8,813 per life-year-gained (LYG). Assuming the benefit due to DMP lasting for 5 years after the end of the actual intervention would lead to additional 1.5 life months and reduce ICER to 4,021 EUR/LYG.

Conclusion: Based on our decision analysis, DMPs prolong life, but increase life-time costs. A cost-saving effect of DMPs (i.e., more effective and less costly) as suggested in some original studies could not be confirmed in our decision analysis. However, even under conservative assumptions regar- ding the duration of DMP, these programs are cost-effective when compared to other well-accepted medical interventions in heart disease.

ORAL CONCURRENT SESSION B - CLINICAL EFFECTIVENESS AND QUALITY OF LIFE
COST-EFFECTIVENESS OF IMATINIB IN PATIENTS NEWLY DIAGNOSED WITH CHRONIC MYELOID LEUKEMIA
Reed S1, Amstrong T2, Ludmer J3, Gledemning G4 and Schulman K5
1Duke Clinical Research Institute, Durham, NC; 2Novartis Pharmaceuticals Corporation, East Hanover, NJ
Purpose: The International Randomized Study of Interferon-α vs. STI571 (imatinib) (IRIS) demonstr- ated significantly improved clinical measures among patients newly diagnosed with chronic myeloid leukemia (CML) in the chronic phase who were initially randomized to imatinib relative to those initially randomized to interferon-α plus low-dose cytarabine (IFN-α + LDAC). The aim of our study was to estimate the incremental cost per life-year saved (LYS) and the incremental cost per quality-adjusted life-year (QALY) of first-line treatment with imatinib relative to IFN-α + LDAC in chronic phase CML patients.

Methods: A simulation model was developed to incorporate clinical resource utilization and utility data collected in IRIS with data from the literature to estimate lifetime costs, survival and quality adjusted survival. The model was designed so that patients initially treated with imatinib could switch to IFN-α + LDAC, and then to hydroxyurea. Patients initially treated with IFN-α + LDAC could switch to hydroxyurea. Long-term survival estimates were based on proportional haz- ard relationships estimated between age- and gender matched individuals from the general popula-
tion and historical CML patients who attained or did not attain complete cytogenetic response on treatment with IFN-α. The simulation model incorporated first and second order uncertainty. Probabilistic and traditional sensitivity analyses were undertaken to evaluate uncertainty.

Results: In the base-case analysis, undiscounted mean survival was estimated at 13.30 years for patients receiving first-line imatinib and 9.07 years for patients receiving first-line IFN-α + LDAC. Undiscounted lifetime costs were estimated at approximately €48,500 for imatinib-treated patients and €162,800 for IFN-α + LDAC-treated patients. With an annual discount rate of 3%, the incremental gain in survival among patients treated with imatinib was estimated at 3.91 life years (LY) and QALYs. The incremen- tal cost-effectiveness ratio (ICERs) were estimated at €43,100/LYS (95% CI: 37,600 to 51,100) and €43,300/QALY (95% CI: 38,300 to 48,900). The model’s results were most sensitive to changes in assumptions affecting the relative duration or costs of the treatments. Because lifetime costs closely tracked with survival, the ICERs were relatively consistent when varying assumptions used in sensitivity estimation.

Conclusion: When used as first-line therapy for newly diagnosed CML patients, imatinib ap- pears to be an economically attractive therapy relative to first-line treatment with IFN-α + LDAC.
ABSTRACTS

ORAL CONCURRENT SESSION A - PATIENT AND PHYSICIAN BEHAVIOR/ PREFERENCES 1

VARIATIONS IN PHYSICIANS’ TREATMENT RECOMMENDATIONS FOR THE MANAGEMENT OF ENDOMETRIOSIS PAIN

Ankis S1, Llewellyn Thomas H1, Testosson A, Weinstein M, Hornstein M, Laufer M and Kuntz K4

1Stanford University, Stanford, CA; Dartmouth Medical School, Hanover, NH; Dartmouth Medical School, Lebanon, NH; Harvard School of Public Health, Boston, MA; 2Brigham and Women’s Hospital, Boston, MA

Purpose: Several treatment alternatives for endometriosis pain are similar in effectiveness but differ considerably in their risk profiles and quality of life outcomes. The objective was to determine the distribution of physicians’ treatment recommendations for endometriosis pain and how they vary by physician characteristics.

Methods: We conducted a national mailed survey of a random sample of gynecologists, consisting of generalists (n = 112) and endometriosis specialists (n = 248). The questionnaire presented a hypothetical patient with confirmed endometriosis whose pain was not relieved by first-line therapies had failed. Respondents ranked eight treatments from the most to the least preferred and provided a rating on how likely they were to recommend each alternative (on a 0–5 point scale). We summarized the rank orderings and calculated the percentage of respondents who provided relatively high recommendation ratings ("probably" or "definitely" recommended) for each treatment. Logistic regression analysis of the odds of a high recommendation rating was used to explore associations between recommendations and physician characteristics (generalist/specialist, sex, age, race, U.S./foreign medical school, geographic region). Each model’s ability to discriminate between high versus low recommendation was characterized by the area under the receiver operating characteristic curve (C-statistic).

Results: The percentage of respondents providing a high recommendation rating (ranking treatment as most preferred) were as follows: GnRH agonist 83.3% (95% CI, 81.7–85.0%), laparoscopy 4.4% (23.6%), progesterin 34.5% (5.4%), ‘other’ 30.6% (9.7%), alternative therapies 16.3% (4.0%), danazol 12.3% (1.9%), pain treatment center 11.9% (1.9%), hysteroscopy 2.1% (0.7%). Specialists were significantly more likely than generalists to recommend laparoscopy (OR = 2.0; p = 0.01), alternative therapies (OR = 2.0; p = 0.03), and ‘other’ treatments (12.2%; p = 0.03). Female physicians were more likely to recommend GnRH agonist (OR = 2.0; p = 0.08), progesterin (OR = 1.7; p = 0.05), alternative therapies (OR = 2.0; p = 0.03), and ‘other’ treatments (12.2%; p = 0.03). There were no significant associations between treatment recommendations and race, medical school, or geographic region. The accuracy of the eight models in discriminating between high versus low recommendation (C-statistics) ranged from 0.61 for progesterin to 0.76 for ‘other’ treatment.

Conclusions: We found substantial variations in physicians’ treatment recommendations for endometriosis pain. This highlights the need for significant physician-related effects on recommendations, suggesting the possibility of unwarranted sources of variation.

ABSTRACTS

ORAL CONCURRENT SESSION A - PATIENT AND PHYSICIAN BEHAVIOR/ PREFERENCES 1

LUNG TRANSPLANT CANDIDATES’ ESTIMATION OF POST-TRANSPLANT UTILITIES

Singh I1 and Gordon C4

1University of Toronto, Toronto, ON, Canada; 4University Health Network, Toronto, ON, Canada

Purpose: To determine how accurately lung transplant candidates envisage transplant recipient’s visual analog scale (VAS) and standard gamble (SG) utilities, and how these perceptions influence readiness for transplant listing. Methods: VAS and SG utilities for current health were elicited from 71 transplant candidates (median age 54, 57% male) and 99 transplant recipients (median age 53, 49% male). Additionally, candidates completed VAS and SG utilities imagining both that they had received a transplant with a good outcome (no BOS) and that they had developed chronic rejection, called bronchiolitis obliterans syndrome (BOS). Candidates who predicted higher scores for BOS than no BOS were excluded from the analysis, leaving 64 VAS and 59 SG scores. Candidates also indicated readiness for transplant listing on a Likert scale. Participants were separated into those with BOS (n = 23) and those without BOS (n = 78) for derivation of actual utility values for these health states. Two-sample Wilcoxon rank-sum tests were used to compare candidates’ predicted no BOS and BOS scores with recipients’ actual no BOS and BOS scores respectively for both VAS and SG utilities. Fisher’s exact test was used to compare readiness for transplantation between patient groups.

Results: Candidates overestimated recipients’ actual VAS scores for no BOS (p = 0.03), although the difference was not clinically significant (median for both = 80). Candidates accurately estimated SG utilities for no BOS (median 0.91 for candidates and 0.94 for recipients, p = 0.83). Candidates significantly underestimated recipient’s actual VAS (median 40 for candidates and 75 for recipients, p = 0.00005) and SG (median 0.57 for candidates and 0.93 for recipients, p = 0.00005) scores for BOS. Candidates with current health SG scores exceeding predicted BOS scores were less likely to indicate readiness for transplant listing than those whose SG scores for current health were lower than predicted BOS scores (52% vs. 88% p = 0.015). VAS scores followed a similar pattern.

Conclusions: Lung transplant candidates accurately estimated VAS and SG utilities for no BOS, while significantly underestimating VAS and SG utilities for BOS. Candidates who predicted BOS to be worse than their current health were less likely to be ready for transplant listing. Some medically acceptable candidates decline transplant listing. These results suggest that inaccurate perceptions of post-transplant outcomes for BOS may be a contributing factor for those not ready to be listed.

Results: At a willingness-to-pay threshold of $150,000/QALY, CTA yielded the highest net health benefit, which was only 1 quality-adjusted life-year (QALY) greater than DSA—the next best strategy. The total EVPI was 2.5 QALY per patient. This implies that resolving all decision uncertainty is expected to improve net benefit with on average 2.5 QALD greater than DSA—the next best strategy. Our findings have important implications for medical decision making in cancer and coronary artery disease for which clinical future research is justified.

Methods: We developed a probabilistic Markov model comparing cost-effectiveness from the health-care system perspective for four diagnostic tests for the diagnosis of coronary artery disease: exercise echocardiography, exercise single-photon-emission computed tomography (SPECT), computed tomographic angiography (CTA), and digital subtraction angiography (DSA). The expected value of perfect information (EVPI) was estimated per patient and subsequently for the entire USA, assuming an annual patient population of 250,000 for a duration of 5 years. Finally, the value of obtaining more information for particular sets of parameters was estimated. We considered 12 cohorts of patients defined by age, sex, and type and severity of chest pain.

Conclusion: The willingness-to-pay threshold of $150,000/QALY, CTA yielded the highest net health benefit, which was only 1 quality-adjusted life-year (QALY) greater than DSA—the next best strategy. The total EVPI was 2.5 QALY per patient. This implies that resolving all decision uncertainty is expected to improve net benefit with on average 2.5 QALD greater than DSA—the next best strategy. Our findings have important implications for medical decision making in cancer and coronary artery disease for which clinical future research is justified.
ABSTRACTS

ORAL CONCURRENT SESSION A - PATIENT AND PHYSICIAN BEHAVIOR/PREF- ERENCES I

WOMEN’S DECISION MAKING ROLES REGARDING PROPHYLACTIC MASTECTOMY
Nehkvudov L 1 and Geiger A 2
Harvard Medical School/Harvard Pilgrim Health Care, Boston, MA; 3Southern Californ ia Kaiser Permanente, Pasadena, CA

PURPOSE: Contralateral prophylactic mastectomy (CPM) is the removal of a non-affected breast in a woman with unilateral breast cancer and is an option for women considering methods to prevent the risk of breast cancer. Yet little is known about a woman’s role in decision making regard ing CPM and its effect on long term outcomes.

METHODS: We mailed surveys to 766 women aged 18-80 with contralateral prophylactic mastectomy performed at one of six health maintenance organizations between 1979 and 1999. The survey included measures of decision making control, satisfaction, concern about breast cancer, and depressive disorders (CES-D).

RESULTS: We received 562 surveys (response rate 73%). 502 analyzable surveys from women without recurrent breast cancer are included in this analysis. Respondents’ median age was 49 years and duration since the procedure was 10 years. Most women (97%) reported active roles in decision making regarding CPM but their extent of involvement varied; 42% reported making the final decision themselves (ALONE), 39% made the final decision after considering the doctor’s opinion (OPINION), and 15% shared the decision making responsibility with their providers (SHARED).

Women with the greatest involvement (ALONE) were more likely to have college education compared to those with OPINION and SHARED roles (51%, 45%, 19% respectively, p < 0.001) but were similar by race, age, breast cancer risk factors and concerns about breast cancer. While most women were satisfied with CPM within six months of the procedure, those reporting ALONE and OPINION roles were more likely to be satisfied than those with SHARED roles (68%, 49%, 70% respectively, p<0.04; adjusted for education). Decision making roles did not predict long-term satisfactions change in concern about breast cancer or current depressive disorders.

CONCLUSIONS: Among women with active decision making roles, those with higher education reported greater involvement in decision making regarding CPM and short term satisfactions following the procedure. Women with lower education may need additional support to assume more active roles to achieve similar short term outcomes though the level of involvement does not appear to have long term implications.

ORAL CONCURRENT SESSION A - PATIENT AND PHYSICIAN BEHAVIOR/PREF- ERENCES II

UTILITY-BASED ASSESSMENTS OF QUALITY OF LIFE IN A RANDOMIZED TRIAL OF ANTIRETROVIRAL THERAPY IN ADVANCED HIV DISEASE
Iverson V 1, Araki S 2, Sundaram V 3, Hill A 4, Baker F 4, Munakata J 4, Kyriakides T 4, Sanders GA 5, Griffin S 5, Holodny M 5, Brown S 5, Nieas RP 5, Sculptor M 5, Azis A 6, Wu Y 6, Owens D 7 andInvestigators t 8
1VA Palo Alto Health Care System, Menlo Park, CA; 2Stanford University, Stanford, CA; 3VA Palo Alto Health Care System, Palo Alto, CA; 4VA Connecticut Health Care System, West Haven, CT; 5Duke, Durham, NC; 6University of York, York, United Kingdom; 7Bronx VA Medical Center, Bronx, NY; 8Express Scripts, St. Louis, MO 3University of British Columbia, Vancouver, BC, Canada.

Purpose: Advances in antiretroviral therapy (ART) have dramatically increased survival with HIV. However, the quality of life of patients who have been treated with multiple regimens has not been well studied. We assessed health related quality of life associated with advanced HIV in patients enrolled in a tri-national ART-management trial.

Methods: The OPTIMA (Options in Management with Antiretroviral therapy) trial enrolls patients with plasma viral load (PVL) < 2,000 copies/ml, a CD4 count > 300/mcml, and who are currently on ART. We assessed US baseline quality of life assessments (n = 1231) for current health using time tradeoff (TTO), standard gamble (SG), the Health Utilities Index 2 and 3 (HUI2, HUI3), the Euro-Qol (EQ-5D), and the visual analog scale (VAS). We stratified patients (n = 219) by median PVL (59,614 copies/ml) and CD4 count (128/mm3), and each group’s means were compared utilizing Wilcoxon-Mann-Whitney tests. Correlations and differences across instruments were assessed using Pearson correlation coefficients and Wilcoxon signed rank sum tests.

Results: Baseline mean utilities for US patients, in descending order, were TTO 0.82, SG 0.76, HUI2 0.73, EQ-5D 0.71, VAS 0.68, and HUI3 0.60. The distribution of such instruments’ scores appears to be unimodal with scores concentrated near the healthier end of the scale (suggesting a possible ceiling effect). Mean preferences were generally higher in patients with lower PVL (higher CD4). However, only the EQ-5D and HUI3 revealed a significant difference between patients below and above the median PVL (0.74 vs. 0.67, p = 0.02, 0.65 vs. 0.55, p = 0.05). Stratified by low and high CD4 count, the EQ-5D yielded the sole significant difference (0.67 vs. 0.74, p = 0.05). Correlations between instruments ranged from r = 0.69 (SG & HU12) to r = 0.87 (HUI2 & HUI3). All were statistically significant (p < 0.05) except for one pair (SG & HU52). All mean utility scores between instruments were significantly different (p < 0.01), except between SG & HU12, SG & EQ-5D, and HU12 & EQ-5D.

Conclusions: Patients with advanced HIV disease who have failed multiple ART regimens have a substantially reduced quality of life as assessed by utility-based instruments. In this group, immunologic status did not strongly influence quality of life. Assessments of quality of life vary significantly among instruments, thus highlighting the importance of the choice of methods used for preference assessment.

ORAL CONCURRENT SESSION B - CLINICAL EFFECTIVENESS AND QUALITY OF LIFE A VALIDATIONAL ANALYSIS OF UTILITY ELICITATION METHODS
Nath M 1
State University of New York at Albany, Parker, CO

Purpose: The purpose of this study was to examine the convergent and divergent validity of three utility elicitation methods used commonly in medical decision making: the standard gamble, time tradeoff and visual analog scale.

Methods: A multitrait multimethod (MTMM) matrix was used for this validation exercise. Three health states—blindness, AIDS and stroke—were used for the MTMM exercise. In addition, the relationship between specific individual differences (construct irrelevant) variables and utility responses was tested using a structural model. 101 subjects provided utilities using the (IMPACT3 protocol. Data were also collected about individual differences such as numeracy, time preference and risk preference using several attitudinal instruments.

Results: The MTMM analysis showed the methods achieving convergent validity. In other words, the different methods converged on a given health state. However, the methods failed to achieve divergent validity, i.e., given different health states, methods were unable to discriminate between the health states. This research was the first to study construct irrelevant variables—such as numeracy, experience with illness and risk preference (including several types of risk)—cited in the literature as likely to affect responses to utility elicitation methods. A linear structural model was built. The model achieved a good fit and showed that physical risk was positively associated with Standard Gamble utilities. The model also showed that professional risk was negatively associated with Standard Gamble and Time Tradeoff utilities.

Conclusion: This is the first empirical study to examine the convergent and divergent validity of utility elicitation methods and it finds that the methods fail to achieve divergent validity. The adequacy of convergence and the lack of divergence achieved by the methods point to two important conclusions. Firstly, that the debate as to the usefulness of these methods has been unhelpful perhaps, because of this duality in the nature of these methods. Secondly, that method variance is a predominant characteristic of utility elicitation further strengthening the notion that utilities might be constructed rather than elicited by these methods. This is also the first study to examine the relationship between the different utilities and the relationship between each of the utilities and the reference method. This study highlights the need to study these subtypes of risk more closely.

ORAL CONCURRENT SESSION B - CLINICAL EFFECTIVENESS AND QUALITY OF LIFE REFUSALS TO MAKE TRADE-OFFS: TASK CONTEXT AND EMOTION MATTER
Namasivayam L 1, Zikmund-Fisher B 2, Roberts T 3 and Ubel P 4
VA Ann Arbor Healthcare System, Ann Arbor, MI; 5University of Michigan, Ann Arbor, MI; 6University of Michigan, Ann Arbor, MI; 7VA Ann Arbor Healthcare System, Ann Arbor, MI; 8University of Michigan, Ann Arbor, MI; 9Duke, Durham, NC; 10Vanderbilt University, Nashville, TN

Purpose: Person trade-off (PTO) elicitation measure people’s preferences by asking them to choose between different health care treatments. For example, people might be asked how many patients need to be cured of chronic pain to bring as much benefit as curing 100 patients of paraplegia. Many people refuse to make trade-offs in PTO elicitations for reasons that are not well understood. In this study, we test whether refusals to make trade-offs are caused by people protesting against the “need” to make such rationing choices.

Methods: We explored whether willingness to make trade-offs would increase and decrease across the task. The “need” to make such rationing choices would become more likely due to changes in the “need” to make such rationing choices.

Results: 1,754 subjects participated in our randomized trial via the Internet. As expected, subjects who expressed more outrage about their task were more likely to refuse to trade-off, regardless of task type. In addition, as expected, subjects given the rationing task reported more outrage than subjects given the non-rationing task (p < 0.0002). Surprisingly, however, subjects given the non-rationing task had nearly six times higher odds for refusing to trade-off compared to those receiving the rationing task (p < 0.0001). Also, subjects with low subjective numercy and those who thought the survey was relatively easy had higher odds for refusing to make trade-offs (p<0.004).

Conclusion: Relieving people of the need to make rationing decisions reduced the level of outrage associated with PTO elicitation, but at the same time made them even less willing to make trade-offs. Across all subjects, level of outrage was associated with unwillingness to make trade-offs. Given that unwillingness to make trade-offs plagues other preference elicitation, including standard PTO elicitations, further research is needed to clarify why people refuse to make trade-offs and what should be done to rectify this problem.

E8 • MEDICAL DECISION MAKING/JAN–FEB 2005
ABSTRACTS

ORAL CONCURRENT SESSION A - JUDGMENT AND DECISION MAKING
THE GIST OF RISKS AND BENEFITS IN ADOLESCENT DECISION-MAKING:
EFFECTS OF AGE, GENDER, ETHNICITY, AND EXPERIENCE

Revina V
Metta K
Poirier K
Roberts J
Adam M
and Loccy C

University of Texas at Arlington, Arlington, TX; University of Arizona, Tucson, AZ.

Purpose: To investigate decision-making factors that reduce HIV transmission and premature pregnancy in adolescents, specifically, to investigate theoretical predictions about perceptions of the gist of risks and benefits, as well as effects of age, gender, ethnicity, and sexual history on such gist representations.

Methods: 215 male and female adolescents (37% Anglo, 45% Hispanic, and 18% mixed/other) aged 14-19 years were each administered a control sexual decision-making scenario (male and female characters in the scenario had been abstinence) and two scenarios from the following factorial design: (1) male vs. female had been sexually active X (2) protected vs. unprotected sex X (3) one prior partner vs. many prior partners. A variety of responses were elicited, including ratings of risks and benefits, and gist representations of "medium," "high," and "extreme" levels of sexual risk.

Results: For identical scenarios, as predicted, the gist of perceived benefits differed by gender, ethnicity, and sexual history of subjects. Also as predicted, perceived risks differed by age and objective risk factors (such as protected/unprotected sex or number of partners). Male subjects perceived more benefits than female subjects and male characters were perceived to benefit more from sex, especially among Hispanics. Previously sexually active subjects perceived more benefits for characters in the scenarios, but similar risks, compared to abstinent subjects. Age differences in risk perception emerged first for small, categorical differences in risk, as predicted by fuzzy-trace theory.

Conclusions: Theoretical predictions originally tested in abstract laboratory tasks were extended to adolescents' risky sexual decisions. Data confirmed predictions that gender, ethnicity, and experience shape the perception of identical decision options and that increases in risk perception with age were associated with developmental shifts in reliance on gist-based (categorical) decision processes, in accordance with fuzzy-trace theory.

ORAL CONCURRENT SESSION B - METHODOLOGICAL ADVANCES
LOCAL CONTROL VERSUS COVARIATE ADJUSTMENT IN ASSESSING TREATMENT EFFECTS IN CLINICAL TRIAL DATA

Johnston J
Obenchain R
Rosensap M
Hornung K
and Eckman M

Eli Lilly, Indianapolis, IN; University of Cincinnati, Cincinnati, OH.

Purpose: To explore the relative merits of using a Local Control (LC) approach based upon patient clustering rather than traditional smooth regression models for Covariate Adjustment (CA) in assessing treatment effects.

Methods: We used patient data (n=1658) from a randomized placebo-controlled trial of drotrecogin alfa (activated) (DAA) for the treatment of severe sepsis. For simplicity, we selected as covariates three patient characteristics found to be the important independent clinical predictors of mortality: age, APACHE II acute physiology score, and platelet count. In the LC approach, patients' baseline values for these three continuous variables were used to determine their location in 3-dimensional X-space, and dissimilarities between patients were determined by calculating Mahalanobis distances, which were then used in clustering patients. A range of estimates of treatment effect were generated by varying the size (and number) of clusters and measuring the weighted average within-cluster difference in survival across all clusters containing both treated and untreated patients. These estimates were then compared with those obtained for CA using logistic regression.

Results: For identical randomization, small treatment group differences on important baseline predictors of mortality: age, APACHE II acute physiology score, and platelet count. In the LC approach, patients' baseline values for these three continuous variables were used to determine their location in 3-dimensional X-space, and dissimilarities between patients were determined by calculating Mahalanobis distances, which were then used in clustering patients. A range of estimates of treatment effect were generated by varying the size (and number) of clusters and measuring the weighted average within-cluster difference in survival across all clusters containing both treated and untreated patients. These estimates were then compared with those obtained for CA using logistic regression.

Conclusions: Theoretical predictions originally tested in abstract laboratory tasks were extended to adolescents' risky sexual decisions. Data confirmed predictions that gender, ethnicity, and experience shape the perception of identical decision options and that increases in risk perception with age were associated with developmental shifts in reliance on gist-based (categorical) decision processes, in accordance with fuzzy-trace theory.

ORAL CONCURRENT SESSION B - METHODOLOGICAL ADVANCES
THE VALUE OF IMPLEMENTATION AND THE VALUE OF INFORMATION:
COMBINED AND UNSEEN DEVELOPMENT

Fawcett E
Claxton K
and Sculptor M

University of York, York, United Kingdom.

Purpose: To explore the relative merits of using a Local Control (LC) approach based upon patient clustering rather than traditional smooth regression models for Covariate Adjustment (CA) in assessing treatment effects.

Methods: We used patient data (n=1658) from a randomized placebo-controlled trial of drotrecogin alfa (activated) (DAA) for the treatment of severe sepsis. For simplicity, we selected as covariates three patient characteristics found to be the important independent clinical predictors of mortality: age, APACHE II acute physiology score, and platelet count. In the LC approach, patients' baseline values for these three continuous variables were used to determine their location in 3-dimensional X-space, and dissimilarities between patients were determined by calculating Mahalanobis distances, which were then used in clustering patients. A range of estimates of treatment effect were generated by varying the size (and number) of clusters and measuring the weighted average within-cluster difference in survival across all clusters containing both treated and untreated patients. These estimates were then compared with those obtained for CA using logistic regression.

Results: For identical randomization, small treatment group differences on important baseline predictors of mortality: age, APACHE II acute physiology score, and platelet count. In the LC approach, patients' baseline values for these three continuous variables were used to determine their location in 3-dimensional X-space, and dissimilarities between patients were determined by calculating Mahalanobis distances, which were then used in clustering patients. A range of estimates of treatment effect were generated by varying the size (and number) of clusters and measuring the weighted average within-cluster difference in survival across all clusters containing both treated and untreated patients. These estimates were then compared with those obtained for CA using logistic regression.

Conclusions: Theoretical predictions originally tested in abstract laboratory tasks were extended to adolescents' risky sexual decisions. Data confirmed predictions that gender, ethnicity, and experience shape the perception of identical decision options and that increases in risk perception with age were associated with developmental shifts in reliance on gist-based (categorical) decision processes, in accordance with fuzzy-trace theory.

ORAL CONCURRENT SESSION A - JUDGMENT AND DECISION MAKING
FORMER PATIENTS GIVE LOWER UTILITY RATINGS FOR COLUMSTOY THAN DO CURRENT PATIENTS: EVIDENCE FOR A THEORY DRIVEN RECALL BIAS

Sheriff E
Ubel P
and Smith D

University of Michigan, Ann Arbor, MI.

Purpose: The public often underestimates patients' mood and quality of life, and gives lower utility ratings than do patients. Cologstomy patients provide a unique opportunity to further investigate this discrepancy since many have their colostomies reversed and their bowel function restored. This study examined the difference in quality of life, mood and utility ratings between current and former colostomy patients to determine if a discrepancy similar to that found between patients and the public would still exist.

Methods: 330 patients who had undergone colostomy surgery within the past five years were mailed a survey including measures of subjective well-being, mood, quality of life, and a utility measure: the time tradeoff (TTT). The TTT exercise asked people to imagine that they have ten years remaining in their lives and then asked how much of that time they would give up to rid themselves of a colostomy. 96 subjects who still had ccolostomies completed the survey, as did 100 subjects who had had their colostomies reversed.

Results: Measures of quality of life, subjective well-being and mood were nearly identical for the two groups (all p's > 0.1). However, in the TTT exercise subjects who still had colostomies were willing to give up 18 of their 120 months of life to have normal bowel function restored, while subjects who had had their colostomies reversed were willing to give up 44 months of life on average (p < 0.01). In addition, the former patients, as compared to the current patients, reported that “having normal bowel function” was much more important to them (p < 0.001). Higher values on this item predicted lower utility values (p < 0.001).

Conclusions: Although colostomies do not appear to significantly affect current well-being, it seems that people who no longer have their colostomies believe that the experience is much worse than the experience reported by current colostomy patients, as evidenced by their markedly lower utility ratings. This finding mirrors the typical patient/public discrepancy, but is striking in that the “public” in this case has experienced the patient condition. Rather than accurately remembering their experiences with colostomies, these former patients may have formed a theory that colostomies negatively influenced their lives more than in actuality; this is consistent with other research on a “theory driven recall bias.”

ABSTRACTS
ORAL CONCURRENT SESSION A - JUDGMENT AND DECISION MAKING
DIFFERENTIAL EFFECT OF A BREAST CANCER DECISION AID BY MARRITAL STATUS
Hoss R, Barnato A and Peelo P

University of Pittsburgh School of Medicine, Pittsburgh, PA; University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA.

Purpose: To evaluate the effect of a breast cancer treatment decision aid (DA) on a woman's perceived decision satisfaction, knowledge, uncertainty, and regret.

Methods: 14 oncology practices were randomized to have their consenting breast cancer patients receive either an informational pamphlet about adjuvant therapy (AT) (N = 166 women) or an evidence-based, risk-tailored DA (N = 226 women). Women were recruited upon presentation to the oncologist for discussion of AT after primary surgical treatment. A health educator administered the intervention (pamphlet or DA) after the physician's history and physical, but before completion of the oncology consultation. Immediately after the consultation and 3 months later patients completed an interviewer-administered questionnaire. From these data we used principal components analysis to develop scales for decision satisfaction (7 items, a = 0.933), subjective knowledge of treatment risks and outcomes (5 items, a = 0.790), and decision uncertainty (5 items, a = 0.790) at the time of the initial consultation and treatment decision, and for decision regret 3 months later (5 items, a = 0.850). We used hierarchical ordered logistic regression (patients clustered within oncology practices) to evaluate the effect of the DA on each decision outcome rounded to the nearest integer, controlling for patient demographics, decision making style preference (independent, collaborative, deference to physician), and breast cancer severity. Results: The study acceptance rate was 85%. The 386 women had a mean age of 51.8 years, 81% were white, 37% single (widowed, divorced, or separated), and 65% had low breast cancer severity. As reported elsewhere, the DA resulted in a very large reduction in the use of AT among women with low severity breast cancer. Based on a scale of 1 (strong agreement) to 5 (strong disagreement), the women had high decision satisfaction (1.7) and subjective knowledge (1.8), moderate uncertainty (3.4), and low regret (4.4). The DA had no effect upon these outcomes among married women, but improved satisfaction (p = 0.001) and subjective knowledge (p = 0.024), and decreased uncertainty (p = 0.001) among single women, without affecting regret.

Conclusion: Despite having a large effect on actual treatment decisions among all women, the DA improved subjective measures of decision quality only among single women. This unexpected finding deserves further exploration.

ORAL CONCURRENT SESSION B - METHODOLOGICAL ADVANCES
QUALITY-ADJUSTED YEARS OF LIFE GAINED FROM IMPLANTABLE DEFFIBRILLATOR: COMPARISON OF A NEW METHOD WITH TRADITIONAL ANALYSES IN CLINICAL TRIALS
Novas K, Dick A, Zhao H, Corona E, Hall W, Zorantinger J and Moss A

University of Rochester School of Medicine, Rochester, NY; University of Illinois at Chicago School of Public Health, Chicago, IL.

Purpose: Traditionally published QALY analyses of clinical trials are analyzed using either change in health-related quality of life (HRQOL) from baseline that includes survivors only, not accounting for data censoring, and does not consider intermittent outcomes, on each domain (ICD-control) = 0.054, p = .48.

Conclusions: Despite prolonging survival of severely ill cardiac patients, ICD implantation was associated with a similar change in HRQOL over 3 years compared with patients undergoing medical treatment. Using quality-adjusted survival analysis in the trials with high differential mortality permits evaluation of the intervention effect on all domains of patient health.

ORAL CONCURRENT SESSION A - JUDGMENT AND DECISION MAKING
A DECISION ANALYSIS BALANCING SURVIVAL AND CISPLATIN OTOTOXICITY IN CHILDREN WITH GERM CELL TUMORS
Lynn W, Womer R and Silber J

University of Pennsylvania School of Medicine, Philadelphia, PA; The Children’s Hospital of Philadelphia, Philadelphia, PA.

Objective: In a decision analytic model, we asked parents and childhood cancer survivors whether an increased survival rate in germ cell tumors was judged worth the increased risk of sustaining permanent hearing loss due to cisplatin chemotherapy.

Methods: A decision model was constructed to reflect the therapeutic choices faced by families with children diagnosed with germ cell tumors in the Children’s Cancer Group treatment protocols 8882 and 8891. Treatment options included different dose schedules of cisplatin, their associated event-free-survival rates and their corresponding risks of hearing loss. The decision model also accounted for adverse events such as renal, salvage therapies, and the resulting additional toxicity. Three vignettes of hearing loss were constructed to describe outcomes of surviving cancer with mild, moderate, and moderately severe high-frequency sensorineural hearing loss. Standard Gamble utility scores were estimated for the 3 vignettes from 31 parents of childhood cancer survivors and from 17 teenagers and young adults who had survived childhood cancer (age 13 - 21) were assessed.

Results: Based on the preference utility scores, 28 out of the 31 parents tested (90%) would recommend the high-dose regimen, while only 10 out of the 17 childhood survivors tested (59%), p = 0.015 suggested high-dose treatment. Further analyses of the SR utility scores revealed that this difference was due to parents’ assessment of cancer survival with sensorineural hearing loss as management goals worth the risk of death in average utility (0.60). Instead, parents who survived cancer deemed such outcomes significantly detrimental and were willing to take a risk of death to mitigate the possibility of hearing loss (average SG utility 0.62, significantly different from parents, p < 0.01). The differences in model-recommended decisions were not affected by the respondents’ numerical and/or their propensity to exhibit omission bias.

Conclusion: As childhood cancer survivors enter adulthood and take charge of their own health, some would retrospectively make different treatment decisions than did their parents. This insight might help to promote a better therapeutic decision making process in children with cancer, and may aid our understanding of the complex ethical issues surrounding such treatment decisions.

ORAL CONCURRENT SESSION B - METHODOLOGICAL ADVANCES
VARIABLE CORRELATION SUBSTANTIALLY INFLUENCES UNCERTAINTY AND VALUE OF INFORMATION ESTIMATION IN PROBABILISTIC SENSITIVITY ANALYSIS
Goosereiel D, Glick H and Schwartz J

Philadelphia VA Medical Center, Philadelphia, PA; University of Pennsylvania, Philadelphia, PA.

Purpose: Probabilistic sensitivity analyses have frequently been used to characterize uncertainty in cost-effectiveness model results. However, other than variable correlation inherently introduced by Markov processes, correlations between the input variables are rarely explicitly included in such simulations, particularly for models that incorporate estimates obtained from different sources. We hypothesized that varying assumptions about the underlying correlation among input variables might influence the results of probabilistic sensitivity analyses.

Methods: We created a cost-effectiveness Markov model comparing two hypothetical treatment strategies (treatments A and B), each with an associated annual probability of death and an annual treatment cost. We designated lognormal distributions for costs and logistic distributions for mortalities probabilities with fixed parameters, and we generated 1000 samples for each input variable to represent typical data available for a model. Bootstrap samples from these populations were used to estimate distributional parameters for cost and effectiveness. We then performed Monte Carlo simulations in which different degrees of correlation were assumed between the four patient-level cost and effectiveness model inputs. Two hundred groups of 1000 first-order simulations were generated for each set of correlation assumptions. We subsequently calculated the probability that each strategy was truly the optimal alternative at selected cost-effectiveness thresholds (λ). We also calculated the expected value of perfect information (EVPI) for each threshold.

Results: Probabilistic sensitivity analyses produced cost-effectiveness acceptability and EVPI curves that varied substantially depending on the underlying correlation structure. If costs and effectiveness were assumed to be uncorrelated, the probability that treatment B was optimal at λ = $100,000 per quality-adjusted life-year (QALY) was 0.77, and the EVPI was $10,353 per patient. However, if costs and benefits for treatment A were highly correlated, the probability that treatment A was “correct” at λ = $100,000/QALY was 0.60. In both cases, improving correlation between costs and benefits greatly reduced the EVPI. If benefits and costs of both treatments were highly correlated both within and between treatments, then the probability treatment B was optimal at λ = $100,000/QALY was 0.36, and the EVPI was $18,063 per person.

Conclusion: The underlying correlation structure among costs and outcomes in cost-effectiveness analyses can profoundly influence the results of probabilistic sensitivity analyses. Inaccurate assumptions about correlation structure could greatly bias assessments of cost-effectiveness model uncertainty. The explicit assumptions about input variable correlation should therefore accompany reported results of probabilistic sensitivity analyses.
ORAL CONCURRENT SESSION A - JUDGMENT AND DECISION MAKING
PRACTITIONERS’ INSIGHT INTO THEIR OWN DECISIONS ABOUT PRESCRIBING ANTIBIOTICS IN RESPIRATORY INFECTIONS
Winston R1, Dan C2, Leeman-Castillo B3 and Canales R1
1Univ of Nebraska Medical Center College of Medicine, Omaha, NE; 2University of Colorado at Denver, Westminster, CO; 3University of Colorado at Denver, Denver, CO; University of California, San Francisco, San Francisco, CA
Purpose: Understanding how physicians decide whether to prescribe antibiotics is a key step in reducing antibiotic use in respiratory infections (RI). It is not known, however, whether practitioners understand their own use of clinical information in making this decision. Because recent studies have suggested, contrary to prior reports, that decision makers can recognize their own policies, we hypothesized that primary care practitioners (PCP) could be able to distinguish their own decision policy for prescribing antibiotics in RI from that of other PCPs.
Methods: Each of 81 community PCPs reviewed 20 case vignettes of patients with RI constructed using 8 variables in a fractional factorial design. For each case, they decided whether to prescribe antibiotics. For each PCP, we determined the weight they gave each clinical case in deciding about antibiotics (the decision policy). We used cluster analysis to define 9 groups of practitioners with similar policies and chose a policy closest to the mean case weights of the group to be the group archetype. We showed histograms of the 9 archetypes to each PCP with a histogram of their own policy substituted for the archetype of the cluster to which they belonged. We asked them to select their own policy from the display.
Results: Of 81 PCPs who completed the profiles, 54 participated (67%). When asked to identify their own from the 9 policies displayed, 9 of the 54 participants (11%) correctly identified their own rate no better than chance (p = 0.93).
Conclusions: PCPs could not distinguish their own policy for prescribing antibiotics from archetypes of the 8 other approaches taken by 81 PCPs. These other approaches differed significantly in the pattern of weighting. This indicates that PCPs were not sufficiently aware of how they use clinical variables in deciding about antibiotics—an important consideration in changing prescribing behavior. Thus, approaches to improve prescribing may need to include not only information about the optimal use of clinical information in making the decision for antibiotics but also making practitioners aware of their own decision processes.

ORAL CONCURRENT SESSION B - METHODOLOGICAL ADVANCES
BAYESIAN META-ANALYSIS OF PAPANICOLAU SMEAR ACCURACY
Coom 2, Cox D and inventors S3
1Rice University, Houston, TX; 2The University of Texas MD. Anderson Cancer Center, Houston, TX
Purpose: The purpose of this study is to synthesize a published meta-analysis of Papanicolaou (Pap) smear (Fahey, et al. Am J Epidemiol 1995; 141:586-609) using Bayesian approaches and compare the results.
Methods: The Fahey study considered two subgroups for summary receiver operating characteristic (SROC) curve analysis: one subgroup (N = 31 studies) used Pap smear as a follow-up test after a previous abnormal result; the other (N = 27) used Pap smear for screening. We considered two Bayesian hierarchical models for each subgroup. The first model considered true positives and true negatives as binomial random variables, with sensitivity (7) and specificity (7) as the probability parameters, respectively. We assumed that sensitivity and specificity have independent beta distributions with exponential prior on the beta parameters. The second model considered sensitivity and specificity jointly through the log odds ratio, 8 = log(β) – log(1 – β), where 7 followed a normal distribution, which allowed for possible correlation between sensitivity and specificity. We performed sensitivity analysis to examine the effect of prior selection on the parameter estimates.
Results: We examined the possible existence of publication bias using funnel plots. Plots using the second model had the expected funnel shape. The table compares the estimates of overall sensitivity and specificity from the Bayesian models with Fahey’s SROC results. Results from Bayesian model 1 are similar to the SROC approach but with wider credible intervals. The point estimates from Bayesian model 2 are higher although the credible intervals overlap. Sensitivity analysis showed that the models are somewhat sensitive to the variance of the prior distribution; however, the point estimates are more robust.

<table>
<thead>
<tr>
<th></th>
<th>Bayesian Model 1*</th>
<th>Bayesian Model 2*</th>
<th>SROC Curve**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Follow-up Sensitivity</td>
<td>0.61 (0.52-0.70)</td>
<td>0.70 (0.58-0.80)</td>
<td>0.66 (0.58-0.73)</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.55 (0.53-0.72)</td>
<td>0.73 (0.63-0.81)</td>
<td>0.66 (0.58-0.73)</td>
</tr>
<tr>
<td>Screening Sensitivity</td>
<td>0.35 (0.44-0.66)</td>
<td>0.61 (0.56-0.75)</td>
<td>0.58 (0.49-0.67)</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.65 (0.54-0.75)</td>
<td>0.79 (0.66-0.84)</td>
<td>0.69 (0.62-0.77)</td>
</tr>
</tbody>
</table>

*median (95% credible interval); **mean (95% confidence interval)

Conclusions: We concluded that the Bayesian approach has advantages over SROC in that it accounts for between-study variation and allows for estimating the sensitivity and specificity in a particular trial, taking the results of other trials into consideration, i.e., “borrowing strength” from one another. The log odds ratio model, allowing for correlation, produced higher point estimates for both sensitivity and specificity.

ORAL CONCURRENT SESSION A - JUDGMENT AND DECISION MAKING
PROSPECT THEORY IN THE EVALUATION OF ARTHRITIS HEALTH STATES
Suarez-Almanza M1, Souchek J1, Byrne M2, Kelly P3, Richardson M3, Pak C4 and Nelson H5
1University of Colorado, Boulder, CO; 2University College of Medicine/E Michael E. DeBakey VA Medical Center, Houston, TX; 3University of Pittsburgh, Pittsburgh, PA
Purpose: In prospect theory (PT) value is defined in terms of gains and losses from a reference point, with a value function which is convex and relatively steep for losses and concave and not so steep for gains. Our objective was to examine PT postulates in the valuation of a hypothetical osteoarthritis (OA) state, using current health status as the reference point for valuation.
Methods: We surveyed 101 subjects (108 patients with OA and 103 community members). Participants were interviewed and asked to rate their own health, as well as vignettes describing patients with OA using visual analog scale (VAS), standard gamble (SG), time trade-off (TTO) and willingness to pay (WTP). For this analysis we used a scenario based on EQ-SD domains, describing an individual with mild to moderate OA (York tariff 0.69). Subject valuations of this state were examined in relation to the difference score between the state and their own health (gains/losses) as determined by EQ-SD York tariffs.
Results: In the public, 27.2% had worse health status than the vignette, 9.9% had similar health, and 63.9% had equal health. For the patient group, the percentages were 50.5%, 16.7%, and 23.8%, respectively. The difference scores between the scenario and current health was related to the valuations of the states as follows (Spearmans correlation: VAS: r = -0.06 (p = 0.22); SG: r = 0.35 (p < 0.001); TTO: r = 0.35 (p = 0.004); and WTP: r = -0.10 (p = 0.39). An astute post-IT, individuals perceiving the hypothetical state as a gain in health (positive difference) were less likely to trade or gamble than individuals who perceived the hypothetical state to be a lose (negative difference). With SG, the slope was steeper for losses than for gains, with no major differences seen with the other techniques. Overall however, the predictive value of current health as a reference point was low.
Conclusions: Our findings suggest that subjects’ valuation of arthritis states follow PT principles as a whole using VAS, and to a lesser degree TTO. Yet, the explanatory power of current health as a reference point is weak, suggesting that other unrelated factors may play a more important role in patients’ preferences as elicited by these techniques.

ORAL CONCURRENT SESSION B - METHODOLOGICAL ADVANCES
IMPACT OF HEART FAILURE ON 1-YEAR COSTS AFTER MYOCARDIAL INFARCTION RELATION TO HOSPITALIZATION: AN APPLICATION USING A FLEXIBLE LINK AND VARIANCE FUNCTION MODEL
Basu A1, Aronkole B3 and Rathouz P4
1University of Chicago, Chicago, IL; 2University of Illinois at Chicago, Philadelphia, PA
Background: Increased incidence of myocardial infarction (MI) and improved post-MI survival have led to an increase in the incidence of post-MI heart failure (HF) and in the cost of care attributable to post-MI HF. Objectives: 1) To estimate the potential 1-year cost savings if HF could be prevented in post-MI patients. 2) To illustrate the biases that may arise in using alternative estimators to model expected medical expenditures data. Methods: We use data for years 1998, 1999 and 2000 from a large claims database. After applying the exclusion criteria, we group eligible patients into those with and without HF subsequent to MI using inpatient, outpatient and prescription records. We model the total medical expenditures for each patient over the 1-year period post index date. Covariates include HF, age, sex, death and comorbidities at index hospitalisation, type of insurance, procedures performed, year and the type of MI. The estimators that we consider are the ordinary least squares (OLS) regression, log-transformed GLS regression with and without heteroscedastic smearing, the gamma regression with log-link and the extended estimating equations (EEE) model that estimates both the link and variance parameters for the data along with the regression coefficient. Results: 15,118 patients were eligible for the study. Computed one-year follow-up data were available for 7621 patients. No significant differences were found in observed variables between those with complete one year data and those without. Significant differences in estimated cost-savings were found between estimators. The EEE model estimator was found to be more appropriate than other estimators based on a broad set of goodness of fit tests and tests of over-fitting. Based on this estimator, the potential 1-year cost-savings due to preventing HF in post-MI patient who develop HF was estimated to be $1,470 (113.3). The cost-savings decreased with increasing age. Conclusions: Careful selection of estimator is important for modeling cost data. The EEE estimator appears to outperform alternative estimators studied. We find that preventing HF in post-MI patient can produce substantial savings in health care costs. The estimates provided here can serve as a guide in conducting cost-effectiveness analysis of the new treatments that are aimed to prevent heart failure in patients with myocardial infarction.
ORAL CONCURRENT SESSION A - PATIENT AND PHYSICIAN BEHAVIOR/PREFERENCES

Sullivan P1 and Lawrence W1

1Pharmaceutical Outcomes Research Program, University of Colorado School of Pharmacy, Denver, CO; Agency for Healthcare Research and Quality, Rockville, MD

Purpose: The variability in preferences used in QALY estimation jeopardizes the comparability of cost-effectiveness analyses and has led the Panel on Cost-Effectiveness in Health and Medicine (the PCEHM) to call for a catalogue of “off-the-shelf” preference weights associated with conditions that can be used by health researchers without the burden of collecting primary data.

Methods: The current research responds to the call by developing a nationally representative catalogue of preference weights for chronic conditions and associated sociodemographic characteristics. The authors report the EQ-5D scores associated with chronic conditions and associated sociodemographic characteristics in the nationally representative Medex Panel Survey (MEPS). Chronic conditions were coded using “Quality Priority Conditions” (QPC), Clinical Classification Categories (CCC) and the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9). In order to provide accurate age-adjusted estimates in the face of the ceiling effect exhibited by preference measures, OLS, Tobit and censored least absolute deviations (CLAD) regression models were employed and compared.

Results: As with many health status measures in population health surveys, it is clear from these results that the EQ-5D is not normally distributed, exhibits a significant ceiling effect and the mean underestimates the population EQ-5D score. In addition, errors from OLS estimation exhibit significant heteroskedasticity. Given these factors, the CLAD estimates (or median) appear to be a more appropriate measure of central tendency of EQ-5D scores. Unadjusted and age-adjusted EQ-5D scores as well as the age, gender, race, ethnicity, poverty status and education level distribution associated with each QPC, CCC and ICD-9 code are presented. EQ-5D scores for older age categories were lower than younger categories, female scores were lower than males, certain racial groups had lower scores than others, and EQ-5D scores were higher for individuals with higher education and income levels.

Conclusion: Use of the mean may not be the most appropriate measure of central tendency when estimating population EQ-5D scores. The chronic condition scores reported in this research are nationally representative and may be useful to researchers as preference weights that can be used to calculate QALYS for cost-effectiveness analyses without the burden of primary data collection.

ORAL CONCURRENT SESSION B - METHODS IN JUDGMENT AND DECISION MAKING

WHEN TO ACCEPT A CADAVERIC LIVER OFFER FOR TRANSPLANTATION? A MARKOV DECISION (MDP) MODEL APPROACH

Altmann-L1, Schaefer A2, Roberts M1

1University of Pittsburgh, Pittsburgh, PA; 2Case Western Reserve University, Cleveland, OH

Purpose: Although there is a shortage of cadaveric livers, 45% of all liver offers are initially declined by the transplant surgeons/patients. Organs are rejected in the hope that they will receive a better organ offer in the future. Our objective is to solve the decision problem faced by these patients: should an offered organ of a given quality be accepted or rejected? The decision depends on two major components: the patient’s current and likely future health as well as the current and future prospects for organs. This extends our previous work that determined when to accept a living-donor liver.

Methods: Markov decision processes (MDPs) are analytic tools for sequential decision-making under uncertainty. We developed an infinite horizon discounted MDP model that determines whether to accept or decline a liver offer. The state of the system is described by patient health and organ quality. The Model for End-Stage Liver Disease (MELD) score, a risk prediction score, represents patient health. We estimate the transition probabilities using the natural history of liver disease and the national liver offer rates. We found the optimal policies using a policy iteration algorithm. Results: The figures show an optimal policy based on a simplified definition of organ quality, where the quality is determined by donor characteristics such as age, gender, race, etc. Organ scores are ordered in a decreasing quality, that is, Organ 1 is the best and Organ 14 is the worst organ. Note also that the higher the MELD score, the sicker the patient is. As expected, the optimal accept/decline decision is a function of organ quality and patient health. For example, when the patient has a MELD score of 18, the policy suggests that the patient should not accept liver offers that have a lower quality than 5 (that have a higher figure than 5).

Conclusions: Given functions that represent patient illness and organ quality, this analytic model can be used to determine the accept/decline decisions for cadaveric organs.
ORAL CONCURRENT SESSION A - PATIENT AND PHYSICIAN BEHAVIOR/ PREFERENCES 2

THE EFFECT OF A DECISION-ASSISTING TOOL ON PREGNATAL TESTING INCLINATIONS AND BEHAVIORS

Kuppermann M, Gregorich S and Neas R

1University of California, San Francisco, San Francisco, CA; Express Scripts, Inc., and Washington University School of Medicine, Maryland Heights, MO

Purpose: Decisions regarding prenatal testing for Down syndrome are value laden and should reflect informed patient preferences. We explored the effect of an interactive computerized decision-assisting tool (“PT Tool”), which we have previously shown to increase knowledge and decrease decisional conflict, on women’s prenatal inclinations regarding prenatal testing in a randomized clinical trial.

Methods: 464 pregnant women were randomized to either PT Tool or the computerized version of California’s educational pamphlet (control). Baseline testing inclinations were assessed prior to randomization and 5.2 weeks later by asking: “If you were offered amniocentesis free of charge in your current pregnancy, would you choose to have it?” with response options ranging from “definitely would” to “definitely would not.” Utilization was assessed at a 30 gestational week interview or via chart review.

Results. The overall rate of utilization of invasive testing (10%) did not differ between PT Tool viewers and controls among women aged < 35, to whom these tests are not routinely offered. A follow-up, inclination to undergo diagnostic testing was lower among women who viewed PT Tool (24% versus 33%, p = .03), suggesting that all else being equal, viewing PT Tool would attenuate the uptake rate of invasive testing in younger women if these tests were offered to them free of charge. When it came to utilization of invasive testing among women aged 35 and older, women in the control and PT Tool intervention groups also had almost identical invasive test rates (66%), after controlling for baseline inclination. However, an interaction between the experimental intervention and baseline invasive test inclinations was observed.

Conclusion: The effect of a decision-assisting tool (PT Tool) on prenatal testing inclinations is related to baseline inclination levels. PT Tool is a useful and unbiased decision aid that helps women make informed choices about prenatal testing.

ORAL CONCURRENT SESSION B - METHODS IN JUDGMENT AND DECISION MAKING RANDOMIZATION TURNS DIAGNOSTIC RESEARCH INTO INTERVENTION RESEARCH

Moons K, Bleshurev C and Grobbée D

University Medical Center Utrecht, Utrecht, Netherlands

In almost every system to grade epidemiological studies according to their level of evidence, randomized studies or meta-analyses of randomized studies receive the highest classification. Although the use of such hierarchies may help to separate the wheat from the chaff, it has also led to misconception and abuse. The popular belief that only randomized studies produce results applicable to clinical practice does a disservice to patient care, clinical investigation and the education of health care professionals. Indeed the ultimate goal of medical care, including diagnostic testing, is to improve patient outcome. Hence, it has widely been advocated to establish whether a tool’s diagnostic accuracy, the impact of the test on patient outcome must also be quantified. As studies on patient outcome commonly require a randomized approach to prevent confounding, a randomized design for diagnostic test evaluations is increasingly advocated. However, to demonstrate the beneficial effect of a diagnostic procedure or strategy on patient outcome, we believe that randomization is by no means a prerequisite. Using randomized studies in diagnostic research certainly changes an essential characteristic of this type of clinical research. It turns diagnostic accuracy (or classification) research into intervention or efficacy research. The nature of this change is clear from the definition and the object of research determine the appropriate study design. In our view, a test’s effect on patient outcome can be inferred and indeed considered as quantified 1) if the test is meant to include or exclude a disease for which an established reference is available. 2) if a cross-sectional accuracy study has shown the test’s ability to adequately detect the presence or absence of that disease based on the reference, and 3) if the randomized studies have provided evidence of efficacy of the optimal management of this disease. In such instances diagnostic research does not require an additional randomized comparison between two or more test–treatment strategies (one with and one without the test under study) to establish the test’s effect on patient outcome. We will additionally discuss when a randomized design is needed to properly infer on a test’s value to change patient outcome. We conclude that in many instances, a cross-sectional accuracy study will be sufficient to quantify the clinical value of diagnostic tests on patient outcome.

ORAL CONCURRENT SESSION B - METHODS IN JUDGMENT AND DECISION MAKING MEMORY FOR AND USE OF THE OTTAWA ANKLE RULES

Brehaut J, Stoll I, Visentin L, Graham I and Grimshaw J

Ottawa Health Research Institute, Ottawa, ON, Canada

Purpose: The Ottawa Ankle Rules (OAR) are extremely sensitive for detecting foot and ankle fractures, and are reportedly used by over 90% of Canadian emergency physicians. However, little is known about how this tool is used in practice. We surveyed physicians to determine whether they report using the OAR consistently and/or exclusively, and whether the rule can be remembered consistently. We also examined the effect of a computerized decision-support tool on the use of OAR.

Methods. We surveyed members of the Canadian Association of Emergency Physicians (CAEP) about their use of the OAR in practice, and tested their memory for the rule via a single multiple choice item in which they were asked to pick out the 5 components of the OAR from a series of 5 plausible alternatives. Conducted between March and May 2003, the 4-page survey included 24 closed-ended questions and 3 sections: one asking a series of questions about the OAR, another examining use of other clinical decision rules, and one on practice details and demographics. The sample frame included 400 active CAEP members, and excluded those listed as non-MD’s, retired, or not currently residing in Canada. The survey was conducted according to standard survey methodology (Dillman, 2000).

Results. Of our initial sample of 400 names, 376 were eligible potential respondents with corrected response rate of 94.2%. The survey team received 284 completed surveys (71% response rate). We found that 281 (99.6%) had heard of the OAR. Of these, most respondents (92%) were familiar with the OAR’s rule, and 91% were aware of the 3 clinical scenarios for which the OAR is used. We found that 130 respondents (46.1%) reported using the OAR consistently, and 116 respondents (41.0%) reported using the OAR exclusively. Of the respondents who used the rule exclusively, 65 respondents (55.3%) reported using the rule for all clinical scenarios, and 51 respondents (44.7%) reported using the rule for fewer scenarios. There were no differences in the reported consistency and/or exclusivity of the OAR use among the three sections of the survey.

Conclusion: Almost all respondents reported being familiar with the OAR and most report using it consistently. Most report not using the rule exclusively, but consider other factors (some of which may be inappropriate) in a decision that might reasonably be governed by the rule alone. Memory for this simple rule was imperfect in the majority of respondents. How these considerations change the effectiveness of clinical decision rules warrants further study.
ORAL CONCURRENT SESSION A - PATIENT AND PHYSICIAN BEHAVIOR/PREFERENCES 2

PREFERENCES: VIEWS OF PRIVACY AND RESEARCH: DELIBERATIVE DEMOCRACY AT WORK.

Danziger, L. Kohn, A. and Hayward, R.
VA Ann Arbor Healthcare System, Ann Arbor, MI. Georgetown University, Washington, DC. The Ohio State University, Columbus, OH

Purpose: The HIPAA Privacy Rule went into effect April 2003, giving IRBs only vague guidelines on when researchers may access medical records without permission from patients. Patients’ opinions on this issue are essential but difficult to obtain because the topic requires considerable understanding of privacy violation risks and knowledge of minimal risk research. We used a novel mixed-methods approach to provide deliberative responses to this complex policy question.

Methods: We used a deliberative democratic approach, coupled with baseline and follow-up phone surveys, to obtain recommendations about optimal criteria for allowing researchers access to medical records without permission from patients. A representative sample of veterans, drawn from two rural and two urban geographically dispersed regions, was invited to participate in an all-day deliberation session and baseline and follow-up phone surveys. 231 veterans participated in the sessions and were randomly assigned to deliberation subgroups of 4-7 people each. An additional 295 veterans participated in the phone surveys alone, as a comparison group.

Results: The majority of groups and individuals thought that researchers should ask patients at least once before using their medical records for research. Whether veterans trust that researchers will keep their medical information confidential depends on who is conducting the study. Over 80% of veterans trust that the VA will ensure confidentiality of their medical information always or most of the time and said they would be willing to allow VA researchers access to their medical records compared to only about 48% who held these views related to pharmaceutical company researchers. Veterans who participated in the all-day deliberation became more willing (p = .008) to allow VA researchers access to their medical records after having done the deliberation, while veterans who did not attend the session were no more willing (p=.53) at the times of the follow-up surveys. 96% of veterans felt the deliberation process was fair and 94% felt they were listened to. One veteran said that “with more exposure and thought, my decisions are more in line with my moral values.”

Conclusion: Veterans want a say in how their medical records are used in research. Deliberative methods appear as an effective way to obtain informed recommendations from patients for complex and value-laden policy issues in a way that traditional focus groups and large-scale surveys cannot.

ORAL CONCURRENT SESSION B - METHODS IN JUDGMENT AND DECISION MAKING

“What’s Better Than You? Then I’m OK”: Comparative Information and Risk Prevention Strategies.

Fagerlin, A., Kizmazk-Fisher B and Ubel, P.
University of Michigan, Ann Arbor, MI. VA Ann Arbor Healthcare System, Ann Arbor, MI

Purpose: Imagine you have a 6% chance of developing cancer: A medication can reduce your risk of cancer to 3%, but it has serious side effects. Would you take the medication? Now, imagine that you just learned that the average person’s risk of cancer was 3% and your risk of cancer would drop to 1.5% if you took the medication. Would this information change your preference for the medication? Should it change your preference? This study tested whether attitudes toward cancer prevention drugs change when people know how their risk of cancer compares to an average person.

Methods: 504 women were randomly assigned to receive one of four hypothetical scenarios in which they imagined that their risk of developing breast cancer in the next 5 years was 6%. These survey versions differed only in whether the average woman’s risk of breast cancer was said to be the same as the respondent’s (6%), lower (3%), or higher (12%), while the fourth group did not receive comparative information. All respondents were told about a pill that could reduce their risk of breast cancer by 50%, although the pill has some serious side effects. Respondents answered questions about their attitudes toward breast cancer and the cancer reducing pill.

Results: When told their risk of breast cancer was higher than the average woman’s, women were more worried about breast cancer (p < .02), expressed greater desire to take the pill (p = .003), and perceived the risk reduction caused by the pill to be more significant (p < .01) than women described as average risk or below average risk for breast cancer. Women who did not see average risk information behaved similarly to those at average or below average risk. Most women perceived being able to compare their risk to that of the average woman as being helpful.

Conclusions: Providing comparative risk information can significantly influence people’s decisions about prevention strategies. If a prevention strategy reduces a person’s risk by half, it should not matter whether others receive greater or lesser benefit from the same strategy, yet this knowledge changes behavior. This tendency of people to focus on comparative, rather than absolute, benefit occurs in many contexts and has implications for clinical practice and research (including design of decision aids).

ORAL CONCURRENT SESSION B - METHODS IN JUDGMENT AND DECISION MAKING

Probing Probabilistic Information in a Randomized Clinical Trial Consent Document: Do Format and Preference Matter?

MacMillan, K. and Llewellyn-Thomas, H.
University of Toronto, Toronto, ON, Canada. Dartmouth Medical School, Hanover, NH

Purpose: Effective communication of risk/benefit information requires the recipient to understand probabilistic information. Comprehension may be affected by whether the format in which the information is presented (i.e., using numbers or words that matches peoples’ preferences) and the source of the information (whether the information format “match” or “mismatch.”

Methods: A convenience sample of 228 subjects, recruited from various clinical and community sites, received pre-assembled randomized packages. These packages contained a) a sham consent form containing risk and benefit information in either numeric or verbal format; and b) questionnaires assessing a) the primary outcome of comprehension; and c) underlying format preference.

Results: The study design identified emergent “match” (n = 105) and “mismatch” (n = 119) subgroups, based on whether or not the received format was congruent with the subjects’ reported preference. Overall, 32.5% of subjects achieved correct responses to all 8 comprehension items, and comprehension scores did not differ between the subgroups. However, the type of match/mismatch did demonstrate an effect. Among the 48 subjects who preferred the verbal format, comprehension scores were lower, regardless of whether they received a “match” (n = 29, mean 3.80) or a “mismatch” (n = 26, mean 5.46). Among the 178 subjects who preferred the numeric format, comprehension scores did not significantly differ between those who received a “match” (n = 85, mean 6.83) and those who received a “mismatch” (n = 93, mean 6.74).

Conclusions: This study is unique, in that it investigates the effects of format/preference match or mismatch on the comprehension of probabilistic information that is required for genuine consent to RCT entry. A notable proportion would not have provided a fully informed RCT consent/refusal. Those who preferred verbal format were not helped by receiving a format match and their comprehension scores were lower. Results imply that individualized assessment of format preferences and numeracy skills could point to strategies to foster adequate levels of comprehension. The study design could guide future exploratory investigations in other contexts.

E14 • MEDICAL DECISION MAKING/JAN–FEB 2005
ABSTRACTS

ORAL CONCURRENT SESSION A - SIMULATION

THE CASE FOR LIMITED MALIGNANT POTENTIAL BREAST CANCER


$University of Wisconsin, Madison, WI; National Yang-Ming University, Taipei 11221, Taiwan

Purpose: The sharp rise in early stage breast cancer incidence in the past 20 years is generally attributed to screening mammography. However, sustained high incidence levels with plateaus in late-stage incidence led us to ask where all these cancers were before the advent of widespread screening. Accounting for this, we can now flesh out an explanation sketched over a decade ago: the Limited Malignant Potential breast cancer hypothesis.

Method: A discrete event simulation of female breast cancer epidemiology from 1975-2000 was constructed. The simulation incorporates data about mammography dissemination, dissemination of adjuvant therapy, and improvements in mammography opening characteristics as well as secular improvements in female longevity over this time period. The simulation parameters for breast cancer natural history were heuristically manipulated in the context of these other simulation components in order to fit observed age- and historical stage-specific breast cancer surveillance data across the 25 years from 1975-2000. Computationally intensive parameter sampling experiments were conducted to assess likelihood of alternatives to the heuristically “best” solution.

Results: A large pool of occult breast cancers must exist to become incident breast cancers. The simulation balances the size of this pool and the assumed natural history of breast cancer against long run patterns of care data for diffusion of screening mammography and surveillance data regarding disease incidence and mortality. The “best fit” lead us to conclude 42% of all biologically initiated breast cancer is LMP. Sampling experiments rule out substantial likelihood this fraction is under 30% or more than 55%. LMP tumors start with small foci (<2 cm diameter) and grow to a maximum of approximately 1 cm diameter. If not detected in a span of about 2 years, they will reseed and disappear. They never present a lethal threat to the host. In the year 2000 LMP breast cancers accounted for 30% of incident, clinically localized breast cancers and 44% of incident in situ disease. LMP breast cancers presently cannot be discerned histologically from in situ or early localized invasive cancer.

Discussion: Our calculations support the conclusion that a substantial fraction of diagnosed early stage breast cancer is LMP representing over-diagnosis. If true, the LMP hypothesis makes the value of developing a test to distinguish LMP from non-LMP breast cancers large.

ORAL CONCURRENT SESSION A - SIMULATION

EXPLAINING THE SECULAR TRENDS IN COLORECTAL CANCER INCIDENCE AND MORTALITY USING A POPULATION-BASED MICROSIMULATION MODEL


Harvard School of Public Health, Boston, MA

Purpose: To examine the relative contribution of changes in risk factors, screening and treatment on the overall population trends in colorectal cancer (CRC) incidence and mortality.

Methods: We developed a dynamic first-order Monte Carlo model to simulate the US population aged 25 years and older from 1970 to 2000. The model tracks the development of adenomas and their progression to invasive CRC as a function of demographic characteristics and risk factors (e.g., smoking history, body mass index, and folate intake). Secular trends in risk factors were estimated from the National Health and Nutrition Examination Surveys, dissemination of CRC screening was derived from the National Health Interview Surveys, and advancements in treatment were obtained from SEER and SEER-Medicare linked data.

We used the model to simulate the observed trends in CRC incidence and mortality from 1970 to 2000 and generate estimates of the number of cases and deaths that would have occurred under alternative scenarios regarding changes in risk factors, screening, and treatment over this time period. We then compared the results of these scenarios and decomposed the overall changes in incidence and mortality into the proportions explained by each factor.

Results: During the period 1970 to 2000, approximately 2.54 million Americans were diagnosed with CRC and 1.28 million died from the disease. The number of cases represents a 4.7% reduction from the estimated number that would have occurred if there had been no secular trends in risk factors and no dissemination of screening over this time period (8.6% reduction in 2000). Changes in risk factors alone account for 31.5% of the overall reduction, and 62.3% is attributable to screening.

The number of cancer deaths represents a 33.7% reduction from the estimated number that would have occurred in the absence of changes in risk factors, screening, or treatment (22.6% reduction in 2000). Advancements in treatment alone account for 59.9% of the reduction, while risk factors and screening account for 7.8% and 38.7% of the decline, respectively.

Conclusions: Screening and advancements in treatment have played significant roles in the declines in CRC incidence and mortality. Population-based models can provide insight into observed disease trends and shed light on areas in which public health interventions could have a large effect.

ORAL CONCURRENT SESSION B - SCREENING IN CHRONIC DISEASE

COST-EFFECTIVENESS OF HIV RESCREENING DURING LATE PREGNANCY TO PREVENT PERINATAL HIV TRANSMISSION IN A RESOURCE-LIMITED COUNTRY


$VA Palo Alto Health Care System, Palo Alto, CA; St. Michael’s Hospital, University of Toronto, Toronto, ON, Canada; Stanford University, Stanford, CA

Although Human Immunodeficiency virus (HIV) infection is more prevalent in people under age 45, a significant number of infections occur in older age groups. For example, in the Department of Veterans Affairs (VA), which is the largest provider of HIV care in the U.S., approximately 20% of HIV infections occur in patients older than 55. We assessed the cost-effectiveness of rescreening for HIV and treatment with antiretroviral therapy for the elderly.

We developed a Markov model that evaluated costs, quality of life, and survival for patients in an HIV screening program compared to current practice. In both strategies, symptomatic patients could be identified through case finding. Identified patients started treatment when their CD4 count dropped to 350 cells/μL or their viral load exceeded 50,000 copies/mL. Disease progression was based on CD4 and viral load levels. Sexual transmission was based on viral load, knowledge of HIV status, and efficacy of counseling. In the base-case analysis, we considered the potential societal cost savings and health benefits associated with decreased HIV transmission after diagnosis and treatment.

For a 50-year-old patient, one-time screening for HIV costs $83,700 per quality-adjusted life year (QALY) as compared to current practice. Because of the increased risk of age-specific mortality (related to HIV infection), one-time screening for HIV of an 80- or 90-year-old patient increases to $36,400/QALY and $99,900/QALY, respectively. If transmission to sexual partners is excluded, one-time HIV screening cost $45,500/QALY, $98,800/QALY and $210,300/QALY in 50-, 80-, and 90-year-olds, respectively. If the cost-effectiveness threshold for implementing a screening program were $50,000/QALY, screening would be offered to under 84 years or 62 years of age depending on whether the effects of transmission were included or not.

Screening for HIV is cost effective in patients up to about 65 years of age, even if they are not sexually active. Screening is cost effective in patients older than 65 if they are sexually active and have a partner at risk for HIV transmission. Thus, HIV screening is cost effective in a substantially broader age range than is recommended in current guidelines.
ORAL CONCURRENT SESSION A - SIMULATION
PROJECTING HEALTH OUTCOMES AFTER THE FOLIC ACID Fortification POLICY IN THE UNITED STATES

Buntel T, Weinstein M, Willett W and Kunz K
Harvard School of Public Health, Boston, MA

Purpose: For countries deciding on folic acid fortification policies, the tradeoff of focus has been on benefits for neural tube defects (NTDs) versus risk of masking B-12 deficiency. However, folate consumption is also associated with reduced risks of cardiovascular disease and colon cancer. We used a decision-analytic model to quantify the projected health outcomes associated with changes in folate consumption following the 1998 mandate to fortify enriched grain products with folate acid in the US.

Methods: We developed a Markov model to simulate the effect of changes in folate consumption on disease incidence, comparing pre-fortification vs. post-fortification levels. The starting folate distributions before and after fortification were based on separate analyses of two National Health and Nutrition Examination Surveys (1988-1994 and 1999-2000). We analyzed cohorts by age, sex, and race (non-Hispanic white, non-Hispanic black, and Mexican-American). The model distributes a cohort of disease-free individuals into four folate intake states: <200, 201-300, 301-400, and >400 micrograms per day. In each one-year cycle, persons face gender, age, race, and folate-specific risks based on published associations of developing an NTD-affected pregnancy, myocardial infarction (MI), colon cancer, or masking of B-12 deficiency, with multiple events allowed.

Results: In a 25-year-old US cohort, the increase in folate intake after fortification was predicted to prevent the following disease burden over 10 years: 705 NTD, 74 NTIs, and 51 colon cancers; there is no risk of B-12 deficiency in the young cohorts. The greatest percent change was in Mexican-Americans, while the largest population-based benefits accrued in non-Hispanic whites. In a 65-year-old US cohort over a 15-year period, 897 persons face gender, age, race, and folate-specific risks based on published associations of developing an NTD-affected pregnancy, myocardial infarction (MI), colon cancer, or masking of B-12 deficiency, with multiple events allowed.

Conclusions: The potential risks of a small number of cases of B-12 masking were weighed against a substantial predicted benefit of NTDs, MI, and colon cancers prevented. This knowledge can help policymakers—both in the US and other countries—who continue to debate the risks and benefits of fortification.

E16 • MEDICAL DECISION MAKING/JAN–FEB 2005

ORAL CONCURRENT SESSION B - SCREENING IN CHRONIC DISEASE
IMPLEMENT OR ANALYZE? A VALUE OF INFORMATION ANALYSIS OF ULTRASOUND FOR BREAST CANCER SCREENING

Schleinitz M1, Blume J2 and Berg W2
1Brown University and Rhode Island Hospital, Providence, RI; 2Brown University, Providence, RI; Berg Enterprises and American College of Radiology Imaging Network, Philadelphia, PA

PURPOSE: Observational series have suggested that adding annual whole breast ultrasound (US) to mammographic screening improves detection of early stage cancers, particularly for women with mammographically dense breasts. We performed a value of information analysis to determine whether further study of screening breast ultrasound is necessary prior to implementing this strategy.

METHODS: We constructed a Markov model from a societal perspective comparing screening with mammography plus US to mammography alone for women with mammographically dense breasts. In both arms, we estimated the benefits of perfect information by calculating the average difference between the expected net economic benefit given perfect information on all model variables and the expected net economic benefit given current data in 5,500 Monte Carlo simulations at cost-effectiveness thresholds ranging from zero to $100,000 per quality adjusted life year (QALY). In addition to perfect information on all model variables, we also independently assessed the value of information on the effectiveness of US, cost of US, quality of life measures, efficacy of mammography and disease related variables. We assumed that $50,000 women would meet our criteria for US screening over the next 10 years.

RESULTS: In the base case, the addition of US to mammography improved quality adjusted life expectancy by 0.056 QALYs at an incremental cost-effectiveness ratio (ICER) of $38,000 per QALY. The maximum value of simultaneous, perfect information on all variables was $375 per person, or $176 million for the eligible population. Considered individually, the incremental efficiency of US had a maximum value of $142 million, perfect information on the cost of US and on utilities each had a value of $988 million. Value of information decreases as cost-effectiveness thresholds exceed the ICER. At $100,000 per QALY perfect information has an expected value of $24 million.

CONCLUSIONS: Screening breast US may improve the lives of eligible women, but its cost-effectiveness is uncertain. Further research would decrease this uncertainty and has a maximum expected value of $176 million for the eligible population. Considered individually, the addition of US to breast cancer screening.

ORAL CONCURRENT SESSION B - SCREENING IN CHRONIC DISEASE
A QUALITATIVE STUDY OF AFRICAN-AMERICAN'S DECISION TO SCREEN FOR PROSTATE CANCER: BUST NOT TO KNOW?

Goldberg I
University of Illinois at Chicago, Chicago, IL

Purpose: There is no consensus regarding the risks and benefits of screening for early-stage prostate cancer. Medical organizations encourage shared decision making both to improve understanding and reach individualized decisions. The question then becomes, what is necessary for an “individualized” decision? An additional complexity is that low-income, African-American (A-A) men are even less likely to be screened and at greater risk of being diagnosed with advanced stages of disease. The aim of this research was two-fold: (1) identify issues and values associated with this decision from two sources: A-A patients-at-risk for prostate cancer and A-A patients’ wives; and (2) examine specific cognitive processes that may differentially impact the screening decision.

Method: A triangulation methodology was used involving two qualitative. First, 4 focus groups each with 4-6 A-A male patients over 40 with no history of prostate cancer, and 4 focus groups with 4-7 wives, followed by semi-structured interviews. The focus groups were used to map the domains of relevant issues. The follow-up interviews confirmed and validated the findings with participants from the original focus groups and a new set of participants.

Results: Preliminary results identified the importance of both cognitive and affective factors. Cognitive themes included an anticipated shift with aging in the trade-off between quality and quantity of life (as one gets older, the quality of one’s sex life may be less important than living long enough to play other social roles, such as grandfather), and anticipated diminishing marginal loss (recognizing that the incremental loss of sexual functioning at an older age may be less painful). Emotional themes included both regret and uncertainty avoidance, especially pronounced for wives as vicarious decision makers; unexpected benefits of learning one has cancer (from the viewpoint of strengthening relationships and valuing every day of life); the unique contributions of family members to the decision; and the significant role of spirituality, prayer, and fatalism.

Conclusions: Preliminary findings demonstrate the complexity of this choice for A-A men and women. Theoretically, this work allows us to understand the cognitive and emotional processes underlying this decision. From an applied standpoint, this work points us towards unexplored areas that need to be researched in order to improve management decisions for early-stage prostate cancer detection for A-A men at greater risk for the disease.
ABSTRACTS

ORAL CONCURRENT SESSION A - SIMULATION

BALANCING PATIENT AND PAYER PREFERENCES: AN EFFICIENT FRONTIER FOR BREAST CANCER SCREENING

Hyung June Keum
University of Michigan, Ann Arbor, MI

Purpose: Determining when to screen for breast cancer with mammography is a complex problem involving multiple decision makers with competing objectives. The goal of this research is to develop a model for balancing patient and payer preferences to determine policies for mammography screening as a function of patient age and perceived condition.

Methods: We use the theory of partially observable Markov processes (POMDPs) to develop efficient frontiers for balancing patient and payer preferences for breast cancer screening and treatment. A POMDP is a generalization of a Markov decision process that allows for incomplete state information. The condition of a patient is a function of many unobservable physical characteristics. There is uncertainty about any underlying disease and uncertainty associated with the response of a patient to a given treatment. In addition, different diagnostic and treatment procedures entail varying costs. In this case, the disease is observed only indirectly via a collection of incomplete, uncertain observations. We use a POMDP to incorporate uncertainty associated with the partial observability of the disease by the decision maker, and the uncertainty associated with the treatment outcome in determining the effectiveness of screening.

Results: We present a POMDP that can be used to determine: 1) when to recommend a mammogram and, 2) given the information provided by the mammogram, what treatment to provide. Further, we use the POMDP structure to develop a medical decision making tool for determining a "cost-effective" plan for mammography screening and breast cancer treatment. We develop efficient frontiers in order to explore the relationship between patient and payer preferences and to determine conditions for mammography screening to be cost-effective.

Conclusions: We have developed a model for determining cost-effective policies for breast cancer screening and treatment under conditions of uncertainty. This model incorporates uncertainty associated with the partial observability of the disease by the decision maker, the uncertainty associated with the treatment outcome, and the conflicting preferences of the patient and payer decision makers in determining the effectiveness of screening. The results show great promise as an alternative means for determining cost-effective monitoring and treatment policies for breast cancer.

ORAL CONCURRENT SESSION A - SIMULATION

CHOICE OF ANTIRETROVIRAL THERAPY AND PROJECTED LIFE EXPECTANCY AMONG SOUTH AFRICAN WOMEN EXPOSED TO NEVIRAPINE FOR PREVENTION OF MOTHER-TO-CHILD TRANSMISSION OF HIV

Holmes C1, Friedberg K1, Lesina E1 and Walensky R1
(1) Massachusetts General Hospital, Boston, MA; (2) Boston University School of Public Health, Boston, MA

Purpose: Nevirapine is an antiretroviral agent given to HIV-infected women as a single dose for the prevention of mother-to-child transmission (PMTCT) of HIV in resource-poor settings. Although resistance to nevirapine after this single dose decreases the virologic response to later combination antiretroviral therapy (ART) that includes nevirapine, long-term outcomes are unknown. Our objective was to evaluate the choice of ART regimen (nevirapine vs. protease inhibitor-based) on the life expectancy of HIV-infected South African women exposed to nevirapine for PMTCT.

Methods: We developed a Markov model to compare the life expectancies of women previously exposed to nevirapine for PMTCT starting either nevirapine-based, or protease inhibitor-based combination ART. We assumed that women with the following CD4 cell distribution at baseline started ART once the CD4 count fell below 200 cells/μl: 30% <200 cells/μl, 30% 200-500 cells/μl, 30% 501-200 cells/μl, and 10% >200 cells/μl. We modeled decrease in efficacy of nevirapine-based therapy in the presence of drug resistance, and increased efficacy as time from delivery increased. Base case probabilities were from published literature and included 18% nevirapine resistance at 10 days post delivery, efficacy of viral suppression for those with and without nevirapine resistance of 85% and 87% at 3 months, and 68% and 60% at 6 months. We performed a wide range of sensitivity analyses on ART efficacy, prevalence of baseline nevirapine resistance and other parameters.

Results: For a single ART regimen, PI-based therapy increased overall projected life expectancy by 1.6 months (Table). Women starting nevirapine-based regimens with nevirapine resistance at baseline had lower projected life expectancy than women starting nevirapine therapy without resistance. Results were sensitive to the efficacy of antiretroviral regimens and the initial distribution of CD4 cell counts, but not sensitive to varying the prevalence of baseline nevirapine resistance from 5% to 72%.

Conclusions: Life expectancy in South African women previously exposed to nevirapine is likely improved with protease inhibitor-based therapy. These results should be considered in developing guidelines for both initial and subsequent therapy for HIV in South Africa.
ABSTRACTS

ORAL CONCURRENT SESSION A - COST EFFECTIVE ANALYSIS: METHODS
A PILOT STUDY OF VALUE OF INFORMATION ANALYSIS TO SUPPORT RESEARCH RECOMMENDATIONS FOR THE NATIONAL INSTITUTE FOR CLINICAL EXCELLENCE

Claxton K, Rigging S, Ginnelly L, Griffin S, McCabe C, Phillips Z, Tappenden P and Wailoo A
University of York, York, United Kingdom; University of Sheffield, Sheffield, United Kingdom; University of Nottingham, Nottingham, United Kingdom.

Purpose: To demonstrate the benefits and feasibility of value of information analysis to support research recommendations made by the National Institute for Clinical Excellence (NICE).

Methods: A series of six case studies was selected from recent technology appraisals conducted by NICE. The case studies included: clopidogrel and dipyridamole in secondary prevention (CD); glycoprotein antagonists (GPA); screening for age related macular degeneration (AMD); neuraminidase inhibitors (NH) for influenza; liquid-based cytology (LBC) and beta interferons for multiple sclerosis (BMS). The case studies were broadly consistent with the recent NICE guidance on reference case analysis and included a probabilistic decision analytic model. In each case a re-analysis using value of information analysis was conducted. The Expected Value of Perfect Information (EVPI) surrounding each decision problem for the UK population, and the EVPI associated with particular model inputs was established using appropriate non-parametric methods.

Results: The re-analysis of each case study was completed within 4 weeks and the results were used to inform the recommendations of the NICE Research and Development Committee. The value of research differed substantially across the 6 technology appraisals (EVPI ranging from £2.8m to £905m). In some cases the analysis indicated that the original research recommendations should not be regarded as a priority, e.g., the EVPI surrounding LBC was zero (£2.8m). In other cases it indicated that additional research should be commissioned, e.g., the EVPI surrounding CD for stroke patients, and amyotrophic lateral sclerosis (£175m respectively). The analysis also indicated which comparators should be included, e.g., clinical trials of clopidogrel and ASA-MR-dipyridamole maybe worthwhile (EVPI=£600m) but MR-dipyridamole could be excluded. The analysis also indicated which patient sub-groups should be included, e.g., further trials of treatment for AMD were more valuable for those with lower starting visual acuity. The case studies highlighted a number of general methodological issues including consideration of all comparators, synthesis of direct and indirect evidence, and considering structural as well as parameter uncertainty.

Conclusions: Value of information analysis can be conducted in a timely way, which can inform the research recommendations made by NICE. It can be used to identify circumstances where decisions should be made on additional evidence and indicate the type of research which should be required including: the comparators; patient subgroups; and follow-up.

ORAL CONCURRENT SESSION A - COST EFFECTIVE ANALYSIS: METHODS
THE RETURN OF THE FIVE YEAR PLAN: MATHEMATICAL PROGRAMMING FOR OPTIMAL ALLOCATION OF RESOURCES
Eggington S, Claxton K and Sculptor M
University of York, York, United Kingdom; London School of Hygiene and Tropical Medicine, London, United Kingdom.

Purpose: To develop a mathematical programming framework to optimise the allocation of resources across healthcare technologies subject to inter-temporal budgetary and other constraints and apply it to policy relevant decisions.

Methods: Healthcare decisions based on the incremental cost-effectiveness ratio (ICER) cannot identify the true opportunity cost the technologies which should be displaced of implementing a new, more costly technology. A mathematical programming framework was developed to optimise the allocation of resources across health technologies subject to inter-temporal budgetary and other constraints. This is applied to a policy problem using examples relevant to the National Institute for Clinical Excellence (NICE) for England and Wales. To determine the optimal level of implementation of each health technology across patient groups, the optimisation problem is characterised in two ways: (i) as a 0-1 Mixed Integer Linear Program (0-1 MILP) or (ii) as a 0-1 Mixed Integer Non-Linear Program (0-1 MINLP). Annual budgets for pharmacueticals and other healthcare costs are defined for each time period. Horizontal equity concerns are incorporated as indivisibility constraints by restricting the decision variables to be 0-1 integer values. Non-linearity arises where fixed costs or other non-constant returns to scale are identified. Shadow prices are generated for each constraint. In addition, the value gained by trading budgets over time can be explored.

Results: The data required to implement the model was extracted from the 6th and 7th wave of appraisals considered by NICE, showing that the model is feasible. The optimal level of implementation across patient groups for each health technology was obtained. The shadow price of each budget for each year was estimated and can be interpreted as the value, in terms of health benefits of relaxing (or trading) these constraints. The opportunity costs of a range of possible horizontal equity concerns were expressed in terms of health benefit foregone. Alternative budgetary rules were also evaluated and showed that gains in population health can be made by allowing budgets to be traded over time.

Conclusion: The mathematical programming framework which has been developed to optimise the allocation of resources across health technologies can be applied to policy-relevant situations. The framework is shown to provide a robust and transparent process for social decision making in public health.
ORAL CONCURRENT SESSION A - COST EFFECTIVE ANALYSIS: METHODS LIMITATIONS OF ACCEPTABILITY CURVES FOR PRESENTING UNCERTAINTY IN COST-EFFECTIVENESS ANALYSES

Groot Koerkamp B, Hunink M, Stijnen T, Hammitt J, Kunst K and Weinstein M
Harvard School of Public Health, Boston, MA; Erasmus Medical Center, Rotterdam, Netherlands

Purpose: Clinical journals increasingly present uncertainty about the cost and effect of health care interventions using cost-effectiveness acceptability curves (CEACs). CEACs present the probability that each competing alternative is optimal for each value of the cost-effectiveness threshold. Our objective is to evaluate the limitations of CEACs for presenting uncertainty in cost-effectiveness analyses.

Methods: We explored many hypothetical parametric distributions of incremental cost and effect, allowing asymmetrical distributions as well as between cost and effect. For each joint distribution we compared different presentations of uncertainty, such as: CEACs, credible intervals on the incremental net benefit, and value of information.

Results: (1)Minimizing the probability of cost-effectiveness may result in a different ranking of interventions than maximizing expected benefits when distributions of net benefits are asymmetric. A risk-neutral decision maker is interested in the latter ranking. (2)A risk-averse decision maker cares about both the probability and the consequences of making the wrong decision. Therefore, he may prefer an intervention with a lower probability of cost-effectiveness to avoid a small probability of a catastrophe. CEACs do not inform about the consequences of making the wrong decision. (3)CEACs mix the magnitude and the precision of the mean incremental net benefit. Consequently, the medical importance suggested by, for example, a 90% probability of cost-effectiveness is ambiguous: it can reflect a huge, though imprecise difference in net benefit (e.g., a promising new cancer treatment) or a small but precisely defined difference hardly justifying the cost of implementation. (4)Because magnitude and precision are mixed, evidence presented as CEACs is difficult to synthesize with other qualitative, quantitative or subjective evidence and risk-attitude. Credible intervals on the incremental net benefit present magnitude and precision separately. (5)CEACs are typically equivocal about the value of information: the exact same CEAC can represent a decision with a high or a low value of information.

Conclusion: Both for guiding immediate decisions and for prioritizing information collection, these considerable drawbacks of CEACs should make us rethink their use in communicating uncertainty. A more informative presentation of uncertainty would be the credible/confidence intervals of incremental cost, effect and net benefit, together with the total expected value of perfect information.

ORAL CONCURRENT SESSION B - PUBLIC HEALTH 2 PREDICTING OPTIMAL ANTHRAX RESPONSE PARAMETERS: IMPACT OF INCIDENCE

Huppat N, Neukermans C, Cuomo J, Koshy M and Callahan M
Weill Medical College, Cornell University, New York, NY; RAND, Santa Monica, CA

Purpose: The incubation period and consequent epidemic curve caused by a large-scale anthrax exposure is a critical parameter. The extent to which a civilian population setting remains poorly understood and hotly debated. We investigated the impact of different epidemic curves to determine the minimum safe detection time and response time to achieve 95% disease prevention through mass antibiotic prophylaxis.

Methods: Using a state transition model of medical prophylaxis for bioterrorism agents (MDM’03) we estimated the proportion of patients exposed to an infectious dose of aerosolized anthrax. We explored many hypothetical parametric distributions of incremental cost and effect, allowing asymmetrical distributions as well as between cost and effect.

Results: (1)Minimizing the probability of cost-effectiveness may result in a different ranking of interventions than maximizing expected benefits when distributions of net benefits are asymmetric. A risk-neutral decision maker is interested in the latter ranking. (2)A risk-averse decision maker cares about both the probability and the consequences of making the wrong decision. Therefore, he may prefer an intervention with a lower probability of cost-effectiveness to avoid a small probability of a catastrophe. CEACs do not inform about the consequences of making the wrong decision. (3)CEACs mix the magnitude and the precision of the mean incremental net benefit. Consequently, the medical importance suggested by, for example, a 90% probability of cost-effectiveness is ambiguous: it can reflect a huge, though imprecise difference in net benefit (e.g., a promising new cancer treatment) or a small but precisely defined difference hardly justifying the cost of implementation. (4)Because magnitude and precision are mixed, evidence presented as CEACs is difficult to synthesize with other qualitative, quantitative or subjective evidence and risk-attitude. Credible intervals on the incremental net benefit present magnitude and precision separately. (5)CEACs are typically equivocal about the value of information: the exact same CEAC can represent a decision with a high or a low value of information.

Conclusion: Both for guiding immediate decisions and for prioritizing information collection, these considerable drawbacks of CEACs should make us rethink their use in communicating uncertainty. A more informative presentation of uncertainty would be the credible/confidence intervals of incremental cost, effect and net benefit, together with the total expected value of perfect information.

ORAL CONCURRENT SESSION B - PUBLIC HEALTH 2 USE OF CATALYTIC MODELING TO ESTIMATE HEPATITIS A INCIDENCE IN A LOW ENDEMICITY COUNTRY: IMPLICATIONS FOR MODELING IMMUNIZATION POLICIES

Chen H, Bauch C, Poty R, Pham B and Kahn M
Glaxo SmithKline Canada, Oakville, ON, Canada; University of Guelph, Guelph, ON, Canada; University Health Network, Toronto, ON, Canada

Purpose: In order to evaluate the benefits of Hepatitis A (HA) universal immunization programs, it is important to estimate the true incidence of infection. Because HA infection is significantly underreported, it is necessary to use serologic data and epidemiological methods to arrive at an estimate of true incidence.

Methods: The catalytic model used (e.g., Armstrong & Bell, 2002) involved several steps: 1) obtaining seroprevalence estimates; 2) obtaining reported incidence estimates; 3) fitting a model that relates prevalence to reported incidence (i.e., factors for underreporting, and change in incidence over time); and 4) deriving estimates of the latter factors from the fitted model.

Results: HA incidence data were obtained from the National Notifiable Disease Registry (Step-2). Average reported incidence (1988-2000) was adjusted for the probability of remaining at risk (i.e., 1 - prevalence of past infection) and being asymptomatic when infected (i.e., derived from a published model estimating the age-specific likelihood of jaundice, given HA infection) (Step-3).

Conclusion: Using the catalytic model for estimating incidence in Canada is feasible. The catalytic model can be further refined and validated for use in other countries to improve our understanding of HA transmission and disease trends.
ORAL CONCURRENT SESSION A - COST EFFECTIVE ANALYSIS: METHODS
PARAMETRIC SENSITIVITY ANALYSIS FOR CANCER SURVIVAL MODELS USING LARGE-SAMPLE NORMAL APPROXIMATIONS TO THE BAYESIAN POSTERIOR DISTRIBUTION

Hazen G
Northwestern University, Evanston, IL

Purpose: Decision-analytic models of cancer screening and treatment must include sub-models of cancer survival post diagnosis. One simple model postulates a probability, p, of permanent cure and an excess annual mortality rate n for those not permanently cured. Analysts can choose a combination (p, n) to accurately fit cause-specific survival data from the SEER cancer registry. However, if a model sensitivity analysis on (p, n) is desired, it is not obvious what neighborhood of the estimated (p, n) should be explored. We address this question by using Bayesian analysis to derive the posterior distribution of (p, n) given survival data.

Methods: Bayesian theory states that the prior distribution of (p, n) is approximately bivariate normal with mean equal to the posterior mode and covariance matrix equal to the Hessian of the log-posterior density. This approximate posterior distribution can be used to guide a sensitivity analysis.

Results and Conclusions: For stage-II ovarian cancer, the posterior distribution of (p, n) given SEER survival data is approximately bivariate normal with mean and variance 0.62/0.08 for p and 0.14/0.30 for n, and correlation 0.85. For sensitivity analysis, a representative one-dimensional posterior approximation can be explored by varying the largest principal component of this distribution within 2 SDs of its mean with the smaller principal component fixed at its mean. The band about the resulting survival curve for a 65-year-old female is shown below. We present similar results for other stages of ovarian cancer.

ORAL CONCURRENT SESSION B - PUBLIC HEALTH 2
DOES LEAD-TIME FROM INFECTION TO PID DEVELOPMENT MATTER IN CHOOSING AN STD SCREENING INTERVAL? A COST-EFFECTIVENESS ANALYSIS

Smith K, Cook R and Roberts M
University of Pittsburgh, Pittsburgh, PA

Purpose: Pelvic inflammatory disease (PID) is a major cause of disease among young women. As a preventive measure, most experts recommend screening for sexually transmitted disease (STD) every 12 months, while some argue for 6 month screening in higher risk women. However the lead-time from STD acquisition to PID development is unknown and its influence on screening interval and cost-effectiveness is unclear.

Methods: Using a Markov decision model, we estimated the incremental cost/QALY gained by chlamydia and gonorrhea screening every 6 or 12 months compared to no screening. Our base case analysis examined high-risk young women over a 2-year time horizon (25% infection risk, 70% asymptomatic), 5% PID risk with 12 month screening, and 25% PID complication risk based on population-based data) using peaked PID risk distributions for lead-times from 1-12 months. Lower risk women and differing PID risk distributions were examined in sensitivity analyses. Other risk, cost, and utility data were obtained from the literature.

Results: Compared to no screening, 12 month screening is cost saving for lead-times of 9-12 months and costs $21000/1000 QALY at 18.300/1000 QALY gained at shorter lead-times. Compared to 12 month screening, 6-month screening costs less than $47,000/QALY for PID lead-times from 1-12 months (range $15,100 to $46,700 per QALY gained at shorter lead-times). When the incidence of infection decreases to 5%/yr, costs/QALY for 6 month screening are greater than $75,000 for all lead-times (range $75,000 to $129,200). Other PID risk distributions showed similar relative insensitivity to PID lead-time.

Conclusion: Based on our analysis, uncertainty about the lead-time between infection and PID development or the PID risk distribution is not a significant factor in choosing a screening interval; the baseline infection rate is most important. Screening for chlamydia and gonorrhea every 6 months is economically reasonable in high-risk women, but expensive in lower risk groups.

ORAL CONCURRENT SESSION A - COST EFFECTIVE ANALYSIS: METHODS
$50,000 PER QALY: INERTIA, IDENTITY, OR IRATIONALITY?

Beatrice R and Roberts M
University of Pittsburgh, Pittsburgh, PA

Purpose: $50,000 per Quality-Life Year (QALY) is a threshold commonly used to delineate cost-effectiveness, but if it is controversial because it lacks theoretical basis, is not inflation-adjusted, and does not consider important societal factors (e.g. resources, technology growth). Therefore, our objective was to determine what cost-effectiveness thresholds are compatible with varying assumptions regarding health care resources and technology growth.

Methods: We developed a computer simulation that represents how individuals in the US interact with health care services. Use of a service produces an incremental gain in duration of life commensurate with its expenditure and cost-effectiveness. The simulation permits a cost-effectiveness threshold to be designated, which then constrains which available services are offered. If needs for offered services exceed allocated resources, services are excluded randomly without further regard to cost-effectiveness, and allocated resources may not be spent efficiently. If allocated resources exceed the needs for offered services, some resources may be unused. Data sources were based on published cost-effectiveness analyses of 128 interventions, and US government surveys. We examined thresholds of $200,000, $50,000, $100,000, and unlimited $/QALY. All costs were in 2002 US dollars.

Results: 44% of health care services cost less than $200,000/QALY, and 78% cost less than $100,000/QALY. In the base case analysis (allocation resources = 4% of current US GDP), mean life expectancy (LE) was highest (74.9 years) without use of any cost-effectiveness threshold because each threshold resulted in some allocated resources being unused. When allocated resources were decreased to 10% of GDP, LE was highest using a threshold of $100,000/QALY (73.9 years) because resources were partially unused with lower thresholds. (e.g., 33% of resources unused with $200,000/QALY LE 78.5 years) and were used inefficiently with higher thresholds (unlimited $/QALY LE71.2 years). When allocated resources were decreased to 5%, LE was highest (76.2 years) with the lowest examined threshold ($20,000/QALY) because resources were used completely and more efficiently than with higher thresholds. When technology growth doubled the volume of available services but resources stayed constant, LE was highest (76.8 years) with a threshold of $50,000/QALY.

Conclusion: Optimal cost-effectiveness thresholds decrease with fewer resources or greater technology growth. Thresholds established without regard to these factors may lower life expectancy substantially compared to optimal thresholds, and sometimes compared to no threshold at all.

ORAL CONCURRENT SESSION B - PUBLIC HEALTH 2
EVALUATING THE U.S. FOLIC ACID FORTIFICATION POLICY: DID WE SUCCEED?

Bentley T, Willett W, Weinstein M and Kurtz K
Harvard School of Public Health, Boston, MA

Purpose: The 1998 mandate to fortify enriched grain products with folic acid in the US aimed to help prevent neural tube defects (NTDs) by increasing 50% the proportion of women of childbearing age consuming greater than 400 micrograms (mcg) per day of folic acid. Limits on fortification levels are necessary because of concerns about masking of B-12 deficiency in older populations. Our analysis estimates the increase in folate intake after fortification.

Methods: We analyzed food, supplement, and total folate intake by gender, age (15-34, 35-64, and 65+), and race/ethnicity (non-Hispanic whites, non-Hispanic blacks, and Mexican-Americans) from two National Health and Nutrition Examination Surveys (NHANES): 16,794 subjects from NHANES III (1988-1994) and 4,831 from NHANES 1999-2000. We used a one-way analysis of variance on a two-measure subsample from NHANES III to adjust food folate intake distributions for measurement error. We compared pre- and post-fortification population-based distributions of total folate intake and proportions of the population consuming more than 400 and 1,000 mcg/day of total folate.

Results: Overall, daily food and total folate intake increased by approximately 100 mcg/day after fortification. The proportion of younger women consuming greater than 400 mcg/day of folate has increased since fortification, but has not yet reached the 50% target: 28% (pre-fortification) vs. 33% (post-fortification) of 15-34-year-old whites had intake >400 mcg/day; 19% vs. 23% of blacks; and 15% vs. 28% of Mexican-Americans. Among older populations (ages 65+) who may be at risk of B-12 deficiency masking, the percent who are consuming over 1,000 mcg/day (the tolerable upper limit) has increased after fortification for whites and black males, but has remained unchanged for black females and has decreased for Mexican-Americans: 2% vs. 4% (pre vs. post) for white males and females; 1% vs. 3% for black males; 5% (no change) for black females; 6% vs. 2% for Mexican-American males; and 3% vs. 0% for Mexican-American females.

Conclusions: Since fortification, folic acid intake among the U.S. population has increased, with substantial variations by age, gender, and race. Targeted supplement-use interventions among women of childbearing age may be needed to further increase the proportions of these women consuming greater than 400 mcg/day of folic acid.

© 2005 Society for Medical Decision Making. All rights reserved. Not for commercial use or unauthorized distribution.
ORAL CONCURRENT SESSION A - CLINICAL STRATEGIES AND GUIDELINES
SHOULD PROPHYLACTIC SURGICAL FUSION BE OFFERED TO PATIENTS WITH RHEUMATOID ARTHRITIS AND CERVICAL INSTABILITY BEFORE DEVELOPMENT OF NEUROLOGICAL DEFICITS? A DECISION ANALYSIS
Lisa Y. Sandor B and Mantal A
University Health Network, Toronto, ON, Canada

Background: Surgical fusion is routinely proposed to patients with rheumatoid involvement of the cervical spine in the presence of neurological symptoms. Successful outcomes of surgery have prompted surgeons to advocate intervention in patients with pain alone in the absence of neurological signs.

Objective: To evaluate the benefits of a prophylactic measure for patients with painful cervical instability in the absence of neurological signs (Ranawat I) [PROPH], versus only for those whose disease progresses to neurological involvement (Ranawat >=II) [NEURO].

Methods: We developed a decision analysis model with a comprehensive representation of the transitions of patients with cervical instability between Ranawat stages, surgical complications and death. A systematic search of Medline was performed to obtain estimates of events in the model. We excluded studies that contained less than 10 patients, and were published before 1985, after which CT and MRI imaging modalities were routinely used. Primary outcomes considered were the progression from Ranawat I to Ranawat >=II and mortality. We adopted a time horizon of 7 years, the average time of observation in the supporting literature. Sensitivity analysis was performed on key variables.

Results: Seven studies met eligibility criteria, providing input to the following baseline variables: 1) progression from Ranawat I to >=II: 57% [NEURO] vs. 8% [PROPH]. 2) chance of non-permanent surgical complications: 14% [both PROPH and NEURO]; 3) chance of permanent surgical complications: 2% [PROPH] vs. 5% [NEURO]; 4) improvement to Ranawat I with NEURO: 30%. Neither was associated with increased mortality directly related to surgery. Adopting PROPH would lead to a higher proportion of patients remaining in Ranawat I or 0 (free of pain) (92%) as compared to NEURO (57%). Sensitivity analysis shows that preference of PROPH over NEURO is robust.

Conclusion: Results of this decision analysis indicate that PROPH is a realistic option, to be discussed with patients who present with cervical instability. With NEURO, RI is realized only by a minority of patients once they progress to RI or RII. High-quality, prospective, observational studies are needed to corroborate these findings. Cost-effectiveness and quality of life assessment would provide added values.

ORAL CONCURRENT SESSION B - HEALTH ECONOMICS
HEALTH, WEALTH, AND HAPPINESS: FINANCIAL RESOURCES BUFFER SUBJECTIVE WELL-BEING AFTER THE ONSET OF A DISABILITY
Smith D, Langa K, Kabets M and Ubel P
University of Michigan, Ann Arbor, MI

Purpose: Studies have shown that money has much less effect on happiness than people believe it does. For example, real income has quadrupled in the US over the past 5 decades, and yet the overall happiness level remains about the same. We examined the relationship between financial status and subjective well-being for people with new disabilities (e.g., acute myocardial infarction). We compared people with new disabilities who were in poorer health, and especially after the onset of a serious physical disability. Specifically, greater net worth was associated with a decline in well-being that was several times larger than the modest decline experienced at the decline in well-being that was several times larger than the modest decline observed for people above the median.

Results: People who report a new disability, accounting for 3 times as much variance in well-being as for people who were not disabled (p < .01).

Conclusions: Financial resources are a larger predictor of well-being for people who are in poorer health, and especially after the onset of a serious physical disability. Specifically, greater net worth appeared to buffer well-being after the onset of a disability, people below the median in net worth experienced a decline in well-being that was several times larger than the modest decline observed for people above the median.

ORAL CONCURRENT SESSION A - CLINICAL STRATEGIES AND GUIDELINES
EVALUATING CLINICAL GUIDELINES AND PREDICTIVE INSTRUMENTS FOR ACUTE CARDIAC ISCHEMIA
1University of Texas at Arlington, Arlington, TX; 2Yale University, New Haven, CT; 3Mayo Clinic, Rochester, MN

Purpose: To compare clinical guidelines and predictive instruments for acute cardiac ischemia.

Methods: Two samples of emergency-room patients presenting with non-traumatic chest pain or pressure were classified as low, intermediate, or high risk according to American Heart Association/American College of Cardiology unstable angina guidelines, Agency for Health Care Policy and Research (AHCPR) guidelines, Acute Cardiac Ischemia Time Intensive Predictive Instrument (AC- TIP), and physicians’ triage decisions (discharge, ward or monitored bed, or cardiac intensive care). Sample 1 consisted of every fifth consecutive patient eligible for enrollment (N = 1004) and Sample 2 consisted of patients for whom physicians had returned a detailed questionnaire (N = 1028). Mortality was 1.9% overall, 7.2% White-Non-Hispanic, 18% White-Hispanic, and 10% other. Of the 1028, 333 patients returned to the hospital for a cardiac procedure (e.g., percutaneous transluminal coronary angioplasty or coronary artery bypass graft); and 75 received a cardiac diagnosis (e.g., acute myocardial infarction, unstable angina, or stable angina) within a year of their initial visit. Of the 1004, 200 patients returned within a year. Sequence logistic regression analyses with occurrence of a cardiac procedure as a criterion variable were constructed, inputting coronary artery disease (CAD) risk level for the older and newer guidelines, acute myocardial infarction (AMI) risk level for the older and newer guidelines, as well as level of triage for initial visit and AC-TIP scores as predictors. Cardiac diagnosis was similarly analyzed.

Results: Occurrence of a cardiac diagnosis was predicted by the older AHCPR guidelines but not by the newer guidelines. Physicians’ initial triage decisions, AC-TIP scores, AHCPR risk levels, and newer guidelines’ AMI risk level predicted procedures—and such contributed unique variance. Examining procedures and diagnoses separately, the older guidelines consistently outperformed the newer guidelines in assigning higher levels of risk to cardiac outcomes.

Conclusions: Although the newer guidelines were designed to improve on the older guidelines, empirical comparisons revealed that the newer guidelines led lower predictive validity for the occurrence of subsequent cardiac diagnoses. Pandemicly, although the newer guidelines increased levels of risk, lack of resolution contributed to difficulty in predicting cardiac outcomes. These results indicate that changes in guidelines should be empirically evaluated before widespread implementation. They also indicate that the use of multiple, related guidelines in concert with physicians’ judgments can improve prediction of cardiac risk.

ABSTRACTS
E21

© 2005 Society for Medical Decision Making. All rights reserved. Not for commercial use or unauthorized distribution.
ORAL CONCURRENT SESSION A - CLINICAL STRATEGIES AND GUIDELINES

CLINICAL PREDICTORS DETERMINE THE VALUE OF GENETIC TESTING IN THE PREDICTION OF SEVERE PROGRESSION IN PATIENTS WITH RHEUMATOID ARTHRITIS

Janssens A, Steyerberg E, Van Duijn C and Criswell L

Erasmus MC, Rotterdam, Netherlands; University of California, San Francisco, San Francisco, CA

Objective: Genetic factors influence disease progression, but may also predispose to other, non-genetic factors, such as age at symptom onset. The additional value of genetic information to the prediction of prognosis may therefore differ in the presence or absence of those clinical predictors. We aim to evaluate the additional value of HLA-DRB1 alleles encoding the rheumatoid arthritis (RA) shared epitope (SE) in predicting radiographic damage in RA conditional on non-genetic predictors.

Methods: Demographic characteristics, baseline clinical characteristics and SE status were available for 160 Caucasian women with RA. Univariate analyses were performed for the selection of non-genetic predictors. The additional value of the SE to the prediction of radiographic damage was determined by a newly developed method estimating the likelihood ratio (LR) of SE status conditional on other predictors. These LRs were used to calculate the sensitivity and specificity of the SE testing for each woman in the study.

Results: Overall, the LR of SE presence (LR+) was 1.4, the LR of SE absence (LR-) 0.4, the odds ratio (OR) 4.0, sensitivity 0.65 and specificity 0.88. The SE was a significant predictor of radiographic damage (multivariate OR 1.5 [95% CI 1.2, 1.9]; p = 0.001). Next to Health Assessment Questionnaire (HAQ) score (OR 4.4 per 1 point [9.1, 11.0]), rheumatoid factor (RF) positivity (OR 2.4 [1.7, 3.4]; p = 0.001), disease duration (OR 1.5 per 5 years [1.1, 1.9]) and family income (OR 0.6 per 1000 US$ [0.6, 1.0]). Of these predictors, HAQ score (p = 0.01) and RF positivity (p < 0.001) were also associated to the SE. At the individual patient level, taking into account the risk profile of individual women, the LRs varied from 1.0 to 2.3, the LR- from 0.2 to 0.6, the sensitivity from 0.60 to 1.00 and the specificity from 0.89 to 0.78. In women who were RF positive, the average LR+ and LR- were 1.5 and 0.3, and the average sensitivity and specificity were 0.89 and 0.36, whereas in those RF negative, these figures were 1.9, 0.4, 0.78 and 0.55, respectively.

Conclusion: The value of genetic information in the prediction of prognosis depends on the risk profile of the patient. This variation is explained by correlation between the genetic factor and clinical predictors.

ORAL CONCURRENT SESSION B - HEALTH ECONOMICS

THE IMPACT OF DIABETES ON EMPLOYMENT IN A COHORT OF MIDDLE AGED ADULTS: FINDINGS FROM HEALTH AND RETIREMENT STUDY

Tuncelli K, Bradley C, Elston Lafata J and Nerenz D

Henry Ford Health System, Detroit, MI; Michigan State University, East Lansing, MI

Purpose: To understand the economic impact of interventions designed to prevent the onset of diabetes, one needs to understand the impact of diabetes on employment. We examined the effect of diabetes on labor market outcomes among older adults before the usual retirement age of 65.

Methods: Using data from the 1992 and 1994 waves of the Health and Retirement Study, among working individuals aged 51-61 in wave 1, we estimated the effect of diabetes in wave 1 on the probability of working in wave 2 using a multivariable probit model. In addition to socio-demographics, we controlled for health insurance, occupation, and the presence of chronic conditions. Among those who were working in wave 2, we estimated the effect of diabetes on weekly hours worked, and work-loss days using multivariable ordinary least squares regressions, and on the presence of work limitations using a multivariable probit model. We undertook all analyses separately for men (N = 3581) and women (N = 4080). A P value of 0.05 was considered statistically significant.

Results: Men with diabetes were 7.5 percentage points (pp) less likely to work than men without diabetes. The effect of diabetes for women was 4.9 pp. Among both women and men who work, the presence of diabetes did not affect usual weekly hours worked. However, women with diabetes had 2.3 more work-loss days per year than women without diabetes. Furthermore, the presence of diabetes increased the probability of work limitations for both men (6.4 pp) and women (7.2 pp).

Conclusions: Diabetes causes both men and women to leave their job before the usual retirement age. The negative effect of diabetes on employment is larger among men than women. However, the number of work-loss days and the presence of work limitations are greater among women than men. One possibility is that women with diabetic complications might be less likely to have workforce than diabetic men with complications. This study provides partial evidence that diabetes negatively affect work productivity of employees who survive and remain working. Although diabetics do not decrease hours worked, it does increase the number of work-loss days and the presence of work limitations. Diabetes affects patients, employers, and society not only by reducing employment but also by contributing to work loss and limitations.

ORAL CONCURRENT SESSION B - HEALTH ECONOMICS

COMBINING REGISTRY, PRIMARY, AND SECONDARY DATA SOURCES TO IDENTIFY THE IMPACT OF CANCER ON LABOR MARKET OUTCOMES

1 Bradley C, Neumark D, Oberst K, Brennan S, 2 Schenk M

1 Michigan State University, East Lansing, MI; 2 Public Policy Institute of California, San Francisco, CA; 3 Wayne State University, Detroit, MI

Purpose: We developed a research agenda to study the labor market outcomes attributable to detection and treatment for cancer, and developed a research design and data collection strategy that improves upon other research addressing these questions.

Methods: We identified an inception cohort of women with breast cancer (n = 490), men with prostate cancer (n = 294), and corresponding control groups, and followed their labor market outcomes for a period of 18 months. The main outcomes were changes in employment, hours worked and wages, which prior research have shown to be correlated with health, and reductions in which were a major contributor to the productivity losses. We also critically assessed how well our study design and data collection strategy accomplished its objectives.

Results: The study performed well in recruiting subjects from a population-based data source, constructing a control group from the Current Population Survey (CPS), and implementing a longitudinal design, with 90% subject retention rate, that reflects the natural course of cancer’s ill effects within the first year following diagnosis. Using the CPS, we demonstrated that simply reporting employment transitions from a cohort of cancer survivors without a control group overstates the non-employment effect of cancer. Nevertheless, the CPS control sample was not a perfect match to the cancer sample in terms of income and socioeconomic status. Overall, the negative effect of cancer was greatest during months following diagnosis relative to the control sample. At 12 and 18 months following diagnosis, many subjects return to work. Breast cancer’s non-employment effect appears to be twice as large for African-American women.

Conclusions: Although our research is specific to cancer, the issues raised and addressed are relevant to the study of a number of chronic diseases. If the economic consequences of disease are to be seriously examined so that their results are meaningful for policy makers and treatment providers, researchers investigating labor market outcomes must take considerable care in their study design, instrument validity, sample selection, and data collection, as well as in the execution of the analysis. The methods used to estimate the economic and other economic attributes attributable to these conditions require careful scrutiny so that reliable findings can be used to shape health care decisions and policy.
ORAL CONCURRENT SESSION A - CLINICAL STRATEGIES AND GUIDELINES
PEDIATRICIANS’ DECISION MAKING: A RANDOMIZED CONTROLLED TRIAL OF DECISION SUPPORT WITH SENSITIVITY AND SPECIFICITY

Nork C, Koepssel T, Doctor J and Christakis D
University of Washington, Seattle, WA; University of Washington, Seattle, SC

Purpose: To determine how presenting pediatricians with a diagnostic test's sensitivity and specificity affects their estimate of the posttest probability.

Methods: We mailed a four-page questionnaire to 1552 pediatricians practicing in the U.S. randomly selected from the AMA master list, and set up to 4 times to non-responders. The questionnaire presented a vignette of a 3-month old girl with perioral cyanosis and a hacking cough, subjects were instructed to assume that the likelihood that she had pertussis was 30%. They were then told which the direct fluorescent antibody (DFA) for pertussis sent on her was negative. We randomized subjects to one of three decision support groups (control, technical or non-technical decision support). Controls received no additional information. The technical decision support group was presented the sensitivity (50%) and specificity (95%) of the DFA for pertussis. The non-technical decision support group was presented the DFA sensitivity and specificity, and a non-technical explanation of these terms. Subjects were asked to estimate the likelihood that this patient had pertussis.

Conclusions: We conducted a test to evaluate how decision support affected estimates of posttest probability and chi-square tests to compare the proportions in each group who chose a probability of 50%.

Results: 653 subjects returned completed surveys: 202 in the control, 231 in the technical and 220 in the non-technical decision support groups. The estimated response rate was 52%. 85% of participants were board certified in pediatrics and their mean age was 43 years. There were no significant differences between randomization groups in any subject characteristic. The correct posttest probability was 18%, but it was substantially overestimated in all three groups (mean 41, median 50, range 0-100, inter-quartile range 15-50). Subjects who received technical decision support had a mean posttest probability not significantly different from controls (41% vs. 38%, p = 0.16). In contrast, subjects who received non-technical decision support estimated a significantly higher posttest probability than controls (47% vs. 38%, p < 0.001). Subjects who received technical or non-technical support were more likely than controls to choose a probability of 50% (17% vs. 38%, p <0.0005; 17% vs. 41%, p < 0.0005).

Conclusions: Presenting non-technical decision support about the sensitivity and specificity of a diagnostic test worsened pediatricians’ estimates of disease probability. Decision support may have unexpected effects on medical decision making.

ORAL CONCURRENT SESSION B - HEALTH ECONOMICS
COST SHARING POLICIES AND HEALTH CARE DEMAND: LESSONS FROM CANADA

Li X, Goh D, Marra C, Lacalle D, Esoaki J and Anis A
University of British Columbia, Vancouver, BC, Canada; St. Paul’s Hospital, Vancouver, BC, Canada; Vancouver Coastal Health Research Institute, Vancouver BC, Canada

Purpose: Patient cost-sharing usually reduces overall health resource utilization. However, under the Canadian medicare system, where all health services except prescription drugs are free, the converse may be true. To test this hypothesis, we estimate healthcare demand elasticities (percent change in quantity demanded for each percentage change in price) to evaluate the impact of cost sharing policies for prescription drugs on overall health resource use among seniors with rheumatoid arthritis (RA) in British Columbia (BC), Canada.

Methods: Medication and physician visit data between 2001/01/01 and 2002/12/31 for all seniors were selected from a population-based RA cohort. Under the BC drug insurance program, prior to 2002 only seniors 65+ of their dispensing costs for drugs to an annual maximum of $200 (Plan A) after which drugs became free of charge. Starting in 2002, this plan was split into Plans A and A1 (Pre-maximum 2002 to 2003) and Plan B (a maximum of $275 and $400 for prescription to an annual maximum of $275 and $400, respectively). Only seniors who either reached or did not reach the annual deductible in both years were considered. Patients were classified into 4 groups based on reaching the annual deductible and their Plan. Own-price and cross-price elasticities were estimated using mixed effect models controlling for patient demographics and disease-related variables. The monthly rate of prescriptions filled and physician visits were assessed.

Results: A total of 5,227 patients were included in the study. Estimated own-price and cross-price elasticities are presented (Table). All four groups have negative own-price and positive cross-price elasticities. Positive cross-price elasticities of demand on physician visits with respect to the change of drug price suggested that when cost sharing for prescription drugs increased, so did the demand for physician visits. However, with increased cost of prescription drugs, the number of prescriptions filled fell.

Conclusions: In a predominantly publicly funded health care system, the introduction of market-based cost-sharing may not create such a patient cost-sharing might have the unintended impact of increasing overall utilization.

E23
ESTIMATING THE COST AND EFFICACY OF THREE REFERRAL STRATEGIES FOR HIV PARTNER COUNSELING AND REFERRAL SERVICES

We developed a probabilistic decision model to calculate costs and effectiveness. Our analysis included 16 studies from 5 different countries with data from 2868 participants. Cost and resource utilization data were obtained from health departments, national and provincial health services, and cost utilities were obtained from the societal perspective and $1,515 and $1,750 (95% CI 1,477-1,767 and 1,736-1,766) from the provider perspective.

CONCLUSIONS: PCRS, particularly provider PCRS, is an effective method of identifying HIV infection in persons who may have been exposed. Provider PCRS should be favored when considering both cost and effectiveness.

POSTER SESSION - PUBLIC HEALTH: METHODOLOGICAL ADVANCES

SYSTEMATIC META-ANALYSIS ON THE EFFICACY OF DISEASE MANAGEMENT PROGRAMS IN PATIENTS WITH CONGESTIVE HEART FAILURE

Purpose: To assess a partially-effective HIV vaccine in a cohort of 15-year old adolescent women in South Africa over 10 years in terms of HIV infections and deaths prevented in mothers and infants.

Methods: A Markov model is constructed for all 15-year old adolescents in South Africa followed for 10 years. Each adolescent can become HIV infected, pregnant, or die. This model's baseline output without the vaccine is calibrated with the AIDS demographic model used by the Actuarial Society of South Africa. A vaccine is introduced to reduce the HIV incidence rates of adolescents as well as vertical transmission to their infants through breastfeeding. At the end of 10 years, the number of HIV infections and deaths prevented in adolescents and infants is analyzed with a Monte Carlo simulation. In addition, when an adolescent becomes pregnant or develops AIDS, she has a probability of starting prophylactic antiretrovirals or starting highly active antiretrovirals (HAART).

RESULTS: If in four years South Africa were to adopt the use of prophylactic antiretrovirals to prevent perinatal HIV transmission in the peripartum period, an additional 31% of infant infections are prevented at year 1. If the majority of AIDS patients have access to HAART by 2006, the number of HIV-related deaths would dramatically drop (77%) but the number of adolescents living with HIV would increase (7%). In this setting, a 50% effective vaccine would not prevent deaths but decrease the number of adolescents ever infected by HIV by 65,160 (15%) and HIV infected infants by 19,150 (23%).

Conclusion: A partially-effective HIV vaccine consistently lowers the number of infected adolescents and infants by an additional 15-20% in our model in addition to the advantages gained from antiretrovirals. A partially-effective HIV vaccine has an immediate impact in the prevention of adolescents and infants in South Africa no matter the domestic policy implemented.

ABSTRACTS
POSTER SESSION - PUBLIC HEALTH; METHODOLOGICAL ADVANCES

AN ANALYSIS OF STATE DIFFERENCES IN HEALTH-RELATED QUALITY OF LIFE BETWEEN PERSONS WITH AND WITHOUT DIABETES
Imai K and Zhang F
Centers for Disease Control and Prevention, Atlanta, GA

Purpose: To analyze state variations in differences in health-related quality of life (HRQOL) between persons with and without diabetes in the U.S.

Methods: Individual-level data came from the Behavioral Risk Factor Surveillance System (BRFSS) 2001 survey. The BRFSS is a continuous, state-based, random telephone survey of more than 210,000 community-dwelling U.S. adults and aged 18 and older. HRQOL in the BRFSS assesses a person's perceived sense of well-being related to physical health, mental health, and activity limitations. Physical and mental health is measured by “number of days when physical or mental health was not good.” While activity limitation is assessed by “number of recent activity limitation days because of poor physical or mental health.” Data on state health care characteristics were taken from the Area Resource File. We used a two-step estimator where the first step consisted of estimation of state-level differences in HRQOL between persons with and without diabetes (controlling for individual-level demographic and socioeconomic characteristics), while the second step examined factors associated with the state-level differences estimated in the first step.

Results: Differences in HRQOL between persons with and without diabetes varied significantly across U.S. states and territories (see Table 1). Physicians’ and hospital densities were positively associated with differences in HRQOL between persons with and without diabetes. In particular, stronger associations were found for recent days of physical health. In contrast, HMD penetration was negatively associated with differences in recent days of physical health and activity limitation between the two populations.

Conclusions: There are significant variations in state-level differences in HRQOL between persons with and without diabetes, and part of these variations can be explained by state health care characteristics.

Table 1 Adjusted State-Level Differences in HRQOL between Persons with and without Diabetes

<table>
<thead>
<tr>
<th>Unhealthy Days</th>
<th>Physical</th>
<th>Mental</th>
<th>Physical or mental</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (days)</td>
<td>3.9</td>
<td>1.3</td>
<td>4.0</td>
<td>2.1</td>
</tr>
<tr>
<td>Standard Errors</td>
<td>0.9</td>
<td>0.7</td>
<td>0.9</td>
<td>0.7</td>
</tr>
</tbody>
</table>

POSTER SESSION - PUBLIC HEALTH; METHODOLOGICAL ADVANCES

QUALITY OF LIFE AMONG PEOPLE WITH SELF-REPORTED VISION PROBLEMS IN THE US, BRFSS, 2002
Kanjilal S, Beckles G, Narayan K and Saaddine J
Centers for Disease Control and Prevention, Atlanta, GA

Purpose: To describe the quality of life (QOL) of people who self-reported vision problems and to compare it with the QOL of people who self-reported other health problems.

Methods: We analyzed the 2002 Behavioral Risk Factor Surveillance System (BRFSS) survey. The BRFSS is a state-based annual ongoing telephone survey of the civilian non-institutionalized population, ages 18 and older. In 2002, 83,949 people were interviewed in 20 states that used an optometric module consisting of 11 questions about QOL. The analysis was restricted to the 365 people who reported vision problems (VP) as their primary health concern and the 14,812 individuals that self-reported one health condition out of a list of 13 (e.g., hearing problems, arthritis, cardiovascular diseases, diabetes, etc.) as their primary health concern. Data were analyzed in SAs-callable SUDAAN in order to take into account the complex sampling scheme of the survey. Logistic and linear regression models were run with the QOL indicators as dependent variables.

Results: The VP group was older and less employed than the “Other Health Problem” (OHP) group. For the QOL indicators, the VP group reported fewer days of pain (5.74 (SE: 1.09) vs. 11.17 (SE: 2.23), p < 0.0001) and of not enough rest (7.37 (SE: 0.88) vs. 11.54 (SE: 0.20), p < 0.0001) in the past month than the OHP group. The VP group also reported a longer duration of impairment (11.13 years (SE: 1.38) vs. 7.52 years (SE: 0.15), p < 0.01) and a higher percent needing assistance with routine tasks (47% (5.2%) vs. 30% (6.8%), p < 0.01). There was no significant difference in the days of depression in the past month between the VP and OHP groups (6.63 (SE: 0.88) vs. 7.13 (SE: 0.18), p = 0.67).

Conclusions: Though people with vision problems have a better quality of life than those who report a range of other health problems, the difference between the two groups is not very large for many QOL indicators. This suggests that although vision problems in general are often perceived as not being as serious as other health conditions, they result in an equally high burden on quality of life. Further research quantifying the QOL burden in economic utility terms will allow health professionals and policy makers to accurately prioritize the growing problem of visual impairment among many other health conditions.
POSTER SESSION - PUBLIC HEALTH; METHODOLOGICAL ADVANCES

SHOULD WE TREAT PEOPLE WITH OCULAR HYPERTENSION TO PREVENT GLAUCOMA?
Kymu S, Kass M and Gordon M
Washington University School of Medicine, Saint Louis, MO

Purpose: The medical treatment of ocular hypertension (OH) has been shown to be efficacious in preventing primary open angle glaucoma (POAG), the most common form of glaucoma. However, treatment of OH is controversial among ophthalmologists, with some arguing that treatment prior to evidence of POAG is a waste of resources. We conducted this investigation to determine under what conditions treatment of ocular hypertension is cost-effective.

Methods: A cost-utility approach was taken using a Markov decision model and data from the Ocular Hypertension Treatment Study (OHTS) augmented by a literature review. Four alternative treatment strategies were modeled: "Treat all persons with OH," "Treat people with OH and a 2% annual risk of POAG," "Treat people with OH and a 7.5% annual risk of POAG," and "Treat no one." Logistic regression using OHTS data was used to estimate the incidence of POAG among those treated and untreated, as well as the proportion of persons treated for each strategy. Precision of our results were tested with one- and two-way sensitivity analysis and Monte Carlo simulations.

Results: The "Treat 5%" strategy mildly dominated the strategy of "Treat no one." The "Treat 2%" strategy was cost-effective with an incremental cost-effectiveness ratio (ICER) of $80,950/QALY. Treatment of all persons with OH was not cost-effective with an ICER of $332,303/QALY. The treatment of all persons with OH was not cost-effective with an ICER of $332,303/QALY. The treatment of all persons with OH was not cost-effective with an ICER of $332,303/QALY. The treatment of all persons with OH was not cost-effective with an ICER of $332,303/QALY. The treatment of all persons with OH was not cost-effective with an ICER of $332,303/QALY.

Conclusions: The conclusionsvalidated the effectiveness of an additional tax increase on smoking behavior and consumer, and the effect of the tobacco consumption reduction on state and local tax revenues. We also estimate the one-year and long-term health benefits of the resultant tobacco consumption reduction in terms of prevented conditions and avoided health care costs. We also investigate the effect of investing part of the excise tax increase revenue into the state tobacco control program.

Results: A $1.50 cigarette excise tax increase would result in 360,000 new quitters and increase state tax revenues by $1.6 billion and local tax revenues by $78 million. In one year, this tax increase would prevent 104,000 heart attacks, 104,000 low birthweight births, 300 new pediatric asthma cases, and also save $19 million in health care expenditures. Long term, this tax increase would annually prevent 4,000 smoking-related deaths and also save $1.5 billion in health care expenditures. Designating 20% of this tax increase for the state tobacco control program would provide an additional $320 million in the first year to the tobacco control program but reduce the increases in state and local tax revenues to $1.1 billion and $64 million, respectively. However, this investment would create an additional 275,000 new quitters. These additional quitters would further prevent in one year 200 heart attacks, 100 strokes, 200 low birthweight births, 250 new pediatric asthma cases, and further save $14 million in health care expenditures. Long term, these additional quitters would further prevent 3,000 smoking-related deaths and $1.1 billion in health care expenditures.

Conclusions: Increasing the cigarette excise tax in California by $1.50 would significantly increase state and local revenues, increase the state’s health, and reduce health care expenditures. Designating 20% of the cigarette excise tax increase for the state tobacco control program would slightly reduce state and local revenues, but produce more quitters and further increase health benefits and health care cost savings.
POSTER SESSION - PUBLIC HEALTH; METHODOLOGICAL ADVANCES
GIS MAPPING TO MAPPING TO LOCATE ADHD-STIMULANT RISKS


The Ohio State University, Mount Gilead, OH; Children’s Research Institute, Columbus, OH; Children’s Hospital, Columbus, OH; The Ohio State University, Columbus, OH

Purpose: As the therapeutic use of a medication increases, risk of unintentional overdose, medication errors and unintentional overdose increases. The objective of this study is to identify the chorographic geographic areas where exposures to stimulant medications used in the treatment of attention deficit and hyperactivity disorders (ADHD) are prevalent.

Methods: All calls to the Regional Poison Control Center (RPCC) from 1999 to 2001 were logged into a computer database, and non-ADHD stimulant callers were used to control for penetrance of RPCC services in the area. Patient demographics, the nature of the call, site of exposure, clinical presentation and treatments received were documented. Patient zip codes were used to identify county of the caller, county classification was based on standard criteria (Department of Health) and 2000 census data was used to estimate the socio-economic parameters of the population. Geographic mapping (Maptitude®) was used to identify the populations at greatest risk (e.g., age, gender, geographic location) and to estimate the number of children in the community at risk based on age and published prevalence of ADHD in the United States.

Results: All calls (n = 153,190), exposures (e.g., ingestion, inhalation, dermal contact) (n = 116,888) and information calls (n = 14,361) reported to the RPCC were used in the analysis. Controlling for the total number of calls and population ADHD exposure calls were greatest in the largest urban county. Other large metropolitan areas, with populations greater than 50,000, were more likely to report ADHD stimulant exposures.

Conclusion: Through the combination of incident reporting and medical mapping, communities most in need of interventional programs to decrease the number of intentional and unintentional ADHD stimulant medication exposures may be identified.

POSTER SESSION - PUBLIC HEALTH; METHODOLOGICAL ADVANCES
PERINATAL SYPHILIS TRANSMISSION OUTCOMES OF RPR VERSUS RAPID TESTING IN FIELD CONDITIONS IN HAITI: RESULTS OF A DECISION ANALYTIC MODEL

Schackman BR, Neukermans C, Naette S, Joseph P and Fitzgerald D

Weill Medical College of Cornell University, New York, NY; GHESKIO Center, Port-au-Prince, Haiti; Weill Medical College of Cornell University, New York, NY

Purpose: In the United States 6-7000 women infected with HIV give birth every year. Although routine prenatal care includes HIV testing, 5% of women do not receive prenatal care and should have rapid HIV testing at the time of delivery because anti-retroviral chemoprophylaxis during labor or within 12 hours of birth reduces maternal-fetal HIV transmission from 25% to 0-12%. The OraQuick rapid HIV test has a 99.6% sensitivity and 100% specificity, however multiparous women can have false positive results.

Methods: To examine the risks and benefits of the OraQuick rapid HIV test in pregnant multiparous women.

Results: Data were obtained from searching Medline, CDC, National Vital Statistics Reports, the Cochrane Library and the Red Book to determine national data regarding annual pregnancies, absence of prenatal care, multiparity, HIV prevalence in women of child-bearing age, efficacy of chemoprophylaxis and cost of medications. OraQuick test characteristics (sensitivity 99.6% and specificity 93.3%) were based on FDA data. Results: In the US in 2002, 39.107 women received no prenatal care, and in an urban setting, 65.3% of pregnant women not receiving prenatal care were multiparous. The CDC reported HIV prevalence in women of childbearing age varies from 0.8% to 4.4%. For this prevalence range, of 33,103 multiparous women not receiving prenatal care, 267-1,335 women will be HIV-positive. The OraQuick would identify 206-1,329 of those HIV-infected, but also 2,318-2,207 uninfected would be false positives. Assuming a 25% maternal-fetal HIV transmission rate and that antiretroviral therapy reduces transmission by one-half, treating all OraQuick positive individuals lowers HIV transmission from 66 to 33 and 332 to 166 respectively. However, by treating all who test positive, 21 to 74 women would need to be treated with antiretroviral therapy to prevent one case of HIV transmission (NNT). Furthermore, 1.7 to 8 women without HIV would be treated for every women appropriately treated. Cost of $112 [SDV = $41.15, OraQuick = $83.9, Western Blot = $82] per positive test for a total of $317,084-846,752 or $2,811-9,608 per maternal-fetal HIV transmission prevented. One to six HIV-infected mothers would miss an opportunity to receive chemoprophylaxis.

Conclusions: In multiparous women, the OraQuick Rapid HIV test has decreased specificity, so some women without HIV would likely be treated to prevent maternal-fetal HIV transmission. Local are prevalences data should be obtained.
ABSTRACTS

POSTER SESSION - PUBLIC HEALTH: METHODOLOGICAL ADVANCES
MEASURING HOW PATIENTS USE INTERNET-BASED DECISION AIDS
Weber C1, Fortin J1, Landau C1, Cyr M1 and Col N1
1Harvard Medical School, Boston, MA; 2Rhode Island Hospital, Providence, RI; 3Brown Medical School, Providence, RI
Purpose: The Internet is an important vehicle for disseminating decision aids, yet little is known about how web-based decision aids are used. We sought to assess how women use a comprehensive web-based, interactive decision aid.

Methods: We developed the web-based Women’s Interactive System for Decisions on Menopause (WISDOM), which generates personalized risk reports (cardiovascular disease, breast cancer, hip fracture) and shows the impact of prevention options on absolute risks. WISDOM includes an embedded Markov model and a database of menopausal treatments that are accessed via an interactive symptom/treatment chart that can be sorted by treatment type, symptom concerns, and efficacy. Tiedot text explanations about treatments are included. Summary reports can be sent to clinicians. Healthy women between 45 and 65 years were recruited as part of a randomized trial evaluating the impact of WISDOM versus standard care. The navigation patterns of WISDOM users were captured in log files.

Results: To date, 69 women have enrolled, 35 were randomized to WISDOM, and 25 visited the website. The mean age was 51 years, 97% were white, and 86% were college graduates. Most (80%) completed all aspects of WISDOM, of those, 80% chose to send summary reports to their clinicians. Patients spent an average of 33 minutes (m) using WISDOM, with 11 m allocated to inputting data, 4 m interacting with the risk report, 3 m with the symptom/treatment chart, and 1.5 m with prevention options. Users were most interested initially in learning about diet and vitamins (43%), followed by lifestyle changes (33%), hormones (14%), and prescription meds (10%). Six (24%) clicked the chart an average of 12 times to read about a treatment-symptom interaction; 2 viewed more detailed information an average of 3 times. The 19 participants who viewed the prevention options page averaged 38% on breast cancer and hip fracture, and 24% on cardiovascular disease.

Conclusions: Participants interacted with a wide range of features and spent over 30 m on the site. Most, but not all, women sent summary reports to their clinicians, yet few were primarily interested in information about prescription meds. Entering the detailed personal information needed to provide tailored recommendations consumed a substantial portion of users’ time, but participants used the feedback it generated.

POSTER SESSION - PUBLIC HEALTH: METHODOLOGICAL ADVANCES
EVS1 FOR SURVIVAL TRIALS
Brennan A, Chilcott J, O’Hagan A and Kharrour S
University of Sheffield, SHEFFIELD, United Kingdom
Purpose: To develop methods to understand expected value of sample information (EVSI) calculations for 2 treatment survival trials using cost-effectiveness modelling.

Methods: Cox proportional hazard methods are common to evaluate the relative risk of survival, but the approach is non-parametric, and we sometimes need to extrapolate baseline survival for cost per QALY calculations. Within an illustrative (but representative) model, we use a Weibull survival curve for prior baseline survival (favorable lognormal to characterize uncertain Weibull parameters) and a lognormal distribution for prior relative risk. Monte Carlo simulation of notional patients = n. Data are censored for a pre-specified duration of follow-up d. For each simulated trial design survival trials. Such methods could be used alongside traditional sample-size calculations to compare results in practice. Further work is needed on the issue of possible correlation between baseline survival and relative risk.

POSTER SESSION - PUBLIC HEALTH: METHODOLOGICAL ADVANCES
DATA-DRIVEN BAYESIAN BELIEF NETWORK FOR FORECASTING EMPLOYMENT ONE YEAR AFTER TRAUMATIC BRAIN INJURY
Caster J; Doctor J, Dikmen S and Temkin N
University of Washington, Seattle, WA
Purpose: To compare the forecasting accuracy of Bayesian networks to other standard methods for determining the chance of employment one year after traumatic brain injury (TBI).

METHODS: This study compared four predictive models: (1) an independent Bayesian network that assumes independence of all predictor nodes from each other, (2) a complex Bayesian network where dependencies are allowed between predictor nodes, (3) a logistic regression model, and (4) a classification and regression tree (CART).

Data on demographic profile, pre-injury work information, injury severity, and neuropsychological test scores at one month post-injury collected from three longitudinal studies on TBI conducted at the University of Washington from 1980 to 1987 was analyzed. Data for 337 workers who were followed up for one year was incorporated. The two Bayesian networks were constructed using Netica. Structure was determined manually while the parameters were learned from data. Logistic regression was performed using Stata. CART analysis was performed using DIReCT.

Conclusions: Each model was trained and tested using a 10-fold cross-validation procedure. Probabilities of being employed at one year post-TBI were predicted for each subject in each model with probabilities of 0.5 or higher indicating employment.

One year after traumatic brain injury. Automated decision support systems should consider this approach when making forecasting judgments.
ABSTRACTS

POSTER SESSION - PUBLIC HEALTH; METHODOLOGICAL ADVANCES

APPLICATION OF A BAYESIAN MARKOV MODEL TO DECISION MAKING

Ding M, Xing Y, Cox D and Cormier J

Rice University, Houston, TX; M.D. Anderson Cancer Center, Houston, TX

Purpose: It has been suggested that Bayesian approaches offer superior analytical methods for decision-makers. We sought to demonstrate the advantages of Bayesian methodology to quantify the treatment strategies.

Methods: A continuous-time Markov model was designed with 5 health states (no evidence of disease, local-regional recurrence, distant recurrence, death of disease and death of other causes) to assess quality-adjusted life years (QALYs) for patients treated with IFN- compared to no adjuvant treatment following surgical resection. Patient data were retrospectively collected from a tertiary cancer center. A subgroup analysis was performed based on age cohorts (<40, 40-49, 50-59, and >59 years old). The survival time in each state was assumed to have an exponential distribution. The models were implemented in Splus.

Results: Using 5000 replications, four methods returned similar results for QALYs. With 5 million replications, results were near identical. However, the difference method repeatedly gave estimates substantially different to the other methods. The single MCS and UNLI methods were the least complex methods to use but are not restricted in their appropriateness. The two stage MCS and quadrature based methods are complex and time consuming.

Conclusions: The difference method is not rooted in the mathematical definition of EVPI and is clearly an inappropriate method for estimating EVPI. Thus, where appropriate EVPI should be estimated using either the single MCS or UNLI method. However, as often in cases where none of these methods is appropriate, the two stage MCS and quadrature methods should be used.

TABE 1

<table>
<thead>
<tr>
<th>Age</th>
<th>Control</th>
<th>IFN-</th>
<th>Incremental Effectiveness</th>
<th>95% Credible Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;40</td>
<td>7</td>
<td>10.40</td>
<td>1.93</td>
<td>-3.60, 8.52</td>
</tr>
<tr>
<td>40-49</td>
<td>5.42</td>
<td>11.21</td>
<td>5.79</td>
<td>-1.98, 2.22</td>
</tr>
<tr>
<td>50-59</td>
<td>6.54</td>
<td>5.25</td>
<td>-1.29</td>
<td>-4.01, 1.50</td>
</tr>
<tr>
<td>&gt;59</td>
<td>4.17</td>
<td>4.68</td>
<td>0.50</td>
<td>-1.86, 2.67</td>
</tr>
</tbody>
</table>

Conclusions: Bayesian approaches produce results that incorporate variability in patient histories and uncertainties in model parameters. In this case study, adjuvant IFN- was effective overall and in patients less than 50 years of age. This approach provides a solid theoretical framework for decision-making.

POSTER SESSION - PUBLIC HEALTH; METHODOLOGICAL ADVANCES

A NEW RESPONSE FORMAT FOR WTP: RESULTS FROM TWO PSYCHOMETRIC STUDIES

Hackenbruch J, Reis F, Urb P and Smith D

University of Michigan, Ann Arbor, MI; Princeton University, Princeton, NJ

Purpose: Using a traditional, open-ended willingness to pay (WTP) method can result in a highly skewed distribution with difficult to interpret outliers (e.g., $10,000,000,000,000), and can be confounded with respondents' financial resources. In addition, in health contexts the traditional WTP measure can be relatively insensitive to objective differences in health conditions. We conducted two studies testing a newly devised open-ended format where participants were asked to express WTP in terms of a percentage of their available income. Participants read descriptions of quality of life with a below-the-knee amputation and a spinal cord injury condition and were asked their WTP for a hypothetical treatment that could restore functioning for each condition. Order of presentation of the two conditions was counterbalanced.

Results: While results from study 1 directly compared this new measure to the traditional open-ended dollar format, study 2 added a manipulation to determine if the advantage of the percentage format would persist when changed to a per month format.

Methods: We distributed written, anonymous surveys to members of the general public in a hospital cafeteria (study 1 n = 230, study 2 n = 288). The participants were randomly assigned to either a traditional open-ended WTP question or to the new measure using percentage of available income. Participants read descriptions of quality of life with a below-the-knee amputation and a spinal cord injury and were asked their WTP for a hypothetical treatment that could restore functioning for each condition. Order of presentation of the two conditions was counterbalanced.

Conclusions: An alternative to the traditional open-ended WTP format in which respondents indicated their willingness to pay a percentage of their available income showed better distribution properties and was not correlated with income (p < .05), the traditional WTP was strongly correlated with income, r = .46, p < .001. This method proved to be more sensitive to the difference between the amputation condition and the more severe spinal cord injury condition than the traditional dollar amount method (p < .05). The difference in sensitivity persisted even after attempts to transform the traditional measures to address the distributional properties. Study 2 demonstrated that the advantage in sensitivity to severity persisted when approached as a per month format (p < .01).
APPLIED TO DETERMINE THE SENSITIVITY AND SPECIFICITY OF A NEW TEST WHEN NO GOLD STANDARD EXISTS.

LATENT CLASS ANALYSIS, A TECHNIQUE THAT USES MAXIMUM LIKELIHOOD ESTIMATION, CAN BE APPLIED TO DETERMINE THE SENSITIVITY AND SPECIFICITY OF A NEW TEST WHEN NO GOLD STANDARD EXISTS.

METHODOLOGY:

In this study, latent class analysis was used to determine the accuracy rates of three commonly used measures of carpal tunnel syndrome: the Time’s Test, the Phalen’s test, and the nerve-conduction velocity test. Results: It was found that the ‘Time’s Test and the Phalen’s test are both highly sensitive, 97 and 92 respectively and specific. 91 and 88 respectively.

CONCLUSION: The estimation of the sensitivity and specificity of these common tests for carpal tunnel syndrome support their widespread clinical use.

POSTER SESSION - PUBLIC HEALTH; METHODOLOGICAL ADVANCES

USING LATENT CLASS ANALYSIS TO ESTIMATE TEST ACCURACY FOR COMMON MEASURES OF CARPAL TUNNEL SYNDROME

LaJoie A1, McCabe S1 and Egedal S1
1School of Public Health and Information Sciences, University of Louisville, Louisville, KY; University of Louisville, Louisville, KY

Purpose: The accuracy of a diagnostic test used to classify a patient as being diseased or disease-free is a valuable piece of information to be used by the physician when making treatment decisions. If a gold standard reference test is available, determining the sensitivity and specificity of a new test is straightforward. However, if that reference test is incorrectly assumed to be perfectly sensitive and specific, the errors of the reference test can result in an under-estimation of the accuracy of the test being evaluated. Latent class analysis, a technique that uses maximum likelihood estimation, can be applied to determine the sensitivity and specificity of a new test when no gold standard exists.

METHODS: In this study, latent class analysis was used to determine the accuracy rates of three commonly used measures of carpal tunnel syndrome: the Time’s Test, the Phalen’s test, and the nerve-conduction velocity test. Results: It was found that the ‘Time’s Test and the Phalen’s test are both highly sensitive, 97 and 92 respectively and specific. 91 and 88 respectively.

CONCLUSIONS: The estimation of the sensitivity and specificity of these common tests for carpal tunnel syndrome support their widespread clinical use.

POSTER SESSION - PUBLIC HEALTH; METHODOLOGICAL ADVANCES

WEB-BASED BAYESIAN COMMUNICATION: THE BAYESIAN Z-TEST

Lehmann H1, Barshay A1, Grimm L1, Robinson K1 and Sheffield C2
1Johns Hopkins, Baltimore, MD

Purpose: Bayesian Communication (BC) provides an explicit and quantitative way to combine a reader’s preconceived notions with data from a study to help in making decisions, and thus implements the decision-analytic paradigm in the setting of interpreting and adapting research results. To date, BC has been available only for statistics on or end-users, interpreted through statisticians. The current research addresses whether BC can be provided over the Web to non-statistical clinicians using a knowledge-based paradigm.

METHODS: The domain was clinical trials whose outcomes in two groups were expressed as proportions (i.e., the Bayesian z-test for proportions). A User Panel of clinicians who read research reports (as well as generate them) was employed to elicit front-end specifications and to provide formative evaluation. Usability tests with other end-users were further employed during development.

RESULTS: A three-tier architecture was implemented; see http://www.hopkinsmedicine.org/bayes. The Web-based front-end guides users through the assessment process: First, the user minimally clinically important difference for the absolute difference in percentage is elicited. Then, the user chooses the appropriate statistical model, depending on which parameter they have the most prior knowledge: absolute difference alone; absolute difference plus baseline control rate; absolute difference plus experimental rate; relative difference plus control rate; and relative difference experimental rate. For each parameter, they are then asked to specify the mean and level of certainty. The application then displays the posterior distributions, after the system’s back-end utilizes the BUGS Bayesian-statistical updater to perform the calculations. Finally, the system further displays one-way sensitivity analyses for each prior mean and certainty, again, using BUGS, but using novel methods for this purpose.

CONCLUSIONS: This is the first attempt at delivering non-trivial BC to non-statisticians and provides a model for carrying the decision-analytic agenda forward into the domain of clinical trials.

POSTER SESSION - PUBLIC HEALTH; METHODOLOGICAL ADVANCES

PROBABILITY MODELING OF MEDICAL ERRORS IN RADIOTHERAPY

Lea B1, Ekaette E1, Kolly K2, Newcomb C3, Craighead P4 and Duncombe P4
1University of Calgary, Calgary, AB, Canada; 2Tom Baker Cancer Centre, Calgary, AB, Canada

Purpose: The process of treating cancer with ionizing radiation (radiotherapy) is complex and subject to medical errors; potentially resulting in morbidity and mortality to multiple patients, and litigation or criminal charges in some cases. Existing mechanisms to prevent such errors tend to be based on caregiver experience as opposed to using a systematic approach to identify and manage risks. Our objective is to develop and implement a quantitative risk and decision analysis model to elucidate risks and inform evidence based risk management and resource allocation decisions in radiotherapy.

METHODS: A team of oncologists, medical physicists, and risk and decision analysts first qualitatively mapped the radiotherapy system, and defined categories of events. Based on this, a probabilistic risk and decision analysis (PRADA) model was developed. This model employs linked influence diagram and Bayesian network calculations in a user friendly environment. The model allows estimation of risks and reduction of those risks using different quality assurance/quality control (QA/QC) patient safety interventions. Multiattribute utility functions are used as outcomes, and risk aversion is explicitly addressed. The value of different means of informing cancer staging and subsequent radiotherapy decisions is estimated using Bayesian methods. Model variables have been defined using a combination of literature values and expert judgement.

RESULTS: The qualitative mapping process identified four major categories of activities: Assessment, Preparation, Treatment, and Follow-up, and hierarchical levels of physician, physicist, dosimetrist, and technician activities. Influence diagrams have been defined for Assessment and Follow-up (as these activities involve decisions), and Bayesian networks for Preparation and Treatment (as these activities are simple sequences of events). Preliminary results indicate that the potential for catastrophic errors is high in the Assessment and Preparation stages, and relatively low in Treatment due to existing QA/QC procedures at that stage. Risk management options that increase resources to the early stages appear to have a potentially beneficial impact.

CONCLUSIONS: We provide a quantitative multiattribute model describing the potential sources of medical errors in radiotherapy and their consequences. The model suggests alternative strategies for risk management that may not be routinely implemented in cancer centres. This model represents a novel application of risk and decision analysis methods in the healthcare field. The methods should be generalizable to many forms of technologically intensive forms of healthcare such as surgery.

POSTER SESSION - PUBLIC HEALTH; METHODOLOGICAL ADVANCES

MEASURING ADAPTATION: THE TIME TRADE-OFF UTILITY MEASURE CAPTURES IMPROVEMENT IN QUALITY OF LIFE AFTER AMPUTATION

Sarr B1, Ubel P2 and Smith D3
1University of Michigan, Ann Arbor, MI

Purpose: An important area of quality of life research is the impact of illness and disability on patients’ quality of life. However, measurement of quality of life is complicated by many factors, including the fact that people tend to adapt physically and emotionally to their illness or disability. In general, it has been suggested that quality of life improves with time since the onset of disability, as physical functioning and symptoms improve, and as emotional adaptation occurs. We wanted to see whether a utility measure (the time trade-off—TTO) would be able to capture this improvement, compared to other self-report measures of well-being.

METHODS: We identified 273 individuals who had undergone an amputation at the University of Michigan in the last 5 years. Of these individuals, 172 amputees completed and returned the survey. Subjects were asked questions about their type and level of amputation, overall quality of life, happiness, physical functioning, health utility, and demographics. The TTO measure asked participants to indicate how many months of life they would trade in exchange for having their limb back, healthy and fully functional (the TTO score was calculated from answers to a series of 10 forced choices, beginning with a “10 year” of extreme choices to make the trade-off evident).

RESULTS: The TTO was correlated with measures of health; subjects who were willing to trade more time to have their limb restored reported more pain, and lower physical functioning (p's < 0.05). Pain declined with time since the amputation, while physical functioning increased (p's < 0.05). The time trade-off also captured this overall improvement: subjects were less willing to trade more time to have their limb restored reported more pain, and lower physical functioning (p's < 0.05). Various measures of subjective well-being, such as life satisfaction and mood, overall quality of life, and happiness were not correlated with time since amputation.

CONCLUSIONS: With amputees, we observed a decrease in pain and an increase in physical functioning with time since amputation. This improvement was apparently reflected in higher time trade-off utility values (indicating less willingness to trade life years for perfect health), but not with other subjective rating scales of well-being, including the 0-100 quality of life, the Diener Life Satisfaction scale, or measures of mood.

POSTER SESSION - PUBLIC HEALTH; METHODOLOGICAL ADVANCES

PDF COPY OF ABSTRACTS
ABSTRACTS

POSTER SESSION - PUBLIC HEALTH; METHODOLOGICAL ADVANCES
USING COMBINED PATIENT CHARACTERISTICS TO PREDICT DIRECT AND INDIRECT COSTS IN PARKINSON'S DISEASE

Sibbritt, D., Bomscehin, I., Sopottie, A., Berger, K., Oertel, W. and Dodel, K.
Massachusetts General Hospital, Harvard Medical School, Boston, MA; Ludwig Maximilians University, Munich, Germany; Munich, University of; Bonn, Germany; MRRG Medical Economics Research Group, Munich, Germany; University of Marburg, Marburg, Germany

Purpose: To identify independent predictors of disease-specific drug costs, direct non-drug costs, and indirect costs in Parkinson's disease.

Methods: Data from an ongoing prospective cost study of the German Competence Network for Parkinson Syndromes (n=152) were analyzed using multivariate regression. Potential predictors were sociodemographic factors, clinical variables from the Unified Parkinson's Disease Rating Scale [UPDRS] including disease stage [Hoehn & Yahr classification] and quality-of-life parameters [EuroQol, EQ-5D, Parkinson's Disease Questionnaire 39 [PDQ-39]]. Data for disease-specific drug costs and direct non-drug cost were log-transformed. Indirect costs were calculated using the human capital approach. Modeling of indirect costs proceeded without transformation in two steps: first, the probability of presence of indirect costs was predicted by logistic regression, second, estimation of costs was performed by linear regression in those with non-zero indirect costs.

Results: Predictors for disease-specific drug cost were age (p < 0.001), sex (p < 0.001), UPDRS (p = 0.0001), and quality of life [EQ-5D, p = 0.02]. The model for the prediction of other direct costs included disease stage [Hoehn-Yahr scale, p = 0.05 and p = 0.01] and PDQ-39 (p = 0.03). The probability for the presence of indirect costs was dependent on age (p < 0.001), UPDRS (p = 0.04), sex (p = 0.04), and depression [p = 0.02]. The magnitude of indirect costs was a function of disease stage (p = 0.003) and falls (p = 0.087). Variance explained by the models [adjusted R-squared] ranged from 24% to 28%.

Conclusions: We identified UPDRS and quality-of-life as the most important predictors of costs in Parkinson's disease. Drug costs also depended on age and sex. However, these factors explained only about a fourth of the total variance in costs.

POSTER SESSION - PUBLIC HEALTH; METHODOLOGICAL ADVANCES
COMBINING PROPENSITY SCORE WITH CLASSIFICATION AND REGRESSION TREE TO EVALUATE THE EFFECT OF INSULIN TREATMENT FOR TYPE II DIABETES

Theraul, E.*, Nicol, M. and Patel, B.*
University of Southern California, SAN DIEGO, CA; University of Southern California, Los Angeles, CA; MEDIMPACT, INC., SAN DIEGO, CA

Purpose: To estimate the effect of insulin treatment on the health care utilization of type II diabetes patients by developing a parsimonious propensity score-based model.

Methods: The data consists of the medical and pharmaceutical insurance claims for 370 type II diabetes patients. The estimation procedure uses a classification and regression tree (CART) analysis to determine the optimal set of explanatory variables for the propensity score weighted least square regression. The propensity score is calculated with logistic regression. This approach reduces the number of predictors to a small, manageable number without relying on any assumptions about the distributions of predictors or interactions between predictors. In addition, CART can determine a split value for continuous predictors, like age. Models relying on the logistic regression stepwise selection method are contrasted with models built with CART, and the results regarding the effect of insulin treatment are compared.

Results: Logistic regression models created with CART have a higher sensitivity and specificity than parametrically constructed models, for the same number of predictors. Both specification methods produce similar results, namely reductions in total cost and drug cost. The CART-based models, however, show a larger effect for insulin treatment on health care utilization, particularly patients over the age of 53. This may indicate that CART-based models can better correct for selection bias.

Conclusion: Combining the CART and propensity score methods produces more accurate and parsimonious models than parametric model selection alone.

POSTER SESSION - PUBLIC HEALTH; METHODOLOGICAL ADVANCES
IS THERE A SPECIAL AWARD FOR THE EXPERIMENTAL DRUG IN RANDOMIZED CONTROLLED TRAILS? THE CASE OF INTERFERON FOR CHRONIC HEPATITIS C

Thiebaud, P., Pagliaro, L. and Attanasio, M.
V. Cerrello Hospital, Palermo, Italy; University of Palermo, Palermo, Italy

Purpose: To identify independent predictors of disease-specific drug costs, direct non-drug costs, and indirect costs in Parkinson's disease.

Methods: ALT normalization sustained for at least 6 months was the abstracted measure of benefit, such effect was analysed in experimental vs control group for association with a set of variables of design and reporting. A fixed effects, arm level, logistic meta-regression accounted for the sources of heterogeneity reported in the individual studies [PROC LOGISTIC-SAS Institute, Cary, NC].

Results: Main characteristics of arms were mean age 46 years, male 65%, cirrhosis 21%, genotype type 1-63%, yearly schedule 28%, high quality score [nadas < 3] 28%, multicenter RCT 52%, unblinded RCT 72%, oriental country 19%, large trial (> 200 pts) 19%, funding for profit 51%, non-profit 19%, not declared 30%. Arms used as experimental included older patients, more cirrhotics, unblind RCT 72%, oriental country 17%, large trial (> 200 pts) 19%, funding for profit 51%, non-profit 19%, not declared 30%. Arms used as experimental included older patients, more cirrhotics, almost all originated from monocentric RCTs and were published mainly before 1995. Overall, 826 out of 4386 patients were sustained responders (benefit 17%). Only four variables maintained statistical significance after forward selection: mono vs multicentric trial [OR 1.9, 95% CI 1.1-3.0], experimental vs control arm [OR 1.5, 95% CI 1.1-1.9], yearly vs six-months schedule [OR 1.3, 95% CI 1.1-1.9] and trial first author from Asia [OR 4.4, 95% CI 1.6-11.9]. The Model Chi-square was 89.8, df 4, p < 0.0001.

Conclusion: arms wherein Interferon was used as experimental drug obtained higher rates of benefit. Methodological issues (local randomization procedure in the monocentric RCT or absence of double blind in all) or other factors might account for or contribute to the association between experimental role and likelihood of benefit. If present in other set of trials, this association might suggest a sort of special award for the experimental drug.
Poster Session - Public Health; Methodological Advances
Comparison of meta-analysis approaches: summary receiver operating characteristics (SRC0) curve versus Bayesian hierarchical models for assessment of diagnostic tests

Xing Y, Cong X, Foy M, Ding M, Cox D, Hunt K and Cormier J
M.D. Anderson Cancer center, Houston, TX; Rice University, Houston, TX

Purpose: To contrast various meta-analysis techniques for the assessment of the diagnostic test performance of sentinel lymph node biopsy (SNB) following neoadjuvant chemotherapy in patients with breast cancer.

Methods: A systematic review was conducted of studies which examined the results of SNB following neoadjuvant chemotherapy. Inclusion criteria required completion axillary lymph node dissection as a test "gold standard." Robust resistant regression method was used to construct ROC curves and compared to results using Bayesian hierarchical models. Two distinct Bayesian models were considered. A beta model using sensitivity derived from each study as random draws from a beta distribution was analyzed using an exponential prior for the beta parameters. Within a study, the observed number of SNB test positives out of the true positives was assumed to be a binomial random variable. The second model, logit model, assumed that the studies were derived from a population of studies in which the log odds of the sensitivity was a normal distribution. Both Bayesian models used prior parameters derived from published data. Sensitivity analyses were performed to examine the effects of prior selection on posterior estimates.

Results: Fourteen studies were identified. The range for reported sensitivity was 61-100%. The specificity was 100% in all studies. Pooling of data resulted in the sensitivity of SNB of 89% (95% confidence interval, 82-93). The logit model estimated the sensitivity of SNB at 90% (95% credible interval, 84-94). The logit model showed little sensitivity to prior parameters, while the beta model was more sensitive.

Conclusions: The estimation of sensitivity for SNB following neoadjuvant chemotherapy derived from meta-analysis of published studies varies from 83% to 90% depending on the analytic approach. Model assumptions are important in deriving summary estimates. Both Bayesian hierarchically generated a wider variation in the estimate because between-study variation was incorporated into these models. Bayesian approaches provide a flexible framework to incorporate trial heterogeneity, realistically assess uncertainty, and may result in better input for decision models.

Poster Session - Public Health; Methodological Advances
Calibration of a natural history model of cervical cancer using longitudinal primary data

Harvard School of Public Health, Boston, MA

Purpose: The multi-step model of cervical cancer pathogenesis involves, as the first step, infection with high-risk types of human papillomavirus (HPV). Many women with transient HPV will develop cervical abnormalities, although low-grade (LG) lesions have a high rate of spontaneous regression. Persistence of high-risk HPV types is a prerequisite for the development of high-grade lesions (HG) and cancer. Our objective was to use primary longitudinal data on HPV infection, and subsequent detection of LG and HG lesions to inform underlying transition probabilities in a natural history model of cervical cancer.

Methods: We developed a first-order Monte Carlo simulation model of underlying HPV and cervical disease to calibrate to outcomes in a study of 2,400 high-risk Brazilian women who received HPV testing and cytology screening at enrollment, 8 months, 12 months, and every year thereafter for 6 years. We simulated the underlying disease process for an initial cohort of 13-year-old girls, who were assigned a study entry age and screening schedule based on primary data from the study. We assumed that clearance of HPV depended on age and HPV type, and progression to LG and HG lesions depended on age, duration of HPV infection, and HPV type (high-risk or low-risk). Model outputs consisted of "results" from the screening tests, which depended on their sensitivities and specificities. Hazard ratios for detection of LG and HG lesions were compared with those from the analysis of the primary data.

Results: We identified a set of model parameters that describe the underlying transitions among HPV and cervical disease states that calibrate well to estimated hazard ratios for the association of low-risk and high-risk HPV status at enrollment with detection of LG and HG cervical lesions. Within 24 months, hazard ratios for association between high-risk HPV status and detection of LG lesions was 5.97 and of HG lesions was 5.88. Within 5.85 in the Brazil study, respectively). When allowing for time-varying HPV status, however, model derived hazard ratios were significantly lower than the primary data, suggesting a re-parameterization of the way that HPV is modeled over time.

Conclusions: Leveraging primary data from longitudinal studies together with model simulation methods provides unique opportunities for parameterizing the unobservable and transient nature of HPV infection and its role in the development of cervical cancer.

© 2005 Society for Medical Decision Making. All rights reserved. Not for commercial use or unauthorized distribution.
ABSTRACTS

POSTER SESSION - CEA: METHODS AND APPLICATIONS; HEALTH SERVICES RESEARCH

ESTIMATING TRANSITIONS BETWEEN SYMPTOM SEVERITY STATES OF SCHIZOPHRENIC PATIENTS OVER TIME: A BAYESIAN META-ANALYTIC APPROACH

Amin A.*, Meltzer H* and Dukic V.
University of Chicago, Chicago, IL; Vanderbilt University, Nashville, TN

Purpose: To develop a formulation of schizophrenia progression by modeling the 1-year time profile of severities of positive and negative symptoms associated with the disease under the influence of one of the three standard pharmacological treatments.

Methods: We develop a representative model for the natural course of schizophrenia in 18-65 year old patients. The model not only accounts for the positive and negative domains of the disease but also accounts for the severities within each domain. We employ Bayesian synthesis of published clinical trials and observational studies to estimate the transition probabilities between symptom severity states over time, while properly reflecting the overall uncertainty in the parameters that guide these transitions. Lower transition probabilities to moderate and severe symptom states indicates better control of symptoms.

Results: Based on the posterior mean estimates, we find several detailed transition patterns between severities of positive, negative and joint symptoms over time and by treatment, that were masked in published transition rates. Namely, the transition probabilities indicate that risperidone is best in controlling severe positive symptoms while olanzapine is the worst during the first quarter masked in published transition rates. Namely, the transition probabilities indicate that risperidone is best in controlling severe positive symptoms while olanzapine is the worst during the first quarter masked in published transition rates. Namely, the transition probabilities indicate that risperidone is best in controlling severe positive symptoms while olanzapine is the worst during the first quarter masked in published transition rates. Namely, the transition probabilities indicate that risperidone is best in controlling severe positive symptoms while olanzapine is the worst during the first quarter masked in published transition rates.

Conclusions: Various fine levels of detail on the transition probabilities may serve to estimate quality of life of schizophrenic patients and resource utilization in this field more accurately. Bayesian synthesis of published clinical trials and observational studies reveals considerable uncertainty in the time profiles and severity of schizophrenia symptoms over time and thereby the transition rates in the published findings. Estimates of parameter uncertainty by itself have important implications for the practice of cost-effectiveness analysis and future resource allocation policies in schizophrenia treatments.

POSTER SESSION - CEA: METHODS AND APPLICATIONS; HEALTH SERVICES RESEARCH

IMPACT OF SECONDARY CARDIOVASCULAR EVENTS ON HEALTH STATUS

Van Stel H.*, Busschbach J*, Hunink M* and Busemeyer B.
1UMC Utrecht, Utrecht, Netherlands; 2Erasmus Medical Center, Rotterdam, Netherlands; 3University Medical Center Utrecht, Utrecht, Netherlands

Purpose: Presently, reliable estimates on the impact of secondary cardiovascular events on health status in patients treated for cardiovascular disease are unavailable. However, precise estimates are obligatory to reduce uncertainty about the impact of secondary events, particularly for health economic modeling.

Methods: We gathered individual patient data on health status (EQ-5D, range 1 to 0.6) and secondary cardiovascular events (death, myocardial infarction (MI), cerebrovascular accident, endarterectomy, amputation, extracranial bleeding) during follow-up from several completed clinical trials comparing surgical interventions for patients with cardiovascular disease. Included were three trials on patients requiring coronary revascularisation, comparing bypass grafting, balloon angioplasty with stenting and minimally invasive bypass grafting (total n = 1463), and one trial on patients with intermittent claudication comparing angioplasty with or without stenting (n = 245). Re-interventions were not included or accounted for in this analysis. A random coefficients model on the utility scores, correcting for time and event, was fitted using G-Plus.

Results: This analysis included 1650 patients, with follow-up ranging from 12 to 36 months. Patients without secondary events improved after intervention (+0.05, p < 0.001) 285 patients died. Significant effects were found for myocardial infarction (n = 220, impact score -0.03, p = 0.005), cerebrovascular accident (n = 60, impact score -0.05, p = 0.002) and extracranial bleeding (n = 25, impact score -0.05, p = 0.01), but not for amputation (n = 15). Patients who got an event, except for MI, started substantially (but not significantly, p between 0.13 and 0.22) lower than patients without secondary events. Possible heterogeneity was ignored as the heterogeneity-tests F-max and IICE were negative.

Conclusions: Subsequent cardiovascular events have significant impact in terms of utility. Impact of secondary events is to some range as improvement due to surgery. Adding additional trial data and similar analyses performed on SF-36 domains will further increase the precision and validity of the results. These estimates on the impact of secondary cardiovascular events will be useful in reducing the uncertainty in long-term economic modelling. Supported by Netherlands Heart Foundation grant 2002B456.
ABSTRACTS

POSTER SESSION - CEA: METHODS AND APPLICATIONS; HEALTH SERVICES RESEARCH

INDIVIDUAL PATIENT DATA META-ANALYSIS OF COST-EFFECTIVENESS RANDOMIZED CONTROLLED TRIALS

Delaney B, Ford A, Qume M and Moayyedi P

The University of Birmingham, Birmingham, United Kingdom; Leeds Royal Infirmary, Leeds, United Kingdom; McMaster University, Hamilton, ON, Canada

PURPOSE: Cost-effectiveness RCTs, where the principal aim is to determine both the difference in effect and the resources costs of alternative treatment strategies, are increasing common. Whilst accepted methods exist for meta-analysis of measures of effectiveness, combining cost-effectiveness results is more complex, and cannot be performed without access to the individual trial datasets.

RESULTS: Five trials that compared H. pylori 'test and treat' with endoscopy were pooled. No significant difference was found, relative Risk of dyspepsia recurrence = 1.04, 95% confidence interval 0.97 to 1.11), Endoscopy was not evaluated at patient level using INB = $1570 per patient 'treated'. INB remained negative at -$282/-3240/-4590. Five trials that compared empirical acid suppression with endoscopy were pooled. No significant difference was found, Relative Risk of dyspepsia recurrence = 1.04, 95% confidence interval 0.97 to 1.11). Endoscopy was not cost-effective. INB remained negative at -$1560/-2490/-160 at 6 and was insensitive to increasing A.

CONCLUSIONS: Empirical acid suppression with endoscopy is more complex, and cannot be performed without access to the individual trial datasets.

POSTER SESSION - CEA: METHODS AND APPLICATIONS; HEALTH SERVICES RESEARCH

BAYESIAN POSTERIOR DISTRIBUTIONS FOR PROBABILISTIC SENSITIVITY ANALYSIS

Hanen G and Huang M

Northwestern University, Evanston, IL.

Purpose: In probabilistic sensitivity analyses (PSA), analysts assign probability distributions to uncertain model parameters, and use Monte Carlo simulation to estimate the sensitivity of model results to parameter uncertainty. Bayesian methods provide convenient means to obtain probability distributions on parameters given data. We present large-sample approximate Bayesian posterior distributions for probabilities, rates and relative effect parameters, and discuss how to use these in PSA.

METHODS: We use Bayesian random effects meta-analysis, extending procedures summarized by Adeo, Lu and Glaston (2004). We outline procedures for using the resulting posterior distributions in Monte Carlo simulation.

RESULTS: We apply these methods to conduct a PSA for a recently published analysis ofHeight v. non in the early postnatal period to estimate HIV transmission, as well as data for 5 other probability parameters. Given this data, the two parameters of log Risk population mean and log Risk Ratio population mean for HIV transmission have approximate priors flat normal posterior with mean equal to -3.308 12 and -0.020 23, and correlation -0.108. Using these and other posterior distributions for all 5 remaining probabilities in a PSA yields HIV transmission probabilities for 86.5% (0.13%) of the time, and the expected value of perfect information for each parameter and probabilities equal to $15.13 (1.10) per pregnancy. These results concur with the gold standard for HIV transmission followed by zidovudine prophylaxis is not a close call.

EMPIRICAL TESTING AND VALIDATION OF THE WAITING-TRADEOFF (WTO) METHOD

Kongskorn T, Swan J and Sainfort F

Georgia Institute of Technology, Atlanta, GA; Harvard Medical School, Boston, MA

Purpose: The Waiting-Tradeoff (WTO) is an alternative to the chained temporary Time-Tradeoff (TTO) method for short-term health states such as medical testing. It offers a choice between having a potentially noisy test followed by immediate treatment as opposed to waiting for the results of an "ideal" test before having treatment. The result is scaled to a QALY measure using baseline and disease states for a relevant condition. The WTO was motivated by literature showing the importance of short-term preferences for testing in cost-effectiveness analysis. This study involved the testing of the WTO method by exploring key variables with a sample of healthy subjects who were given scenarios involving testing for atherosclerosis carotid vascular disease with t-way angiography (TWA) or magnetic resonance angiography (MRA).

Methods: A series of medical scenarios in the context of suspected carotid atherosclerotic vascular disease was constructed by varying three factors of interest: (1) radiologic technology (two levels of invasiveness - MRA and XRA), (2) 3 levels of current health state - severe stroke, transient ischemic attack (TIA), and asymptomatic with abnormal finding, and (3) probability of improvement to asymptomatic (three levels—small, good, and excellent chance). These three different factors lead to 18 possible scenarios that were presented to each of 100 healthy subjects. To ascertain test-retest reliability, each subject repeated the assessment for one randomly selected scenario for each technology.

RESULTS: Means for WTO across subjects ranged from 1.10 to 9.74 days for MRA, and from 0.75 to 30.51 for XRA. Mann-Whitney tests for paired comparisons revealed that WTOs for XRA were significantly larger than WTOs for MRA in all comparisons (p < 0.001), indicating subjects preferred the less invasive technology. MRA and XRA waiting times significantly shortened as the current health state increased in severity. Thus, assessments for specific future health states strongly depend on current health states. Regarding test-retest reliability, Spearman's rho correlation coefficients and intraclass correlation coefficients were 0.634 and 0.603 for MRA, 0.773 and 0.782 for XRA, suggesting adequate reliability.

Conclusions: The results suggest the WTO method is a reliable, valid and useful method for measuring preferences for short-term health states. Additionally, these data suggest dependence of future health state assessment on the current health state. This has important implications for preference measurement over multi-state profiles.


POSTER SESSION - CEA: METHODS AND APPLICATIONS; HEALTH SERVICES RESEARCH

MEDICAL DECISION MAKING/JAN–FEB 2005

E34
ABSTRACTS

POSTER SESSION - CEA: METHODS AND APPLICATIONS; HEALTH SERVICES RESEARCH

MORBIDITY COST IN COST-EFFECTIVENESS ANALYSIS

Myers J, McCall S and Gohmann S

Yale University, New Haven, CT; Christine M. Kleinert Institute for Hand and Micro-surgery, Louisville, KY; University of Louisville, Louisville, KY

Purpose: The handling of morbidity cost in cost-effectiveness analysis presents a challenge to the current recommendations. The current recommendations suggest that morbidity costs are irrelevant when assessing the quality of life for a health state. While acceptance of the recommendation is not universal, there has been no empirical evidence demonstrating if morbidity cost should be included in the numerator or denominator of the cost-effectiveness ratio. The current study aimed to provide the first empirical evidence demonstrating whether or not people automatically consider morbidity costs when assessing the quality of life for a health state.

Methods: 181 undergraduate students were randomly assigned to one of two groups: (1) those participants who were not informed of morbidity costs and (2) those participants who were informed of morbidity costs. The participants were additionally randomly assigned a level of illness severity (mild, moderate, or severe). Students were asked to read a description of a health state and to assign an assessment of quality of life for the health state described by the use of the paper standard gamble. Two-factor analysis of variance was performed on the instruments completed.

Results: The overall mean Quality of Life (QOL) for the informed group was significantly lower than that of the uninformed group (P = 0.0001, F = 24.2, df = 1, 179). Similarly, there is a significant gamble. Two-factor analysis of variance was performed on the instruments completed.

Conclusion: The current study demonstrated that those subjects informed of morbidity costs scored quality of life lower than subjects unformed of morbidity costs. Morbidity costs are part of the cascade of events that result from an intervention. Therefore, morbidity costs are part of the effectiveness of an intervention and should be included in the denominator of the cost-effectiveness ratio.

To accurately represent the effectiveness of an intervention, we argue morbidity costs should be included in the description of health states. We recommend that descriptions of health states include morbidity costs when conducting a cost-effectiveness analysis.

POSTER SESSION - CEA: METHODS AND APPLICATIONS; HEALTH SERVICES RESEARCH

POTENTIAL COST EFFECTIVENESS OF AGGRESSIVELY TREATING MULTI-DRUG-EXPERIENCED HIV-POSITIVE PATIENTS

Bayoumi A, Sanders G, Munakata J and Neumann P

Harvard School of Public Health, Boston, MA; Columbia University, New York, NY

PURPOSE: Adherence is a major determinant of the therapeutic effectiveness of medications. Despite evidence that adherence can influence the results and conclusions of cost-utility analyses (CUAs), little is known about whether published studies consider it. Our objective was to examine the inclusion of patient adherence in CUAs of medications. Because the exclusion of adherence might oversrate the cost-effectiveness of interventions, we also examined the relationship between pharmaceutical company sponsorship and adherence inclusion.

METHODS: A systematic review of the English-language literature published between 1998 and 2001 identified 294 original CUAs, including 98 pertaining to self-administered medications. Two trained readers independently abstracted detailed data on study methods and results, and completed a consensus form for each item, including an item on whether compliance or adherence to intervention was considered. We estimated rates of adherence inclusion overall and by study characteristics. Association between adherence inclusion and study sponsorship was assessed with a chi-square test.

RESULTS: Among the 98 CUAs of self-administered medications, 40% (n = 39) considered patient adherence to therapy. Adherence was equally likely to be considered in CUAs of long-term versus short-term (<1 month) drug therapy (39% vs. 44%, p = 0.45). Inclusion of adherence varied across the most commonly studied clinical areas: 25% of chronic anticoagulation studies, 57% of cardiovascular risk reduction studies, 50% of neuropsychiatric studies, and 46% of HIV antiretroviral studies. Among the 70 CUAs in which study sponsorship was disclosed, 46% of pharmaceutical sponsored studies (n = 40) and 50% of non-pharmaceutical sponsored studies (n = 30) included adherence (deviation non-significant).

CONCLUSIONS: Despite its potential importance, few CUAs incorporate medication adherence. As decision and cost-effectiveness analyses are meant to explicitly model real-world costs and effects of interventions, investigators would do well to explicitly consider medication adherence in future analyses.
ABSTRACTS

POSTER SESSION - CEA: METHODS AND APPLICATIONS; HEALTH SERVICES RESEARCH

IS MOHS SURGERY A COST-EFFECTIVE TREATMENT FOR FACIAL NONMELANOMA SKIN CANCERS? A DECISION ANALYSIS


Emory University School of Medicine, Atlanta, GA; Kaiser Permanente, Redwood City, CA; University of Connecticut, Farmington, CT; Emory University, Atlanta, GA

Purpose: Nonmelanoma skin cancers (NMSC), including basal cell carcinoma (BCC) and squamous cell carcinoma (SCC), are the most common cancers in the United States, accounting for more than 1,200,000 new cases per year. More than 65% of all BCCs and SCCs affect facial sites with both cosmetic and functional sequelae. Two common treatment modalities for NMSC are Mohs Surgery (Mohs) and traditional surgical excision (TSE), each with a different efficacy and cost. The purpose of this study was to determine whether Mohs is a more cost-effective method of treatment for facial NMSC than TSE.

Methods: The data from our prospective trial of 96 consecutive patients with primary facial NMSC was used to obtain baseline cost (Connecticut Medicare 2002 reimbursements) and efficacy (marginal analysis) information for our CEA. We approached the CEA using a decision analysis model via TreeAge Data 4.0 software. Our model also incorporated efficacy by 3-year recurrence rates from the literature, and outcomes (quality-adjusted life-years (QALY)) using data from a focus group of patients. We performed a sensitivity analysis to determine the influence of key estimates in the model.

Results: Our baseline CEA demonstrated Mohs to be less costly and more effective than TSE ($960.60 vs. $1248.95, and 0.0 QALY gain). The sensitivity analysis showed that varying values for QALYs, recurrence rates, and percentage of frozen and permanent section margin analysis did not change the results of our CEA. However, our results were sensitive to varying the proportion/cost of defect repairs (gumplation, primary closure, flaps, grafts) following the two procedure strategies.

Conclusions: Therefore, before the most cost-effective treatment for facial NMSC can be definitively established, further research into actual practice patterns of defect repair selection for both procedures must be examined.

POSTER SESSION - CEA: METHODS AND APPLICATIONS; HEALTH SERVICES RESEARCH

BIPOLAR OR RESECTION FOR SINGLE SMALL LIVER NODULES IN PATIENTS WITH COMPENSATED CIRRHOSIS: A DECISION ANALYSIS

Brennan K, Bayoumi A, Sherman M and Krah M

University Health Network, Toronto, ON, Canada; St. Michael’s Hospital, University of Toronto, Toronto, ON, Canada

Purpose: To determine the optimal management of small (1-2 cm) liver nodules detected during routine screening for hepatocellular carcinoma (HCC) in cirrhotic patients. These nodules may be, but are not invariably, malignant. Biopsy is often recommended, but its limited sensitivity results in false negative findings. Hepatic resection offers potential cure, but a risk of liver failure and unnecessary surgery.

Methods: We compared two strategies in a Markov decision model: immediate resection, or liver biopsy and resection if positive. The patient cohort, men and women, aged 55-70 years, had compensated cirrhosis, no comorbidities, and a single 1-2 cm liver nodule identified as probable HCC with routine ultrasound screening and CT confirmation. Patients who had resection (immediately or after biopsy) faced the risks of surgery (unnecessary if not HCC), recurrent HCC, and liver decompensation, but had the advantage of early treatment (if HCC-positive). Biopsied patients had the potential benefit of confirmatory diagnosis, but the risks of biopsy, and the chance of missed HCC, and potential progression to an unresectable state before re-screening. Probabilities and utilities were obtained from a comprehensive literature review and local data.

Results: The baseline analysis favored initial hepatic resection, but the gain was small ($4.72 quality-adjusted life months, or 4.49 life month). The model was robust to most assumptions; the decision was not sensitive to probabilities for transitions to, or death from, biopsy, resection, liver decompensation, or HCC recurrence, or the utility values. Only the sensitivity of biopsy changed the decision; if 95% or higher, biopsy was preferred. Patients with false negative biopsies had the poorest prognosis, as they had the risks of biopsy (needle-track seeding, death), and the chance of decompensation or developing other complications to resection prior to accurate HCC diagnosis by imaging in follow-up screening.

Conclusions: Our results suggest that resection of all suspicious, 1-2 cm liver nodules offers larger survival and better quality of life than selection of patients for resection based on liver biopsy. Diagnostic test modalities for HCC must attain high sensitivity (95%), thus assuring fewer missed HCC, to offset the benefits of early treatment.

POSTER SESSION - CEA: METHODS AND APPLICATIONS; HEALTH SERVICES RESEARCH

THE COST-EFFECTIVENESS OF TARGETED THERAPY FOR ADVANCED COLORECTAL CANCER

Elkin E and Schrag D

Memorial Sloan-Kettering Cancer Center, New York, NY

Purpose: Targeted cancer therapies may improve clinical outcomes for the patients most likely to benefit, while sparing those unlikely to benefit the costs and possible harms of ineffective treatment. Among advanced colorectal cancer patients whose tumors express the epidermal growth factor receptor (EGFR), cetuximab in combination with chemotherapy has been shown to extend time-to-progression and survival. Despite the availability of an immunohistochemical assay (IHC) to identify the approximately 75% of patients whose tumors express EGFR, reports suggest that the test is not widely used. The purpose of this analysis was to assess the impact of IHC testing on the cost-effectiveness of cetuximab therapy.

Methods: We constructed a Markov state-transition model to simulate clinical outcomes in a cohort of 65-year-old patients with metastatic colorectal cancer. We estimated the average costs and time-to-progression and survival. The model included the following states: baseline (with and without IHC), cetuximab plus chemotherapy for EGFR-expressers; no test, cetuximab plus chemotherapy for all patients. Transition probabilities related to cancer progression and treatment efficacy were obtained from clinical trials. We assumed that cetuximab had no benefit in patients whose tumors did not express EGFR. Since the test characteristics of IHC for EGFR have not been well-described in colorectal cancer, we used two-way sensitivity analysis to examine the role of test sensitivity and specificity on the cost-effectiveness of alternative clinical strategies.

Results: In the absence of IHC testing, the addition of cetuximab to standard chemotherapy for all patients increased treatment costs by approximately $86,000, and extended average survival by about 1.2 months, yielding an incremental cost-effectiveness ratio (ICER) of more than $860,000 per QALY. Use of a perfect test achieved the same survival benefit, but at a lower cost, with an ICER of $575,000. Assuming the test had 90% sensitivity and specificity, the ICER of targeted cetuximab therapy was less than $560,000 per QALY, and the ICER of universal cetuximab therapy was more than $1.5 million per QALY.

Conclusions: Using IHC to identify cetuximab candidates can substantially improve the cost-effectiveness of this therapy for patients with advanced colorectal cancer. Even though the prevalence of tumor EGFR expression is relatively high in these patients, it is more cost-effective to use an imperfect test for targeting treatment, rather than prescribe cetuximab universally.
POSTER SESSION - CEA: METHODS AND APPLICATIONS; HEALTH SERVICES RESEARCH
VALUE OF INFORMATION IN CAROTID REVASCULARISATION

Janmar M and Berkova E
University Medical Center Utrecht, Utrecht, Netherlands

Purpose: Stenting of the carotid artery (CAS) for treatment of carotid stenosis is increasingly applied. Costs of stents are high and conclusive trials to determine the (cost-)effectiveness of CAS as compared to endarterectomy (CEA) have to be awaited. The latter was already proven effective in symptomatic patients with severe stenosis. In absence of sound evidence and a yet increasing tendency to perform CAS we set out to assess uncertainty regarding costs and clinical outcomes of CAS, determine conditions required for CAS to become cost-effective in the Netherlands and find main targets for further research given the available evidence.

Methods: Cost and effect estimates from various sources were combined in a Markov model. The European Carotid Surgery Trial (ECST) data form the basis for this model. Procedural costs were collected in-house, and late event costs and quality of life estimates were obtained from literature. Estimates for the clinical outcome after carotid stenting were obtained by expert opinion elicitation. The EVPI of several combinations of model parameters were calculated to assess their relevance. 

Results: The costs of CAS were Euro 2071.100 higher than those of CEA. Net Health Benefits (at Euro 2071.100 per QALY) per percent decrease in complications for CAS were estimated at Euro 2071.100 and Euro 2071.725 at 20% and 25% post-operative complications “major stroke,” “minor stroke” and “death” respectively. For long term postoperative major and minor stroke rates these figures were Euro 2071.380 and Euro 2071.160. The global EVPI was estimated at Euro 2071.100 and the EVPI for post-operative and long-term complication rates were Euro 2071.160 and Euro 2071.170 respectively. CONCLUSIONS: At a global EVPI of Euro 2071.100 an annual intervention rate of approximately 1200 interventions implies a total value of Euro 2071.1.000.000 per year for the Netherlands. Looking at the EVPI of peri-operative and post-operative complications it is clear that even if perfect information on peri-operative complications is obtained, there is still Euro 2071.1.000.000 per year left for research on postoperative complications. A value of sample information analysis may provide a definite answer as to the cost-effectiveness of a randomized trial with long term follow-up.

POSTER SESSION - CEA: METHODS AND APPLICATIONS; HEALTH SERVICES RESEARCH
A PROBABILISTIC COST-EFFECTIVENESS ANALYSIS OF ENOXAPARIN VS. UNFRACHTIONATED HEPARIN FOR DVT PROPHYLAXIS FOLLOWING MAJOR TRAUMA

Lynd L, Cooree R and O'Brien B
McMaster University, Hamilton, ON, Canada

Objectives: To use a Bayesian approach to evaluate the cost-effectiveness of enoxaparin (ENOX) versus unfractionated heparin (UH) for the prophylaxis of deep-vein thrombosis (DVT) following major trauma.

Methods: A decision analytic model was used to measure the incremental cost and incremental effectiveness of ENOX vs. UH as the comparator for DVT prophylaxis from the hospital perspective. Outcome data were extracted from the only published clinical trial (MMJ 1996, 315 (10):708-7). The probability of death from a major bleed or pulmonary embolism (PE), and additional model parameters were derived from published data and the Ontario Trauma registry, with hospital costs derived from the Ontario cost-coing project. The rates of venographically detected DVT were adjusted for the sensitivity and specificity of clinically and ultrasonographically diagnosed DVT. The primary outcome measures were the incremental cost per DVT averted and cost per life-year gained. Probabilistic sensitivity analysis was performed using 2nd order Monte Carlo simulation. All costs are in 2003 CDN.

Results: The total cost of treatment with enoxaparin was $13,395 versus $13,229 with UH, resulting in an incremental cost of $97. Enoxaparin resulted in an incremental effect of 0.085 DTHs averted and 0.29 life years gained. These results yielded an incremental cost-effectiveness ratio of $1,139 per DVT averted however, when life years gained was used as the metric of effectiveness, enoxaparin was dominated by UH. Monte Carlo simulation revealed that in the DVT averted model, 98% of the model iterations fell in the NE and SF quadrants, favoring ENOX. Conversely, in the life years gained model, 96% of the model iterations fell in the NW and SW quadrants, favoring UH. At $70,750 per life-year, there was only a 5% probability that enoxaparin was cost-effective.

Conclusions: Only one previous study has evaluated the cost-effectiveness of DVT prophylaxis in this clinical scenario which showed that ENOX was the dominant strategy. However, neither the costs or outcomes related to PE, or mortality related to PE or major bleed were incorporated into the model. This study demonstrates the importance of considering mortality when modeling DVT prophylaxis-related outcomes in the trauma population, and the benefit of a Bayesian modeling approach to the analysis.

POSTER SESSION - CEA: METHODS AND APPLICATIONS; HEALTH SERVICES RESEARCH
DECISION ANALYSIS FOR TUBERCULOSIS CONTROL USING A DYNAMIC EPIDEMIC MODEL

Buech S, Khilari S, Murray M and Weinstein M
1Harvard University, Boston, MA 2Harvard School of Public Health, Cambridge, MA 3Harvard School of Public Health, Boston, MA

Purpose: The optimal control strategy for tuberculosis (TB) in countries with prevalent or increasing multidrug-resistant disease (MDR-TB) is controversial when resources are substantially constrained. Directly-observed treatment (short-course) with first-line drugs (DOTS) has been shown to be cost-effective but not adequate for treating MDR-TB. Second-line drugs can cure MDR-TB, but also prevent its transmission to others, but they are more expensive and require longer treatment duration. We developed a decision model to inform policymakers in deciding whether to allocate limited TB control resources to treating MDR-TB.

Methods: A deterministic dynamic compartmental model of a population of 100,000 was constructed to describe the transmission dynamics of DOTS-susceptible- and MDR-TB, capturing both the acquisition of drug resistance by ineffective-treated cases and the impact of effective TB treatment on reducing transmission. Four treatment strategies that differed by drug regimen type and case identification method were evaluated: (1) DOTS only, (2) DOTS for new cases and standard-second-line-regimen for first-line treatment failures, (3) DOTS for new cases, drug-susceptibility testing (DST) for first-line treatment failures and individualized-second-line-regimen for identified MDR-TB cases, (4) DST for all new patients with DOTS for susceptible cases, and individualized regimens for MDR-TB cases. Model parameters were dynamically calibrated to reflect epidemiological indicators for Peru where annual TB incidence is 120 per 100,000 with 3% MDR-TB.

Results: Over the 40 year time horizon (discount rate 3%), the baseline DOTS strategy resulted in 524 TB deaths and $238,000 of TB-related costs. Performing DST on first-line failures and treating MDR-TB with an individualized-second-line-regimen had an incremental cost per death averted (ICDA) of $900. Use of a standardized-second-line-regimen was dominated. The strategy where DST is performed on all new TB cases had an ICDA of $4400. When treatment's transmission externality is ignored, the ICDA for strategies (3) and (4) increase from $900 to $12,800, and $4400 to $25,600 per death averted, respectively.

Conclusions: Treating MDR-TB after first-line failure improves health outcomes at an affordable cost ($900 per QALY gained). However, DST on all new TB cases on reducing transmission overestimates the total burden of disease and cost, and also makes the cost-effectiveness ratio of treating MDR-TB with second-line drugs appear several times less favorable.
ABSTRACTS

POSTER SESSION - CEA: METHODS AND APPLICATIONS; HEALTH SERVICES RESEARCH

COST EFFECTIVENESS OF THE IMPLANTABLE CARDIOVERTER DEFIBRILLATOR IN THE MADIT-II POPULATION

Sanders C1, Halkty M1 and Owens D1
1Duke, Durham, NC; 2Stanford University, Stanford, CA

The Multicenter Automatic Defibrillator Implantation Trial (MADIT-II) randomized clinical trial demonstrated implantable cardioverter defibrillators (ICDs) improve survival in post myocardial infarction patients with low ejection fraction. We evaluated the cost effectiveness of ICD implantation compared with conventional therapy in such a population. We used a Markov-based cost-effectiveness analysis to estimate the lifetime costs, life expectancy, quality-adjusted life expectancy, and incremental cost effectiveness. We based survival, sudden cardiac death, and the effectiveness of the ICD on the MADIT-II trial population, and used costs and utilities from the published literature. Total mortality was assumed to be composed of three types: sudden cardiac, non-sudden cardiac, and non-cardiac mortality. Assuming an exponential declining life expectancy, we modeled a constant probability of sudden and non-sudden cardiac death to match the trial sudden and total mortalities for conventional treatment over the average trial follow up (20 months). Based on the trial data we assumed a 67% relative risk reduction in sudden cardiac death in the ICD arm. Compared with conventional therapy, over a patient's lifetime, ICD use led to a greater quality-adjusted life expectancy (increase of 1.33 QALY's) but higher costs (increase of $67,900) — resulting in an incremental cost-effectiveness of $50,900/QALY gained compared with conventional therapy. To obtain a cost-effectiveness of less than $100,000/QALY, ICDs must reduce arrhythmic mortality by 31%. If the cost of the ICD device was reduced from $23,000 to $10,000, the cost-effectiveness of the ICD relative to conventional therapy would improve from $56,000 to $33,500/QALY gained. If the ICD improved quality of life, the cost-effectiveness would improve by $11,000 compared to our base-case estimate; however if quality of life is substantially diminished, use of an ICD becomes expensive. There is little evidence to suggest such an effect on quality of life, however. Our analysis indicates that use of an ICD who meet the criteria for the MADIT-II trial may be economically favorable when compared with conventional therapy. The size of population potentially eligible for prophylactic ICD implantation however suggests that future studies may identify subgroups in whom the cost-effectiveness of prophylactic ICD implantation is higher or lower than the average for this clinical population.

POSTER SESSION - CEA: METHODS AND APPLICATIONS; HEALTH SERVICES RESEARCH

EVALUATION OF LONG-TERM CLINICAL EFFECTIVENESS AND COST-EFFECTIVENESS OF THE NEW GENOTYPE-SPECIFIC GUIDELINES FOR CHRONIC HEPATITIS C TREATMENT IN GERMANY

Siebert U1, Stroczynski G2, Aidelburger P2, Rossol S3, Wasem J4, Manns M5, McKhann J2 and Wong J2
1Massachusetts General Hospital, Harvard Medical School, Boston, MA; 2University of Duisburg-Essen, Essen, Germany; 3University of Heidelberg, Rüsselsheim, Germany; 4Medical School of Hanover, Hannover, Germany; 5Duke University Medical Center, Durham, NC; 6Tufts-New England Medical Center, BOSTON, MA

Purpose: The recently developed German guidelines for antiviral treatment (AVT) in patients with chronic hepatitis C (HCV) recommend busing drug dosage, intended treatment duration, and early stopping rules on the genotypes of the hepatitis C virus (HCV). Therefore, we sought to evaluate the lifetime clinical effectiveness and cost-effectiveness of different AVT strategies including the new German guidelines on genotype-specific treatment. Methods: The German Hepatitis C Model (GEHM), a validated and published Markov model reflecting the German health care system and the practice patterns of German physicians, was used to project clinical events, life expectancy, quality-adjusted life years (QALY), and lifetime costs for the following AVT strategies: (1) no AVT (NoAVT), (2) Interferon alfa-2b plus ribavirin for 48 weeks (PEG), (3) Peginterferon alfa-2b plus ribavirin for 48 weeks (PEG), (4) Peginterferon alfa-2b plus ribavirin according to the German guidelines with genotype-dependent AVT duration, dosing and early stoppage in HCV-positive patients after 12 weeks (GUD/E). Incremental discounted cost-effectiveness ratios (ICER) were calculated from a societal perspective. Clinical data and actual drug utilization data were derived from a large multi-centre randomized clinical trial. Detailed data on long-term costs were based on German actual variable costs, reimbursement data, and health resource utilization data from the German Hepatitis C Patient Survey (n = 196).

Results: Compared to NoAVT, combination therapy with peginterferon alfa-2b and ribavirin (PEG or GUD/E) reduced the 20-year risk for decompartmental cirrhosis, hepatocellular carcinoma, liver transplantation, and liver-related death by more than 50%. Compared to NoAVT, PEG increased life expectancy by 0.06 years and QALY gained by 0.003 years. The incremental cost of GUD/E dominated HFN by strong dominance. Compared to NoAVT, discounted ICERs were 1500 EUR/QALY for NoAVT, 1300 EUR/QALY for PEG, 4700 EUR/QALY for GUD/E. Moving from NoAVT to PEG was associated with an ICER >100,000 EUR/QALY.

Conclusion: Administering combination therapy with peginterferon and ribavirin in accordance with the new German guidelines allows tailoring treatment efficiency to HCV genotype, body weight, and early viral response in patients with minimal loss of effectiveness. Antiviral treatment according to the new German guidelines should be cost-effective compared to other well-accepted medical interventions.
POSTER SESSION - CEA: METHODS AND APPLICATIONS; HEALTH SERVICES RESEARCH

PRACTICAL USE OF COST-EFFECTIVENESS FOR COMMUNITY PLANNING OF HIV PREVENTION

Wu S.*, Cohen D* and Farley T†
RAMI, Santa Monica, CA, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA

Purpose: To develop a practical tool to guide local communities to correctly use cost-effectiveness (CE) information in the literature to set priorities and allocate resources to prevent as many HIV infections as possible.

Methods: We developed “Maximizing the Benefit,” a free tool containing a CE calculator, a multi-attribute utility model with optimization capability, and a user manual. The CE calculator is a pre-programmed Excel spreadsheet tool that compares the cost-effectiveness of 25 HIV prevention strategies in the literature, including individual behavior-change, biomedical, and structural interventions. It also allows users to specify their own interventions. Users can input local data on the size and HIV prevalence of the target group, intervention effectiveness, and local costs. The tool then uses Bernoulli models and proportionate-change models to estimate the number of HIV infections prevented and the cost per infection prevented. The multi-attribute utility model then takes the CE information along with other contextual factors that influence the decision of prevention planning in its priority-setting analysis. Users will specify the contextual factors, such as the local acceptability of an intervention, and determine their relative weights using fuzzy logic. An optimization algorithm can be used to provide the tool for decision support of resource allocation. The user manual includes detailed guidance about how to use the tool, the methods for cost-effectiveness estimation, and how to accurately interpret the CE estimate and include it in HIV prevention planning.

Results: The tool shows that the most important factors in determining the cost-effectiveness of the interventions are the local HIV prevalence and the cost per person reached. For low-prevalence populations (e.g., heterosexuals) the only cost-effective interventions were structural interventions (e.g., mass media, condom distribution), whereas for high-prevalence populations such as men who have sex with men (MSM) and injecting drug users (IDUs), individual and small-group interventions were still relatively cost-effective. Among the most cost-effective strategies overall were showing videos to clinics and warning alcohol taxes. When other contextual factors are considered, local communities may generate different portfolios for their HIV prevention programs.

Conclusions: Comparing the CE of HIV prevention interventions provides insight that can help local communities optimize their HIV prevention strategies. Our tool can facilitate the practical use of CE in community planning to maximize the number of infections prevented.

POSTER SESSION - CEA: METHODS AND APPLICATIONS; HEALTH SERVICES RESEARCH

PROPENSITY SCORING: A COMPARISON OF GREEDY VS. OPTIMAL MATCHING TECHNIQUES

Malhotra S., Anstrom K, Kauf T and Schulman K
Center for Clinical and Genetic Economics, Durham, NC

Purpose: To compare the greedy and optimal matching techniques in a propensity score matched-pair sample. The greedy match is the most frequently used matching algorithm to match cases to controls. Once a match is made, it is fixed. The optimal matching algorithm recomputes all previously made matches before making the current match.

Methods: We developed a propensity score model of medication usage in a cohort of 1819 osteoarthritis patients to match 415 cases to 1489 controls. The model included patient demographics, disease severity indicators, and clinically-pleasurable interactions. We used the SAS macro “matchit” to generate a series of optimal and greedy matched pairs based on propensity scores.

We then estimated the absolute difference between the propensity scores of each matched pair. We used this to obtain the mean absolute difference for the matched set. Lower mean absolute differences indicate closer matches and less bias in the matching algorithm. A bias ratio of absolute difference in propensity score between matched pairs (greedy matched/ optimal matched) was calculated to compare the consistency between greedy and optimal matching. Bias ratio > 1 indicates superiority of the optimal match.

Results: Matched pairs created through optimal matching consistently show smaller absolute differences in propensity scores than pairs developed through greedy matching. The bias ratio is greater than 1 for all sets and it increases as the number of matched pairs increases.

Table Absolute Differences in Propensity Scores

<table>
<thead>
<tr>
<th>Number of matched pairs</th>
<th>Absolute Difference in Propensity Scores (x 10^-7)</th>
<th>Bias Ratio (A/B)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Greedy Matching (A)</td>
<td>Optimal Matching (B)</td>
</tr>
<tr>
<td>350</td>
<td>24.38</td>
<td>22.72</td>
</tr>
<tr>
<td>300</td>
<td>12.86</td>
<td>12.24</td>
</tr>
<tr>
<td>250</td>
<td>8.36</td>
<td>8.16</td>
</tr>
<tr>
<td>200</td>
<td>5.66</td>
<td>5.56</td>
</tr>
<tr>
<td>150</td>
<td>3.61</td>
<td>3.58</td>
</tr>
<tr>
<td>100</td>
<td>2.01</td>
<td>1.99</td>
</tr>
<tr>
<td>50</td>
<td>0.90</td>
<td>0.90</td>
</tr>
</tbody>
</table>

Conclusion: Optimal matching provides more closely matched pairs than the more commonly used greedy matching technique. The greedy match performs poorly when there is intense competition for matched pairs, hence the number of number of matched pairs decreases, and competition for controls increases, the greedy matching performs increasingly biased matches as compared to the optimal match.

POSTER SESSION - CEA: METHODS AND APPLICATIONS; HEALTH SERVICES RESEARCH

ARE RELATIONSHIPS BETWEEN SF-36 SUBSCALES AND HEALTH UTILITY MERELY LINEAR?

Yu A, Yu Y and Nichol M
University of Southern California, Los Angeles, CA

Purpose: The underlying relationship between health status scales (e.g., SF-36 domain) and utility is unknown. Previous studies assumed SF-36 scales mapped linearly to utility measures. Such an assumption provides a simple algorithm, but may be inconsistent with derived power functions, and may yield poor model fitting, lower predictability and regression artifacts. This study was to examine the nonlinear relationships between SF-36 subscales and HUI3 utility.

Methods: Data included 6923 Southern California Kaiser Permanente members who filled both SF-36 and HUI2 in year 1994-1995. Missing values were imputed on item level by MCME and propensity score method with a missing at random assumption. In order to relax the assumption of linearity SF-36 subscales on utility, we used restricted cubic spline functions (CSF) with four or five knots for each subscale except for RF, SF and RE, which had less than ten unique values and were kept as ordinal categories. Subscale scores were transformed to population norm t-scores according to the SF-36 scoring algorithm. The heuristic shrinkage estimate was used to test for model overfitting. Nonlinear relationships by CSF were visualized by plotting utility against each subscale when holding other covariates at their medians. Nonlinearity of each subscale as well as total nonlinearity of the model was examined by F test. The final model included the reduced model with significant nonlinear subscales and interactions between subscales and age.

Results: The following subscales presented statistically significant nonlinear relationship with utility: PF (F=8.34, df=3, p<0.001), VT (F=3.50, df=2, p<0.002), MP (F=11.14, df=3, p<0.001), RP (F=42.81, df=2, p<0.001). Total nonlinearity was significant (F=11.76, df=16, p<0.001). These nonlinear relationships were also evidenced on fitted plots. Interaction terms between age subscales were found to be significant and were included in the final model. Even though 55 parameters were presented in the final model, adjusted R2 increased from 0.490 for simple linear model to 0.538 (total R2 from 0.590 to 0.522) and the heuristic shrinkage estimate of the final model revealed no concern for overfitting.

Conclusions: Some SF-36 subscales show a nonlinear relationship with utility. Researchers should consider the use of nonlinear models in mapping health status to utility.

ABSTRACTS

POSTER SESSION - CEA: METHODS AND APPLICATIONS; HEALTH SERVICES RESEARCH

CANT MELANOMA PATIENTS PREDICT THE QUALITY OF LIFE IMPACT OF AN ALTERNATE MELANOMA STAGE?

Emory University, Atlanta, GA

Purpose: The goal of our study was to evaluate whether melanoma (MM) patients could predict the quality of life impact of a different melanoma stage from their own.

Methods: Utility scores were elicited with a computer-based time trade-off technique. We established 6 different melanoma health states based on stage (I, II, and III) and time from diagnosis, new (e.g. one year ago) or old (e.g. 5+ year ago). The patients’ utility scores for their own melanoma health state and an alternate hypothetical melanoma state were elicited after they were shown a brief presentation of each state’s prognosis, treatment, and reactions of other patients with the same health state. One-way ANOVA and post hoc comparisons were used to analyze mean utilities.

Results: A total of 101 patients (mean age 49.8 years, 54.5% female, and 100% Caucasian), recruited from our melanoma clinics and database, participated in the study. The following table shows mean utilities for actual melanoma health states, as well as mean utilities for hypothetical melanoma health states.

| New Stage | Actual MM Health State Mean Utility (SD) | Hypothetical MM Health State Mean Utility (SD) | Current Patient State%
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>0.93 (0.099)</td>
<td>1.0 (0)</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>0.95 (0.074)</td>
<td>0.98 (0.088)</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>0.95 (0.088)</td>
<td>0.94 (0.080)</td>
<td>8</td>
</tr>
<tr>
<td>II</td>
<td>0.97 (0.058)</td>
<td>0.97 (0.060)</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>0.71 (0.24)</td>
<td>0.73 (0.24)</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>0.52 (0.31)</td>
<td>0.54 (0.31)</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>0.49 (0.50)</td>
<td>0.48 (0.50)</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>0.50 (0.30)</td>
<td>0.50 (0.30)</td>
<td>32</td>
</tr>
</tbody>
</table>

Conclusions: Some SF-36 subscales show a nonlinear relationship with utility. Researchers should consider the use of nonlinear models in mapping health status to utility.

POSTER SESSION - CEA: METHODS AND APPLICATIONS; HEALTH SERVICES RESEARCH

E39

Downloaded from http://medscape.org by guest on June 6, 2007
and Haileyesus T.

Centers for Disease Control and Prevention, Atlanta, GA

Purpose: Cost per case prevented is an important component of analyses designed to assess the cost effectiveness of suicide prevention interventions. Yet, costs of suicide-related injuries have been exclusively limited to the average cost of an emergency department visit or inpatient admission. This study attempts to define episodic costs by exploring the marginal utilization and costs of healthcare services leading up to and immediately following a suicide attempt resulting in hospitalization.

Methods: A private-sector, medical claims database was used to identify persons hospitalized between 1998 and 2001 with a suicide-related diagnosis (ICD-9-CM E950-959), and to assess characteristics of the attempters and their utilization and cost of inpatient services. For each person identified, we then linked to a complementary outpatient claims database to assess average utilization and cost of outpatient services for each of the 12 months preceding and following the inpatient admission date.

Results: Preliminary results from 1998 indicate that for the 134 persons attempting suicide for which an inpatient admission was required, average length of stay was 4 days and average costs were $8,623. The utilization and cost of outpatient services significantly increased 8 months preceding the suicide-related inpatient admission and remained significant until 8 months following the admission (Normal distribution; Kolmogorov-Smirnov goodness-of-fit p = 0.14; 95% CI: $148 - $308, mean $228). Total episodic costs of suicide-related injuries ($8,623), which includes the cost of the inpatient admission and the statistically significant marginal increase in costs of outpatient services, is nearly 50% greater than the costs of the hospitalization alone.

Conclusions: Suicide attempts requiring hospitalization represent just the tip of the healthcare utilization and cost iceberg. These results suggest that arbitrarily-defined episodes of suicide-related injury can be inaccurate and inadequate and that incorporating marginal costs of illness and injury over a statistically defined range will provide policymakers with the best assessment of cost per case prevented. Next steps for prevention include exploring the use of medical claims as a monitoring device for detecting aberrant increases in healthcare utilization as a possible predictor of suicidal behavior.

POSTER SESSION - CEA: METHODS AND APPLICATIONS; HEALTH SERVICES RESEARCH

DEFINING EPISODIC COSTS OF INJURIES: THE CASE OF SUICIDE ATTEMPTS

Cagan E and Halleyesyus T

Purpose: Increasing costs associated with percutaneous coronary interventions (PCIs) have been of concern to health care institutions. A physician led intervention resulting in revised practice guidelines and cost-containment efforts was initiated in an academic referral center, at the same time that new coronary intervention technology (stents and glycoprotein IIb/IIIa inhibitors) was being introduced. This study compares the clinical and economic outcomes associated with PCI procedures in pre and post intervention cohorts.

Methods: Clinical and angiographic data on 1426 pre-intervention and 1738 post-intervention patients were derived from the Mayo Clinic PCI registry. Administrative data and a standardized 2000 constant dollar cost estimate was used to value utilization, in particular to estimate total procedural and post-procedural costs and length of stay (LOS). T-tests were used to compare demographic, clinical and angiographic characteristics as well as to compare observed procedural success rates and economic outcomes between cohorts. Logistic regression and generalized linear modeling was used to estimate the impact of the intervention on procedural success and total costs, respectively, while controlling for patient demographic, clinical, and angiographic characteristics.

Results: The two cohorts were similar in terms of age (60 years), % male (70%), CHF on presentation (8%) and diabetes (23%). The post-intervention cohort had a higher % of patients who received stents (88% vs. 77%), urgent PCIs (48% vs. 36%), prior PTCA (71% vs. 27%), glycoprotein IIb/IIIa inhibitors post PCI (54% vs. 39%), hypertension (64% vs. 59%), moderate/severe left main lesion (46% vs. 41%), and ulcer in any lesion (15% vs. 10%). Procedural success was observed in 91% of patients in both cohorts and did not statistically differ in adjusted analyses. Observed costs per patient were, on average $2,031 lower post-intervention (p<0.001). Model results confirmed a significant economic advantage in the post-intervention period, with a predicted LOS difference of 0.09 days (1.4 vs 2.5), and predicted cost savings of $4,667 (95% confidence interval difference: $4218, $5116).

Conclusion: Physician-led practice management efforts were successful at containing PCI related costs in an era of rapid introduction of new technology while maintaining quality of care. This case study on the process and outcome of cost containment efforts may have implications beyond the cardiovascular setting, to other healthcare settings.
the largest effect, improving adherence by 10.4%. Provider behavior is difficult to change. Future effects on provider adherence. Provider reminders were assessed in two studies and seemed to have practices for the evaluation or management of hypertension.

Effects on improvement in proportion of provider adherence (%IMPROVE) to recommended pharmaceuticals can help promote both quality of care and effective management of pharmaceutical use. Routine, multi-criteria clinical assessment of new drugs would help promote optimal use of these drugs. Routine assessment and integration of these considerations into prescribing decisions expanding the model to include all relevant costs from a societal perspective.

Results: Baseline cost-effectiveness analysis showed that a coordinator who manages 500 patients yearly would reduce further hip fractures from 30 to 21, saving the hospital $104,000. A coordinator was cost-saving 11 over reasonable cost ranges. If only half of patients initiated treatment and only half complied, 3) if treatment efficacy reduced fractures by as low as 20% and 4) if only 220 patients were seen annually.

Conclusion: Employment of a coordinator to manage fragility fracture patients may reduce further hip fractures and is cost-effective from the hospital’s perspective. This analysis did not estimate the full benefit of a coordinator on the prevention of other fractures. The results may change, when expanding the model to include all relevant costs from a societal perspective.

Factors affecting response to and completion of physician surveys

POSTER SESSION - CEA: METHODS AND APPLICATIONS; HEALTH SERVICES RESEARCH

TARGETING FRAGILITY FRACTURES IN AN ORTHOPAEDIC TREATMENT UNIT: COST-EFFECTIVENESS OF A DEDICATED COORDINATOR

Sander E, Maetzel A, Elliot-Gibson V, Beaton D and Bogoch E

University Health Network, Toronto, ON, Canada; St. Michael’s Hospital Hospital, Toronto, ON, Canada

Purpose: The orthopaedic unit at a university teaching hospital hired an orthopaedist (OP) coordinator to manage a collaborative program to identify fragility fracture patients and arrange for investigation and treatment of OP, and patient education. This analysis evaluates the cost-effectiveness of a coordinator in avoiding inpatient hospitalizations due to further hip fractures from the hospital perspective.

Methods: A 1-year decision analytic model was developed combining data from the literature and patient-level data from the first year of the program, during which 430 patients entered the study: age 71 ± 14, female n = 333 (77%), index fracture hip n = 85 (19%), wrist (n = 124, 29%), humerus (n = 72, 17%) and other (n = 49, 11%). OF most likely cause of fracture n = 349 (81%). The decision analysis model calculates the annual incidence of a further hip fracture dependent on type of index fracture (hip, wrist, humerus, other), attribution to OP age and gender. Gender, initiation of OP treatment and compliance modified the incidence of further hip fractures in the presence of a coordinator. The relative risk of further hip fracture varied from 2.2 to 9.8 depending on the index fracture.

Average direct hospital cost of $21,800 for the subset of patients with an index hip fracture were used as a surrogate for the cost of a potential further hip fracture; the cost of a coordinator was $60,000 ± 30% benefit.

Results: Baseline cost-effectiveness analysis showed that a coordinator who manages 500 OP patients yearly would reduce further hip fractures from 30 to 21, saving the hospital $104,000. A coordinator was cost-saving no over reasonable cost ranges. If only half of patients initiated treatment and only half complied, 3) if treatment efficacy reduced fractures by as low as 20% and 4) if only 220 patients were seen annually.

Conclusion: Employment of a coordinator to manage fragility fracture patients may reduce further hip fractures and is cost-effective from the hospital’s perspective. This analysis did not estimate the full benefit of a coordinator on the prevention of other fractures. The results may change, when expanding the model to include all relevant costs from a societal perspective.

FACTORS AFFECTING RESPONSE TO AND COMPLETION OF PHYSICIAN SURVEYS

Breathnach I, Graham I, Visentin L and Stiell I

Ottawa Health Research Institute, Ottawa, ON, Canada

Purpose: Little work has examined whether questionnaire formatting or sender affects physicians’ likelihood to respond to surveys. Furthermore, the extent to which physician surveys are filled out completely and carefully has received little study. We conducted a factorial manipulation of printing format and sender, and examined their effects on physician survey response and completion.

Methods: We asked members of the Canadian Association of Emergency Physicians (CAEP) to complete a 4-page mail survey. The survey was carried out according to standard survey methodology (Dillman, 2000). The 2 x 2 factorial design varied Print Format (single- vs. double-sided printing) and Sender (known [IS, well known to this audience] vs. Unknown [JI, unknown to the audience]) as the manipulated variables. Outcomes included response rate and survey completion, i.e. the number of items skipped without any discernible reason. The sample frame included 400 active CAEP members; those listed as non-MD’s, retired, and not currently residing in Canada were excluded.

Results: Of our initial sample of 400 names, 376 were eligible potential respondents with correct addresses. Overall response rate was 281/376, or 69.4%. Tests for response bias showed no effect of gender on likelihood to respond χ2(1)=0.38, p = 0.52, but a marginal effect of location χ2(1)=3.48, p = 0.06; physicians who came from Ontario were marginally more likely to respond than those coming from other provinces. A 2 x 2 factorial ANOVA with Print Format and Sender predicting response rate showed no significant effect of Print Format (65.6% for double-sided vs. 73.0% for single sided; F1,372) = 2.19, p = 0.12, no significant effect of Sender (72.5% Known vs. 63.6% Unknown; F1,372) = 1.76, p = 0.19, and no interaction F1,372) = 1.33, p = 0.25. Chi-square tests showed that Print Format did not predict the likelihood of leaving more fields blank χ2(3) = 0.70, p = 0.40, but that fewer fields were left blank when the Sender was Known χ2(1)=5.50, p = 0.02.

Conclusions: Print Format and Sender did not affect response rates significantly, although effect sizes of 7.2% and 6.2% might prove important (and statistically significant) with larger sample sizes. The extent to which a questionnaire is fully completed can be a useful variable in survey methodology research.
and quality-of-life data were extracted. We calculated means and 95%-confidence intervals for age, gender, and SF-36 scores using random-effects models for the three patient groups separately and we tested for statistically significant differences between the groups.

Results: 36 studies, reporting on the quality of life in 30,465 patients on renal replacement therapy measured with the SF-36, were included in the analysis. Mean age was 59.3 years for hemodialysis, 53.3 years for peritoneal-dialysis and 43.7 years for transplant recipients. The majority of patients were male: 57% of hemodialysis, 55% of peritoneal-dialysis and 60% of transplant recipients. The major- ity of patients had diabetes: 37% of hemodialysis, 55% of peritoneal-dialysis and 60% of transplant recipients.

Comparing the eight SF-36 dimensions, all patient groups reported lowest scores on the Role-Physical dimension (mean scores: 49.9 among hemodialysis, 33.9 among peritoneal-dialysis and 52.1 among transplant recipients). Highest scores were reported on the Mental-Health dimension (mean score 80.1) and peritoneal-dialysis (mean score 69.3) patients and on the Physical-Functioning dimension for transplant recipients (mean score 66.0). In general, SF-36- dimension scores were significantly lower for hemodialysis and peritoneal-dialysis compared to transplanted patients, except for the Role-Emotional dimension (peritoneal-dialysis scores not signifi- cantly different from scores of transplant recipients) and the Mental-Health dimension (hemodialysis and peritoneal-dialysis scores significantly higher than scores of transplant recipi- ents). Scores of hemodialysis compared to peritoneal-dialysis patients were significantly lower for the Physical-Functioning, Bodily-Pain and Mental-Health dimensions. For the other scales no statistically significant differences were found.

Conclusion: Based on this meta-analysis we conclude that hemodialysis and peritoneal-dialysis patients tend to have a lower quality of life than transplant recipients. Quality of life seems equal or slightly worse for hemodialysis compared to peritoneal-dialysis. Age, duration of therapy, co- morbidities, and other potential predictors of quality of life in patients on renal replacement therapies need to be explored in future studies.
ABSTRACTS

POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISION MAKING
A META-ANALYSIS OF RANDOMIZED CONTROLLED TRIALS WITH CORONARY DRUG-ELUTING STENTS
Saulihi M, Jagannath A and Worej J
Tufts-New England Medical Center, BOSTON, MA
Purpose: To quantify current risks and benefits of coronary drug-eluting stents compared with bare-metal stents.
Methods: Literature search of Medline 1996-2004 and abstracts from major conferences were used. Data relating to death, myocardial infarction (MI), target lesion revascularization (TLR), restenosis, and need for repeat percutaneous transluminal coronary angioplasty (PTCA) or coronary artery bypass grafting (CABG) were extracted from randomized, controlled trials of drug-eluting stents. The DerSimonian and Laird random effects model was used to calculate risk differences (RD).
Results: In all, 13 trials involving 4813 patients were identified but not all contributed to each endpoint. Most patients were 58-64 year old, male, hypercholesterolemic, hypertensive, non-smokers and non-diabetic. Drug-eluting stents compared with bare-metal stents did not significantly affect the baseline risk of death of 1% (CI 0.6-1.4%) or death or MI of 4% (CI 3% to 5%). Drug-eluting stents had a baseline risk of restenosis of 8% (CI 7% to 9%) and TLR of 5% (CI 4% to 6%). Drug-eluting stents compared with bare-metal stents reduced the absolute risk of restenosis by 23%, TLR by 9%, repeat PTCA by 12% and CABG by 1%. Paclitaxel-eluting stents significantly reduced (P<0.001) the risk of restenosis to 9% (RD -14%, CI -20% to -20%) and TLR to 4% (RD -7%, CI -3% to -11%) compared with bare-metal stents. Similarly, sirolimus-eluting stents compared with bare-metal stents reduced the risk of restenosis to 4% (RD -33%, CI -25% to -41%) and TLR to 4% (RD -17%, CI -14% to -20%).
Conclusions: Coronary drug-eluting stents significantly reduce the rates of restenosis and the need for repeat revascularization when compared with bare-metal stents.

POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISION MAKING
RELATIONSHIPS BETWEEN TTO, GENERAL HEALTH STATUS, AND DISEASE-SPECIFIC HEALTH STATUS IN PATIENTS WITH AGE-RELATED MACULAR DEGENERATION
Edwards P, Kongsakorn T, Jacko J and Sainfort F
Georgia Institute of Technology, Atlanta, GA
Purpose: The study explores the relationships between health state utility, measured using the Time Trade-Off Method (TTO), and three non-preference based health measures: two general non-preference-based health measure, the SF-12 questionnaire and the GHQ instrument, and a disease-specific non-preference-based measure, the Visual Functioning Questionnaire-25 (VFQ-25), a questionnaire developed by the National Eye Institute (NEI) to measure visual health function.
Methods: Twenty-nine patients with Age-Related Macular Degeneration (AMD) were asked a series of questions to assess (1) the utility of their current vision by using the TTO method; (2) their general health using the SF-12 as well as the General Health Questionnaire (GHQ); and (3) current visual health using the VFQ-25. In addition, co-morbidities and their visual acuity were measured. Patients were stratified into two groups according to the degree of visual acuity loss in the better-seeing eye. Group 1 (n=15) had 20/20-20/50, group 2 (n=18) worse than 20/50. Utility values obtained from the TTO were then correlated with scores from the above surveys including both the physical component scores (PCS-12) and mental component scores (MCS-12).
Results: The mean TTO for group 1 was 0.55 (SD=0.19) and for group 2 was 0.68 (SD=0.21). For Group 1, correlations between TTO, PCS-12, MCS-12, VFQ, Visual Acuity, co-morbidity, and GHQ indicated that only VFQ was significantly correlated with TTO (p=0.002) and this relationship was very strong (r=0.828). This group also exhibited a moderate relationship with PCS-12 (p=0.008, r=0.524) and MCS-12 (p=0.08, r=0.524). Interestingly, these same results did not extend to Group 2. In fact, none of the above health measures was significantly correlated with TTO for this group.
Conclusions: The results suggest the VFQ-25, a non-preference-based disease-specific measure of visual functioning, was a better predictor of utility than general non-preference-based measures of health for patients recently diagnosed with AMD (group 1). For patients who have progressed further in the disease, no such consistent relationship was identified. It is expected that this finding is due to differences between more severely impaired patients in their ability to cope with their loss of vision. This indicates that disease-specific measures that capture how well the patient is coping with the disease in addition to their level of functioning may be better predictors of utility.
POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISION MAKING

DO FALSE POSITIVE MAMMOGRAPHS ADVERSELY AFFECT QUALITY OF LIFE? RESULTS FROM THE DMIST AGC TRIAL


Dartmouth Medical School, Lebanon, NH; University of Wisconsin, Madison, WI; Brown University, Providence, RI; University of North Carolina at Chapel Hill, Liver- pool, NY; University of North Carolina at Chapel Hill, Chapel Hill, NC

Purpose: A substantial proportion of women who undergo routine screening mammography over a 10-year period will experience a false positive exam, requiring additional work-up to rule out breast cancer. Improved specificity for new breast screening modalities will result in fewer false posi- tive exams. To value improved specificity in economic analyses, our objective was to characterize the impact of false positive mammograms on quality of life, anxiety and attitudes toward future mammography screening.

Methods: Digital Mammography Imaging Screening Trial (DMIST) participants were selected for a quality of life sub-study, with equal sampling of those with positive and negative mammo- grams. Telephone interviews were conducted shortly after the initial screening mammogram and one-year later. At both time points women completed a short form of the Spielbergster state-trait anxi- ety questionnaire (STAIx), EQ-5D, and a current health rating scale (RS). At follow-up, women reported attitudes toward future screening mammography and their willingness to travel to avoid a false positive exam.

Results: We report preliminary data for 1,024 (493 false positive and 531 negatives exams) women ages 27 to 85 (mean age 53) who completed both interviews. Among those with false positive exams, the initial interview was conducted before work-up was complete for 222(24%). At baseline, women who required additional work-up had significantly higher anxiety compared to age-matched women who did not (STAIx 35.2 vs 32.7, but had similar health state values (EQ-5D 0.87 vs 0.88); RS 84.3 vs 85.9). At one year, there were no significant differences in STAIx, EQ-5D, or RS between those with false positive and negative mammograms. Women with false positive exams were signifi- cantly more likely to report that they would undergo future routine screening than those with a nega- tive exam (26% vs. 14%); however, they were no more willing to travel and stay overnight to avoid a false positive exam (80% vs. 86%).

Conclusion: Although false positive screening mammograms were associated with increased anxiety and women were willing to travel to avoid such exams, this did not measurably affect health state values measured by EQ-5D.

POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISION MAKING

MORE NUMERATE PATIENTS HAVE MORE ACCURATE IMPEDEDY JUDGMENTS CONCERNING TREATMENTS OF CERVICAL DYSPLASIA

Humm-K, Smith-K, Hushe S, Turner C, Juniper K and McInroy D

University of Oklahoma Health Sciences Center, Oklahoma City, OK; Centers for Dis- ease Control and Prevention, Atlanta, GA

Purpose and Background: To assess whether imputed efficacy (IE) improves on probability (P) judgment per se as a measure of the gist of patient knowledge of health risks and treatment benefits. By subtracting P(outcome | treatment) from P(outcome | no treatment), the effect of correlated judg- ment errors is removed. Higher IE accuracy in patients with higher numerical literacy (numercy) would support the claim IE is a useful concept.

Methods: 122 female patients who had HPV DNA tests for evaluation of abnormal Pap screens were asked 10 probability questions about the progression of high risk HPV infection with and with- out various treatments. Measures of IE for 5 treatments were calculated. For example, from preci- vical cancer patients, and if not screened, we imputed efficacy of Pap screens for pre- venting dysplasia progression. Accuracy of P and IE was measured using the absolute value of the difference from the same judgments made by a panel of expert clinicians. Patient numercy was measured using TOSTHLA, RSALM, arithmetic, and number comfort. Based on factor analysis, patients were assigned to high and low numercy groups.

Results: 8 of patients’ 10 P judgments were higher than experts’, and 4 of their 5 IE were larger than experts’. Mean absolute error of IE was smaller than MAE of both component probabilities for 3 of the 5 effects. Using the categories based on the overall numercy factor, the more numercate pa- tients had less variable responses than the less numercate in 7 of 10 P and for all 5 IE. Accuracy of different P and IE judgments was related to different numercy subscales. The more numercate pa- tients were less accurate for 2 probabilities, and more accurate for 2 probabilities (all p < .05). How- ever, their IE were more accurate (p < .05) for 3 of 5 IE.

Conclusion: The data support that IE is a better measure of patient understanding of the gist of treatment efficacies than the component P judgments. For most concepts, IE is less variable than P judgments and less inaccurate. Compared with the less numerate patients, the more numerate pa- tients’ IE were less variable for all 5 concepts, and significantly less inaccurate for 3 of 5 concepts.

POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISION MAKING

A DECISION-MAKING TOOL FOR DISTRIBUTION OF HIV PREVENTION RESOURCES

Johnson-Maselli A, Brondino M, Reed B and Pinkerton S

McMaster University, Hamilton, ON, Canada; University of Wisconsin - Milwaukee, Milwaukee, WI; Medical College of Wisconsin, Milwaukee, WI

Purpose: The goal of the study was to develop and pilot a test tool that can be used by state health departments as an aid in deciding which HIV prevention interventions ought to be funded. It serves to assist health department officials to organize their thoughts processes by using explicit crite- ria to evaluate proposed HIV prevention interventions ensure that decision-makers consider all im- portant factors in a systematic and comprehensive manner.

Methods: The model chosen for this project was a multi-attribute utility framework, which may best be thought of as a variant of an expected utility model commonly encountered in research on hu- man judgment and decision making. Variables for the model were derived according to the following steps. First, a broad list of organizational, programmatic, and social attributes was created based on focus groups with health department personnel from across the United States (U.S.). These attrib- utes were identified as being of greatest importance in the decision to fund HIV prevention interven- tions. Second, attributes were weighted by decision-makers (HIV Prevention Planning Community Group Co-Chairs) across the U.S., in relation to their relative importance in the funding process. Third, the attributes were rated by two pilot sites (two state health departments in the U.S.) with re- spect to HIV prevention interventions to be funded at those sites. The multi-attribute model was used to combine attribute weights and ratings for HIV prevention interventions being proposed by local agencies thereby yielding a score for each intervention. Once each intervention had a score the next step entailed a linear programming maximization technique that maximized prevention intervention scores subject to each health department’s budget constraint.

Results: For the first pilot site, the multi-attribute model recommended funding all of the pro- posed interventions, except two. In this case, almost all of the interventions could be funded because the health department budget was slightly smaller than the total cost of all the interventions. A simi- lar situation held for the second pilot site: the budget was sufficient to fund almost all HIV prevention interventions.

Conclusions: Our tool could considered as a welcomed addition to the health departments’ way of funding HIV prevention interventions. The model can help decision-makers maximize the value of spending on HIV prevention.
POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISION MAKING

ABSTRACTS

POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISION MAKING

POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISION MAKING

POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISION MAKING

Abstracts
PERCENTAGE OF THE POPULATION THAT WANTED TO GIVE UP PERFECT HEALTH AND 14.3% WERE UNWILLING TO GAMBLE. FOR THE SEVERE SCENARIO, THE PERCENTAGES WERE 9.1% OF THE PATIENTS AND 17.6% OF THE PUBLIC WERE NON-TRADERS (P = 0.01); 11.6% OF THE PATIENTS AND 4.8 AND 10.5% RESPECTIVELY. STATISTICALLY SIGNIFICANT DIFFERENCES WERE OBSERVED BETWEEN THE PUBLIC ASSOCIATED WITH NON-TRAIDER OR NON-GAMBLER STATUS.

HEALTH STATUS (AFTER CONTROLLING FOR BEING A MEMBER OF THE PATIENT OR PUBLIC GROUPS) WERE NOT ASSOCIATED WITH NON-TRAIDER OR NON-GAMBLER STATUS.

THE RESULTS SHOWED THAT PATIENTS TALKED FOR LESS THAN 1 MINUTE WHILE IT TOOK LESS THAN 3 MINUTES TO REACH A DECISION REGARDING A TOPIC. PATIENTS INITIATED 45% OF THE TOPICS AND EACH TOPIC CONTAINED AN AVERAGE OF 5 TOPICS. THOUGH PATIENTS SPENT MORE TIME IN OTHER TOPICS COMPARED TO IN OTHER TOPICS. ROBUST HAZARD RATIOS FROM THE DURATION MODEL WERE OBTAINED AFTER ACCOUNTING FOR THE CLUSTERING EFFECT OF PATIENT-PHYSICIAN PAIRS.

THE RESULTS INDICATED THAT PHYSICIANS OFTEN DID NOT ADDRESS PATIENTS’ KEY CONCERNS. AFFECTIVE ISSUES PARTICULARLY RECEIVED LITTLE ATTENTION AND WERE FREQUENTLY MISAPPLIED. QUANTITATIVE ANALYSES SHOWED THAT PATIENTS TALKED FOR LESS THAN 3 MINUTES WHILE IT TOOK LESS THAN 3 MINUTES TO REACH A DECISION REGARDING A TOPIC. PATIENTS INITIATED 45% OF THE TOPICS AND EACH TOPIC CONTAINED AN AVERAGE OF 5 TOPICS. THOUGH PATIENTS SPENT MORE TIME IN OTHER TOPICS COMPARED TO IN OTHER TOPICS. ROBUST HAZARD RATIOS FROM THE DURATION MODEL WERE OBTAINED AFTER ACCOUNTING FOR THE CLUSTERING EFFECT OF PATIENT-PHYSICIAN PAIRS.

CONCLUSIONS: A SUBSTANTIAL PROPORTION OF INDIVIDUALS ARE UNWILLING TO TRADE-OFF TIME OR GAMBLE FOR PERFECT HEALTH, ESPECIALLY IF THEY HAVE NOT EXPERIENCED THE DISEASE BEING VALUED. THESE DIFFERENCES ARE ASSOCIATED WITH AGE, EDUCATIONAL STATUS AND TEST PERFORMANCE, AND RAISE EQUITY CONCERNS ABOUT ELICITING PREFERENCES WITH THE CURRENTLY USED TECHNIQUES.

RESULTS: The incremental cost-effectiveness ratio of dalteparin compared with warfarin was $192,726 per QALY gained. Dalteparin yielded a quality-adjusted life expectancy of 1.270 QALYs at the cost of $13,481. Although the dalteparin strategy achieved a slightly higher incremental quality-adjusted life expectancy than the warfarin strategy (difference of 0.034 QALYs), this small clinical benefit was offset by a substantial cost increment of $6,500. Cost-effectiveness results were sensitive to variation of the overall mortality associated with dalteparin and warfarin and the pharmacy costs for dalteparin. Dalteparin cost <$50,000 per QALY only if the pharmacy costs for dalteparin were $571 per day (28% of the drug’s 2002 US wholesale price). In probabilistic sensitivity analysis, the warfarin strategy was considered cost-effective in 97% of Monte Carlo iterations and the dalteparin strategy in 3% at a willingness-to-pay ceiling of $50,000 per QALY gained. If the willingness-to-pay ceiling was increased to $100,000 per QALY gained, the warfarin strategy was optimal in 78% of Monte Carlo iterations and the dalteparin strategy was preferred in 12%.

CONCLUSIONS: Based on the best available evidence, a 6-month course of dalteparin is slightly more effective than a 6-month course with warfarin. However, because of the high pharmacy costs of dalteparin, this drug is very uneconomical economically compared with warfarin.

POSTER SESSION - CLINICAL STRATEGIES: JUDGMENT AND DECISION MAKING
VALIDATING AND UPDATING A PREDICTION RULE FOR NEUROLOGICAL SEQUELAE AFTER CHILDMHOOD BACTERIAL MENINGITIS
Biesheuvel C, Vergouwe Y, Grobbee D and Moons K
University Medical Center Utrecht, Utrecht, Netherlands

AIM: Recently, a prediction rule for developing neurological sequelae after childhood bacterial meningitis was developed on a sample of 170 patients derived from two pediatric teaching hospitals. Before implementing in practice, a rule must be tested in new patients (external validation). Our aim was to study the external validity and, if necessary, to update this rule. Methods: The original prediction rule was developed using multivariate logistic regression analysis and included gender, atypical convulsions in patient history, body temperature at physical examination and type of pathogen. The warfarin strategy was considered cost-effective in 97% of Monte Carlo iterations and the dalteparin strategy in 3% at a willingness-to-pay ceiling of $50,000 per QALY gained. If the willingness-to-pay ceiling was increased to $100,000 per QALY gained, the warfarin strategy was optimal in 78% of Monte Carlo iterations and the dalteparin strategy was preferred in 12%.

CONCLUSIONS: Based on the best available evidence, a 6-month course of dalteparin is slightly more effective than a 6-month course with warfarin. However, because of the high pharmacy costs of dalteparin, this drug is very uneconomical economically compared with warfarin.

RESULTS: The incremental cost-effectiveness ratio of dalteparin compared with warfarin was $192,726 per QALY gained. Dalteparin yielded a quality-adjusted life expectancy of 1.270 QALYs at the cost of $13,481. Although the dalteparin strategy achieved a slightly higher incremental quality-adjusted life expectancy than the warfarin strategy (difference of 0.034 QALYs), this small clinical benefit was offset by a substantial cost increment of $6,500. Cost-effectiveness results were sensitive to variation of the overall mortality associated with dalteparin and warfarin and the pharmacy costs for dalteparin. Dalteparin cost <$50,000 per QALY only if the pharmacy costs for dalteparin were $571 per day (28% of the drug’s 2002 US wholesale price). In probabilistic sensitivity analysis, the warfarin strategy was considered cost-effective in 97% of Monte Carlo iterations and the dalteparin strategy in 3% at a willingness-to-pay ceiling of $50,000 per QALY gained. If the willingness-to-pay ceiling was increased to $100,000 per QALY gained, the warfarin strategy was optimal in 78% of Monte Carlo iterations and the dalteparin strategy was preferred in 12%.

CONCLUSIONS: Based on the best available evidence, a 6-month course of dalteparin is slightly more effective than a 6-month course with warfarin. However, because of the high pharmacy costs of dalteparin, this drug is very uneconomical economically compared with warfarin.
Purpose: The Canadian C-Spine Rule (CCR) was recently derived (2001) and validated (2003) and shown to be highly sensitive for c-spine injury. Other clinical decision rules from our group (e.g. Ottawa Ankle Rules) are reportedly used by over 90% of Canadian emergency physicians. Compared to these, the CCR is newer, more complicated, and involves a higher stakes decision. We conducted a survey to determine current use and attitudes towards this new rule.

Methods: We surveyed members of the Canadian Association of Emergency Physicians about their use of the CCR. Conducted between March and May 2003, the 4-page survey included 24 largely closed-ended questions and 3 sections: one on attitudes towards and use of the CCR, one on use of other clinical decision rules, and a final section on practice details and demographics. The sample frame included 400 active CAEP members, and excluded those listed as non-MEMs, retired, or not currently residing in Canada. The survey was conducted according to standard survey methodology (Dillman, 2000).

Results: Out of a sample of 400 names, 376 were eligible potential respondents with correct addresses. Overall response rate was 261/376, or 69.4%. 216 (82.8%) reported already being familiar with the rule, while 98 (26.2%) reported already using it. Of the 98 (26.2%) who did not currently use the rule, most (73.5%) indicated they would consider using the rule in the future. 4.6 % indicated they thought the rule was too complicated to use. Respondents' attitudes towards the CCR, as indicated by strong or moderate agreement on a 6-point scale: Useful in my practice 87.4%, Not unsafe 85.6%, Efficient use of time, 82.5%, Easy to use 76.1%, Would not increase lawsuits 75.3%, and Safety 74.6%, Easy to learn 74.6%, Easy to remember 60.4%. Factors associated with never having seen the rule include older age (P<218)= 4.84, p = 0.009, non-specialist status (P=213)= 5.78, p = 0.005, part time status (P=246)= 0.08 and community hospital setting (P=248)= 0.06.

Conclusions: A surprising number of Canadian emergency physicians report being aware of and using the CCR. Attitudes towards the rule were generally high; memorability was rated lowest. Widespread use of this relatively new, more complicated, higher stakes clinical decision rule appears to be a reasonable goal.

ABSTRACTS
POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISION MAKING

DETERMINING INDICATIONS FOR CARE COMMON TO COMPETING GUIDELINES BY USING THE CLASSIFICATION TREE METHOD. APPLICATION TO THE PREVENTION OF VENOUS THROMBOEMBOLISM IN MEDICAL INPATIENTS;

Labarre P, Bosson J', Belle L' and Robert C'
CHU BP 217, GRENBOLE Cedex 9, France; TIMC-IMAG, Grenoble Cedex 9, France; CH de la region anschizian, Annecy, France; Institut National de Recherche en Informatique et en Automatique, Marbouinot, France

Purpose: Substantial variations have been reported in the advice given by competing guidelines addressing prophylaxis of venous thromboembolism. The aim of this study was to determine positive and negative indications of prophylactic heparin treatment common to four competing guidelines disseminated in France from 1998 to 2000.

Methods: Retrospectively applied the guidelines to data derived from a cross-sectional study of 818 patients hospitalized in the adult medical wards of a university hospital. For each patient, we determined the number of guidelines recommending the use of prophylactic heparin treatment, discretized into three categories: "0" corresponded to an agreement of the four guidelines to recommend no prophylactic heparin treatment (n = 301), "4" to an agreement of the four guidelines to recommend prophylactic heparin treatment (n = 273), and "1-3" to a disagreement between the four guidelines (n = 244). We displayed the level of agreement between the guidelines by using recursive partitioning analysis, with the number of guidelines recommending the use of prophylactic heparin treatment as the dependent variable and venous thromboembolism risk factors as covariates. We used the C4.5 tree-growing algorithm, which relies on Shannon entropy as an impurity measure of a node and on gain ratio as a splitting criterion. The appropriateness of such indication was illustrated with regard to the rate of deep vein thrombosis detected by systematic compression ultrasound examination.

Results: The resulting classification tree involved ten terminal nodes. Its accuracy estimated by performing tenfold cross-validation was 82% (standard deviation = 3). The covariates determining the structure of the tree included history of venous thromboembolism, acute stroke, recent myocardial infarction, congestive heart failure, current cancer, respiratory insufficiency, bedridden status, acute infectious disease, and varicose veins. Five consistent positive indications of prophylactic heparin treatment were identified. They involved 257 patients (31.4%) and were supported by robust scientific evidence. Deep vein thrombosis was detected in 10.5% (27/257) of these patients. Two consistent negative indications involved 347 patients (42.4%). Deep vein thrombosis was detected in 2.6% (9/347) of these patients. Three indications involving 214 patients (26.2%) were discordant across four guidelines.

Conclusion: Classification tree analysis of real patient data is a useful strategy to identify indications common to competing guidelines. These indications should be considered for inclusion when updating guidelines. Further randomized trials are needed to test discordant indications.

POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISION MAKING

CAN DECISION SUPPORT BE SUCCESSFULLY INTEGRATED INTO CLINICAL CARE?

Dartmouth Hitchcock Medical Center, Lebanon, NH; Dartmouth Medical School, Hanover, NH; Dartmouth Medical School, Lebanon, NH; University of Ottawa, Ottawa, ON, Canada

Purpose: The purpose of this study is twofold: to explore the feasibility of integrating decision support into clinical care; and to determine whether such integration yields benefits for both patients and providers.

Methods: From Feb. 2003 - Mar. 2004, 102 early stage breast cancer patients viewed a Shared Decision-Making® video on therapeutic options prior to surgical consultation. Starting March 2004, newly diagnosed patients with Stage 0-2 cancer (n = 35) were additionally asked to complete a computerized questionnaire addressing: physical and emotional health; distress levels; and decision support needs. The resulting classification tree involved ten terminal nodes. Its accuracy estimated by performing tenfold cross-validation was 82% (standard deviation = 3). The covariates determining the structure of the tree included history of venous thromboembolism, acute stroke, recent myocardial infarction, congestive heart failure, current cancer, respiratory insufficiency, bedridden status, acute infectious disease, and varicose veins. Five consistent positive indications of prophylactic heparin treatment were identified. They involved 257 patients (31.4%) and were supported by robust scientific evidence. Deep vein thrombosis was detected in 10.5% (27/257) of these patients. Two consistent negative indications involved 347 patients (42.4%). Deep vein thrombosis was detected in 2.6% (9/347) of these patients. Three indications involving 214 patients (26.2%) were discordant across four guidelines.

Conclusion: Classification tree analysis of real patient data is a useful strategy to identify indications common to competing guidelines. These indications should be considered for inclusion when updating guidelines. Further randomized trials are needed to test discordant indications.

Abstraction: The decennial booster strategy was dominated by the strategy to give a booster once at 50 years of age. This study was done to determine which booster strategy is optimal.

Methods: We developed Markov cohort model to simulate the economic and clinical consequences (quality-adjusted life years) of 2 tetanus toxoid booster strategies using a hypothetical cohort of fully immunized 28 year olds: boosters every 10 years, or a booster once at 50 years. Data inputs were obtained from the literature for tetanus and vaccine adverse events. Costs in the model were limited to those for vaccines, medical care for tetanus infections, and for those related to vaccine adverse events.

Results: The decennial booster strategy was dominated by the strategy to give a booster once at age 50. The booster-at-50 alternative yielded 76.7 QALY and provided 2,000 (undiscounted) QALYs per 100,000 cohort members. It was costing approximately $600/QALY compared to $840/QALY for the decennial strategy. Results were insensitive to the utilities for tetanus or vaccine adverse events.

Conclusions: In settings where the incidence of tetanus is extremely rare, vaccine adverse events must be taken into account when recommending an appropriate booster policy. We show that the alternative booster strategy for once at age 50 is the optimal decision. Ensuring that everyone receives their primary tetanus vaccine series, boosters with wound management, and a single booster in mid adult life appears to be a safer, more cost-effective way to prevent tetanus infections in this country.
was 57% for MR alone (mean accuracy 78%, standard deviation 3.1%). In a MC-simulation with 10,000 samples, the probability of being the optimal strategy identified. The choice of the best strategy depends on prevalence. If the prevalence is less than 23%, US combination of PET and EUS had the worst performance with only 61% of patients correctly classifying 79% of patients correctly in benign, malignant/resectable, and malignant/unresectable. The result is the “Advanced CKD Management Toolkit” consisting of implementation plans, we sought to maximize the number of patients having both the diagnostic and resectability state correctly classified. In the base case analysis, we used the point estimates of the conceptual framework of TOC, and builds on principles of total quality management (TQM), and tools (e.g. a flow-sheet, a referral form to ease communication between providers) and the meta-tool, establishing functional specifications, and reasons to attack root causes of process failures. Literature search, non-nephrologists, patients. Conducting audits and feedback, organizational change, or financial incentives. We compared strategies in terms of the median effects on changes in blood pressure or changes in the percentage of patients achieving a blood pressure goal (%GOAL). Conclusion: Multiple QI strategies are associated with improved hypertension control. Since most studies included more than one QI strategy it is not possible to discern which strategies have the greatest effects. Future research should define the relative contributions of individual strategies within QI initiatives.

### ABSTRACTS

**POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISION MAKING**

**QUALITY IMPROVEMENT STRATEGIES FOR HYPERTENSION: A SYSTEMATIC REVIEW**

Walsh F, McDonald K, Shojaian K, Sundaram V, Lewis R, Nayak S, Mechanic J, Owens D and Goldstein M

University of California, San Francisco, San Francisco, CA; Stanford University, Stanford, CA; VA Palo Alto Health Care System, Stanford, CA

Purpose: We systematically assessed the effect of quality improvement (QI) strategies on hypertension management. Methods: We searched MEDLINE, Cochrane databases, and bibliographies for experimental evaluations of QI interventions targeting hypertension management. Two reviewers abstracted data and classified each intervention into one or more of the following: provider education, provider reminders, facilitated relay of clinical information, patient education, self-management, patient reminders, audit and feedback, organizational change, or financial incentives. We compared strategies in terms of the mean effects on changes in blood pressure or changes in the percentage of patients achieving a blood pressure goal (%GOAL).

### RESULTS:

64 articles reporting 83 comparisons met inclusion criteria. Overall the median reductions in systolic blood pressure (SBP) and diastolic blood pressure (DBP) were 4.5 mm Hg (interquartile range [IQR]: 1.5, 11.0) and 2.0 mm Hg (IQR: -0.2, 6.8), respectively. Median increases in %GOAL for SBP and DBP were 16.2% (IQR: 10.3, 32.2), and 6.0% (IQR: 1.5, 17.5). Organizational change was associated with median reductions in SBP and DBP of 9.7 mm Hg (IQR: 4.2, 14.0) and 4.2 mm Hg (IQR: 0.6, 6.8), and median increases in %GOAL for SBP and DBP of 21.8% (IQR: 9.3, 16.8) and 17.0% (IQR: 5.7, 24.5). Patient education was associated with median reductions in SBP and DBP of 8.1 mm Hg (IQR: 3.3, 11.8) and 3.8 mm Hg (IQR: 0.6, 6.7) and median increases in %GOAL for SBP and DBP of 19.2% (IQR: 11.4, 33.2) and 17.0% (IQR: 11.4, 24.5). Facilitated relay was associated with relatively small changes in reductions in SBP and DBP of 0.8 mm Hg (IQR: 2.5, 12.3) and 1.3 mm Hg (IQR: 2.8, 6.2), and median increases in %GOAL for SBP and DBP of 25.5% (IQR: 17.0, 34.2) and 2.0% (IQR: 3.6, 5.0). Self-management was associated with median reductions in SBP and DBP of 3.3 mm Hg (IQR: 2.6, 10.1) and 2.8 mm Hg (IQR: 0.4, 6.7), and a median increase in %GOAL for DBP of 9.4% (IQR: 5.3, 11.4). Other strategies were associated with modest improvements in blood pressure outcomes.

Conclusion: Multiple QI strategies are associated with improved hypertension control. Since most studies included more than one QI strategy, it is not possible to discern which strategies have the greatest effects. Future research should define the relative contributions of individual strategies within QI initiatives.
POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISION MAKING
COMPARING SIDE EFFECTS & PERCEIVED MEDICATION EFFECTIVENESS USING NON-STEROIDAL ANTI-INFLAMMATORY DRUGS AND CYCLO-OXYGENASE TYPE II INHIBITORS WITH OSTEOARTHRITIS
Mahboubi S, Anstrom K, Kauf T, Chang J, Omar M, Kahler K and Schulman K

Center for Clinical and Genetic Economics, Durham, NC; Health Economics & Outcomes Research, East Hanover, NJ

Purpose: To compare side effects and perceived medication effectiveness among a cohort of osteoarthritis patients using non-steroidal anti-inflammatory drugs (NSAIDs) and cyclo-oxygenase Type II inhibitors (COX-2 inhibitors).

Methods: 4,386 self-reported osteoarthritis patients completed an internet survey providing comprehensive information on their disease. In addition to other characteristics, they reported medication use, perceived medication effectiveness, and frequency of six conditions possibly associated with osteoarthritis treatment—diabetes, nausea/vomiting, heartburn, stomach pain, headache, and diziness.

From this cohort we identified patients who had been regularly taking osteoarthritis medication (at least 3 times/week) for at least 7 months. Patients were classified as NSAID or COX-2 inhibitor users if they were using NSAID or COX-2 inhibitors without other concurrent osteoarthritis medications.

Potential confounding for treatment selection was adjusted for by developing a propensity score model of COX-2 inhibitor use (vs. NSAID use). The model included patient demographics, disease severity indicators, and clinically- plausible interactions and was estimated by logistic regression. We used the ‘greedy match’ approach to match patients sequentially by 5 to 10,000 patients. Using this new group of propensity-matched COX-2 inhibitor and NSAID users, we compared side effect burden and perceived medication effectiveness.

NSAID users employed, experienced more pain, and missed more hours of work/activity due to their osteoarthritis compared to NSAID users (p<0.002). They had greater difficulty with most daily activities (p<0.05), but reported fewer side effects (p<0.01) and perceived their medication to be more effective than did NSAID users (p<0.01).

The propensity matching model yielded 196 matched pairs of COX-2 inhibitor and NSAID users which revealed no significant differences in demographics or disease severity properties. Propensity matched COX-2 inhibitor users experienced fewer side effects than NSAID users (p=0.048), particularly headaches (p<0.0001). This is significant because headaches were a strong predictor of lower HR-QOL activity in a separate study. 70% of COX-2 inhibitor users perceived their medications to be significantly more effective than did NSAID users (p=0.03).

Conclusion: Among a propensity score matched population of osteoarthritis patients, COX-2 inhibitor users report fewer side effects and perceive their medications to be more effective than NSAID users.

POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISION MAKING
PROSPECT: A NEW PROGRAM FOR COMPUTER-ASSISTED UTILITY ELICITATIONS
Bowers AM and Dale W
St. Michael's Hospital, University of Toronto, Toronto, ON, Canada; University of Chicago, Chicago, IL

We have developed and implemented ProSPEQT (Program to Survey Preferences by Evaluating Quality of Life Tradeoffs), a computer program for computer-assisted utility elicitation interviews. ProSPEQT consists of extensible markup language (XML) documents which are transformed into hypertext markup language (HTML). Health state scenarios are kept in a distinct XML document, allowing for rapid customization. ProSPEQT has modules for the standard gamble (SG), time trade-off (TTO), and rating scale (RS) methods. Visual aids include representation of probabilities using a circle with two arcs or a set of faces (SG), representation of timelines using horizontal bars (TTO), and representation of a rating scale using a vertical “feeling thermometer” (RS). We have applied ProSPEQT to research assessing utilities relevant to Human Immunodeficiency Virus (HIV) infection and prostate cancer. We describe the program’s development and initial experience with 102 HIV-positive participants who evaluated 7 health states, each with 3 methods. ProSPEQT has several features including: the ability to incorporate multiple-choice questionnaires with the utility elicitation, querying of states and methods; storing of results as XML files; recording of each step of an individual’s deliberation and the associated time; allowing respondents to retract responses; and incorporation of audio files. Additionally, ProSPEQT allows for many options including: the method for seeking utilities (bisection; ping-pong, or titration); fixed or random values (with range constraints) for initial utility values; control over fidelity of utility range response and allowance for progression from coarse to fine fidelity during elicitation; and customization of presented survival times to patient-specific age, race, and sex categories using United States life tables. To answer SG, RS, and PT elicitation, participants took a median of 30 [interquartile range 12 to 88], 15 [7 to 54], and 28 [12 to 47] seconds, respectively. The initial SG utility elicitation took 136 [40 to 233] seconds compared to the final SG elicitation, which took 17 [0 to 48] seconds. In 2.8% of responses, participants retracted a response. ProSPEQT is a web-based program for utility elicitation that has multiple features, is broadly customizable across a wide range of health conditions, allows users to select a variety of options when constructing utility elicitation interviews, and can provide detailed information about individual participants’ selections during the interview process. Such programs may help to standardize and de-bias utility elicitions.
Purpose Evidence regarding the cost-effectiveness of adenotonsillectomy for the majority of children currently undergoing this intervention in the Netherlands is lacking. The objective of this study was to assess the balance between costs and effects of adenotonsillectomy in children with mild symptoms or adenotonsilular hypertrophy. Methods Economic evaluation alongside an open randomised controlled trial. Setting: Multi-center: 23 general and 3 university hospitals in the Netherlands. Participants 300 children, aged 2-8 years considered eligible for adenotonsillectomy. Excluded were children with 7 or more throat infections in previous year and those with a high suspicion of obstructive sleep apnoea. Interventions: Adenotonsillectomy versus watchful waiting. Main outcome measure: Incremental cost-effectiveness in terms of costs per episode of fever avoided, per throat infection avoided and per upper respiratory infection avoided at one year were estimated. Results Costs incurred in the adenotonsillectomy group were Euro 171,196 as opposed to Euro 170,004 in the watchful waiting group (49% increase, 100% certain). During a median follow-up period of 22 months children in the adenotonsillectomy group experienced fewer episodes of fever (0.21 vs 0.34, CI 0.54 to 0.12), throat infections (0.21 vs 0.36 to 0.06) and upper respiratory infections 0.53 (0.97 to 0.08) per person per year. Pertaining incremental cost per episode avoided were Euro 170,795, Euro 170,341, Euro 170,744 and Euro 170,798 respectively. Conclusion For the majority of Dutch children currently undergoing adenotonsillectomy, it is i.e., with relatively mild symptoms or adenotonsililar hypertrophy, operation results in a significant increase in cost. Additional research is required to identify subgroups in which operation may be worthwhile.

**ABSTRACTS**

**POSTER SESSION - UTILITY THEORY; HEALTH ECONOMICS; PATIENT & PHYSICIAN PREFERENCES; SIMULATION; TECHNOLOGY ASSESSMENT**

**ALCHEMIST-FRONTIER: A WEB-BASED APPLICATION TO FACILITATE ANALYSIS OF DECISION MODELS COMPARING MULTIPLE POSSIBLE SEQUENCES OF DIAGNOSTIC TESTS**

Sander C, Lin D, Chan J and Gould M

Duke, Durham, NC; Stanford University, Stanford, CA; Veterans Affairs Palo Alto Health Care System, Palo Alto, CA

Many clinical problems involve sequential decisions and multiple strategies. Current analytic software enables decision analysts to determine the optimal first decision, but it is difficult to identify how to proceed downstream decisions and to perform cost-effectiveness analyses. The diagnosis and management of solitary pulmonary nodules (SPN) is one such clinical area in which multiple comparisons of sequential diagnostic tests exist. We sought to develop a software tool which would facilitate the evaluation of such clinical problems.

Using Decision Maker, we developed a Markov model to identify optimal strategies for patients with SPN, comparing 40 possible plausible sequences of diagnostic strategies for SPN management (CT, PET, biopsy, surgery, and watchful waiting). We developed ALCHEMIST-Frontier, a web-based tool which uploads Decision Maker models and evaluates both costs and effectiveness of all possible strategies. ALCHEMIST-Frontier identifies strategies that are eliminated by strict or extended dominance, plots the remaining strategies on an efficiency frontier, and calculates incremental cost-effectiveness ratios by comparing strategies with the next most effective, non-dominated alternative. ALCHEMIST-Frontier also enables the user to reduce the number of potential strategies by either requiring a particular test be performed, or by eliminating any strategies that include a specified test.

ALCHEMIST-Frontier evaluated the SPN model and displayed both graphically and in a table the 40 different strategies and their lifetime costs and life expectancy. In patients with low pre-test probability of malignancy, 35 strategies were eliminated either through strict or extended dominance. Assuming a cost-effectiveness threshold of $50,000/QALY, the optimal strategy began with CT and used PET imaging selectively when CT results were indeterminate, surgery when PET results were positive, and needle biopsy when CT results were benign or when PET results were negative.

When the analyst indicated that PET was unavailable, ALCHEMIST-Frontier eliminated 33 strategies that included PET and 3 of the remaining strategies by strict dominance. The most effective strategy that cost less than $50,000/QALY was to perform CT-guided needle biopsy in all patients, to choose surgery when biopsy results were malignant, and watchful waiting when biopsy results were benign or non-diagnostic.

ALCHEMIST-Frontier facilitates analysis of decision models with sequential decisions and multiple strategies. While we have demonstrated here its utility in the management of SPN, it can be readily used to analyze decision models in other clinical domains.

**POSTER SESSION - UTILITY THEORY; HEALTH ECONOMICS; PATIENT & PHYSICIAN PREFERENCES; SIMULATION; TECHNOLOGY ASSESSMENT**

**EFFECTIVENESS OF ADDING ANTICOAGULATION TO MECHANICAL PROPHYLAXIS FOR PREVENTING DVT IN CRANIOTOMY**

Snegal S, Danish S, Burnett M and Stein S

University of Pennsylvania, Philadelphia, PA

Objective: This study employed a decision analytic model to assess the risks and benefits of adding anticoagulants to mechanical prophylaxis for preventing DVT in the craniotomy patient. Methods: We created a decision tree modeling pneumatic compression devices alone, compression plus minidose unfractionated heparin and compression plus low molecular weight heparin. Post-operative outcomes included DVT, PE and intracranial hemorrhage. Probabilities for the tree were obtained from a MEDLINE search of English-language literature. Searches were supplemented by reviewing the bibliographies of selected articles. We calculated weighted averages of expected complications from each treatment option. We included only clinically evident DVT and PE and only cases of post-craniotomy hemorrhage that required emergency reoperation.

Results: The outcomes of all three strategies are close to 1 because complications, no matter how severe, are uncommon. We used sensitivity analyses to examine the robustness of the apparent superiority of the no heparin strategy. The incidence of pulmonary embolism had the greatest influence on outcome. Its incidence (without heparin) was varied from 0 to 8%. It is the only variable which impacts strategy; if the incidence doubles from baseline to 1.4%, adding heparin begins to yield better outcomes than mechanical prophylaxis alone; strong DIF associated with the PE incidence doubles.

Conclusion: The clear advantage of heparin prophylaxis in general surgery is offset by increased risk of intracranial hemorrhage in craniotomy patients. The effectiveness of the methods are similar, but mechanical prophylaxis without heparin looks best unless the PE incidence doubles.

**POSTER SESSION - UTILITY THEORY; HEALTH ECONOMICS; PATIENT & PHYSICIAN PREFERENCES; SIMULATION; TECHNOLOGY ASSESSMENT**

**DEMOGRAPHIC DIFFERENCES IN ITEM RESPONSES TO PHYSICAL FUNCTIONING AND MENTAL HEALTH DOMAINS OF THE SF-36**

Yu Y, Yu A and Ahn J

University of Southern California, Los Angeles, CA

Purpose: Differential item functioning (DIF) is the observation that individuals with equal ability from different groups present unequal probabilities of endorsing a certain item. The purpose of this study is to investigate whether demographic characteristics have uniform DIF (effect extent is same for all ability levels) or non-uniform DIF (effect extent varies with ability level) effect on items in physical functioning (PF) and mental health (MH) domains of the SF-36, and to examine overall validity of PF and MH items from the DIF perspective.

Methods: Those who completed SF-36 between 04/1994 and 02/1995 (n=738) were extracted from a cohort of Southern California Kaiser Permanente members; Ordinal logistic regressions (OLR) were applied separately to investigate uniform and non-uniform DIF effect of seven demographic variables (including age, gender, ethnicity, education, income, marital and employment status through three steps: entering 1) matching variable, 2) group variable, and 3) interaction term of matching and group variables. The first and last two models were compared to determine the presence of uniform and non-uniform DIF using log-likelihood test and the effect size of DIF based on difference of R-square (Æ.035 negligible, 0.035~0.070 moderate, or >0.070 large).

Results: A total of 210 scenarios (combinations of 15 items and 7 group variables) were examined. After matching on the latent ability level represented by total score of PF or MH domain, all PF and MH items displayed uniform (U) or non-uniform (NU) DIF across at least one demographic group. The items were more sensitive to difference of age (12 items-U, 8-NU), ethnicity (11-U, 11-NU), and income (10-U, 16-NU), but less to marital status (7-U, 4-NU). The extent of variation in item responses due to different education (10-NU) and employment status (11-NU) were more likely to change with the level of underlying ability than the extent of variation due to difference of other demographic attributes. As an overall measure, the effect sizes of present DIF across all demographic groups were small (VAR < 0.035), evidencing satisfactory DIF validity.

Discussion: Inclusions in this study, PF and MH items displayed results in several demographic domains. Strong DIF associated with different demographic characteristics. From the DIF perspective, the study revealed the satisfactory validity of PF and MH domains on the item level.
Results: Excluding 38 deaths, there were 91 ARE in 53 patients (e.g., opportunistic infections, pneumocystis carinii pneumonia and 205 SAE in 88 patients (e.g., chest infection, anemia); utility data were available for a total of 101 non-concurrent events. Mean utility scores did not change significantly from baseline (HUI3 0.58, EQ-5D 0.66) to pre-event (HUI3 0.57, EQ-5D 0.63) for patients with ARE and SAE, and were not significantly different from controls' baseline scores (HUI3 0.60; EQ-5D 0.72). Mean HUI3 scores decreased pre-to post-event for patients with ARE (-0.06, p = 0.07) and SAE (-0.06, p = 0.02), while mean EQ-5D scores fell slightly (ARE: -0.04, p = 0.42; SAE: -0.07, p = 0.09). There were no significant changes in utility scores among controls. Comparing the decrement in pre-to post-event scores in ARE/SAE patients relative to the decrement in controls, we found that both ARE and SAE were associated with significant declines in HUI3 (ARE: p=0.05; SAE: p=0.02) but not EQ-5D score.

Conclusions: Both AIDS and non-AIDS clinical events were associated with declines in health-related quality of life in patients with advanced multi-drug resistant HIV. While HUI3 and EQ-5D scores both decreased, HUI3 showed a greater and significant decline associated with serious clinical events in this population. Formal decision analysis will indicate whether the observed utility changes translate into important differences when estimating the quality-adjusted life expectancy and cost-effectiveness of therapies for advanced HIV disease.
ABSTRACTS

POSTER SESSION - UTILITY THEORY; HEALTH ECONOMICS; PATIENT & PHYSICIAN PREFERENCES; SIMULATION; TECHNOLOGY ASSESSMENT
IDENTIFICATION OF MEN WHO WANT HELP WITH PROSTATE CANCER TREATMENT DECISIONS

Knight S1, Chen M1, Sands L1, Grollman J1, Nadler R1 and Bennett C1
1San Francisco VA Medical Center, San Francisco, CA; 2Purdue University, West Lafayette, IN; 3Rosalind Franklin University of Medicine and Science, North Chicago, IL; 4Feinberg School of Medicine at Northwestern University, Chicago, IL

Purpose: We sought to understand the role men with prostate cancer would like to play in treatment decision making and to evaluate a multivariate model to identify those who want help from their physicians in making decisions.

Methods: We recruited a cross-sectional sample of 192 men with biopsy confirmed prostate cancer seen in private and university-affiliated clinics. In a telephone interview, the men responded to a question about their preferences to participate in treatment decision making and completed measures of health-related quality of life, prostate cancer specific symptoms, optimism, and perceived involvement in care. We used descriptive statistics and stratified analysis by demographic characteristics to analyze the results, as well as compiling free text responses by theme.

Results: Over 75% of the men preferred an autonomous or active role in decision making, i.e., 56% wanted to make the decision themselves after considering physician advice, 22% wanted shared decision making. A minority wanted the physician to make the decision after considering their preferences (14%). Few (1%) wanted to make decisions without physician input. White men who were in better health (i.e., better physical function, fewer symptoms) and were more optimistic/predicted greater autonomy in decision making (p = 0.004). When men used their physicians as discouraging shared decision making, they wanted greater autonomy (p = 0.03). However, a multivariate model including these variables did not accurately discriminate among patients according to their preferences for decision making involvement (67% correct classification). Compared to those who wanted a greater patient role in decision making, it was more difficult to classify those who wanted greater physician involvement.

Conclusions: Most men want to play a role in prostate cancer decision making, but do not want to be abandoned with the decisions. Men who want greater physician involvement in treatment decision making are more difficult to identify than those who prefer greater autonomy. Direct methods for assessing patient preferences for involvement in decision making are needed.

POSTER SESSION - UTILITY THEORY; HEALTH ECONOMICS; PATIENT & PHYSICIAN PREFERENCES; SIMULATION; TECHNOLOGY ASSESSMENT
COMMUNICATING EFFECTS OF RISK INTERVENTIONS IN TERMS OF POSTPONEMENT OF ADVERSE EFFECT

Kristiansen I
University of Southern Denmark, Odense, N-8636 Oslo, Norway

Objective: Previous research has shown that effect measures such as relative risk reduction or number-needed-to-treat may be misunderstood by lay people. We hypothesised that, for lay people, postponement of adverse events is a more understandable measure of treatment benefit.

Methods: A random sample of non-institutionalised individuals aged 40 years and older (n = 1,743) was asked to be interviewed in-person, and 1,367 (79%) accepted. The respondents were representative of the target population in terms of sex and age. The respondents were asked to imagine they were at increased risk of heart attack, and were offered a hypothetical pharmaceutical drug that would reduce this risk. The respondents were informed that the drug would postpone the heart attack by, randomly, 1 month, 6 months, 12 months, 2 years, 4 years or 8 years. Information on baseline risk of heart attack was given to half of the respondents by random.

Results: In total, 58% of the respondents accepted the hypothetical therapy, 30% rejected it and 12% were uncertain. The proportions consenting were 35%, 52%, 64%, 67% and 73% with increasing postponement, and it was lower among those informed about baseline risk than the others (54% versus 62%). The information on health benefit was perceived as easy to understand by 61% of the respondents.

Conclusion: Postponement of adverse events seems to be understandable for lay people, and they are responsive to differences in delay. The results of the study indicate that postponement may be a better way than risk measures when explaining the benefits from interventions for chronic disease processes such as atherosclerosis or osteoporosis.

ABSTRACTS

POSTER SESSION - UTILITY THEORY; HEALTH ECONOMICS; PATIENT & PHYSICIAN PREFERENCES; SIMULATION; TECHNOLOGY ASSESSMENT
HEALTH RELATED QUALITY OF LIFE MEASURED BY SF-36 FOR ADULTS WITH DIABETES: A META-ANALYSIS

Neris S, Zhang X, Chowdhury F and Zhang P
Centers for Disease Control and Prevention, Atlanta, GA

Purpose: The objective of this review is to assess health related quality of life (HRQOL) among persons with diabetes mellitus and to obtain pooled estimates for various subpopulations.

Methods: A systematic review of the literature was performed for studies measuring HRQOL among persons with diabetes using the SF-36 questionnaire. For intervention studies, only the HRQOL measure at baseline was used. Pooled estimates were obtained using a random effects model.

Results: A total of 49 studies with 15,823 participants were included in our analysis. The 8 component scores of the SF-36 ranged from 50.3 (general health) to 74.6 (social function), where 0 represents the lowest and 100 the maximum level of health. The mental health summary score was 51.6 and the physical health summary score was 49.4. Among persons with type 1 diabetes, both summary scores, general health status, social function, and mental health were similar to scores among persons with type 2 diabetes. Physical function, role limitation due to physical problems, and role limitation due to emotional problems were scored higher (i.e. less disability) among those with type 1 diabetes (all unadjusted). Persons aged <65 years had lower component scores than persons >65. Persons with diabetes diagnosed <1 year previously (3 studies) had lower summary scores than persons with disease for >1 year. Compared to scores previously reported among persons with diabetes [1], our pooled estimates were lower for scores of social function, mental health, and role limitation secondary to mental problems, and higher for scores of physical function. All component scores were much lower than those of the US general population [1].

Conclusions: We present estimates of HRQOL achieved by pooling data from a large number studies. Diabetes significantly affects both the mental and physical components of HRQOL as measured by the SF-36 and some subpopulations are affected more than others. These estimates provide useful population norms.

ABSTRACTS

POSTER SESSION - UTILIT Y THEOR Y; HEALTH ECONOMICS; PATIENT & PHYSICIAN PREFERENCES; SIMULATION; TECHNOLOGY ASSESSMENT

BREAST CANCER PREVENTION: A FRAMEWORK FOR DECISION MAKING

Perencevich E, Adducu K, Annis C and Isserlin L
Massachusetts General Hospital, Boston, MA; UCSF Medical Center, San Francisco, CA

Purpose: Develop and evaluate a clinical decision aid using a framework for breast cancer prevention care that provides clinical data in context to appropriately motivate women to choose interventions.

Methods: The decision algorithm includes a general health and breast cancer risk assessment using both the Gail and Claus risk models. Standard format for data presentation was implemented, using absolute risk information with consistent graphical presentation. Breast cancer risk over time is presented in the context of age-matched women and evidence-based models using biomarkers as risk discriminators and predictors of benefit from interventions. Physicians providing the prevention care were trained to use the shared decision making framework. Pilot testing in a randomized clinical trial compared physician training only versus physician training with use of the decision aid. Follow-up data was gathered at 6-12 months post consultation to track patient decisions.

Results: The shared decision making framework appeared to increase patient interest in prevention interventions. Initially, 13% of patients were interested in prevention interventions before the consultations, as compared to 23% after the consultation. This rate returned to baseline at follow-up. Similarly, patient interest in risk refinement interventions increased from 7% to 23%. After follow-up, this rate had decreased to 19%. The feasibility outcomes showed that the decision aid did not interfere with the consultation as measured by consultation duration, user satisfaction, patient knowledge and decisional conflict.

Conclusion: Our results suggest that the decision aid is feasible for use in the consultation room. The decision framework provides access to key information during consultations and allows the integration of emerging biomarkers in the prevention setting. As compared to previously studied clinical behavior, the framework increased patient interest in both prevention interventions and learning more about their level of risk. The tendency for these rates to return to baseline at follow-up suggests the need for ongoing prevention decision support. Future applications of the decision aid include a randomized trial of three arms (a control arm, an arm with physician training only, and an arm with physician training and use of the decision aid) to determine impact on decision-making. The integration of tools to store, track, and present data to patients and physicians will be studied.

POSTER SESSION - UTILIT Y THEOR Y; HEALTH ECONOMICS; PATIENT & PHYSICIAN PREFERENCES; SIMULATION; TECHNOLOGY ASSESSMENT

THE HAPPINESS GAP: A MATTER OF CONTEXT?

Pand H, Fagerlin A, Goldstein C, Smith D, Loewenstein G and Ubel P
VA Ann Arbor Health Care System, Ann Arbor, MI; University of Michigan, Ann Arbor, MI; Carnegie Mellon University, Pittsburgh, PA

Purpose: When healthy non-patients are asked to estimate the happiness associated with a given health condition, their estimates typically differ from those made by patients who have experienced that condition. This happiness gap might occur because patients’ experience provides contextual knowledge about the relative severity of the condition that is unavailable to non-patients—they are better able to appreciate the emotional impact of mild versus severe disease. This study tested whether non-patients distinguish more between different lung conditions when provided with contextual information about disease severity and whether increased discrimination reduces the happiness gap.

Methods: We recruited a convenience sample of 248 non-patients from a hospital cafeteria, and 158 patients with chronic bronchitis or emphysema from a hospital pulmonary clinic. Patients and non-patients were randomly assigned either to provide QoL estimates for each of 5 lung conditions of varying severity (cont’d condition), or only 1 of those possible conditions (no context information). QoL estimates were made on a scale of 0 (as bad as death) to 100 (perfect health).

Results: Non-patients’ QoL ratings for the 5 lung conditions were more broadly distributed in the context condition (range = 43.13) than the no context condition (range = 15.81). However, non-patients’ ratings were still more narrowly distributed than patients’ ratings in both the context (range = 66.89) and the no context condition (range = 28.66), and the happiness gap was not eliminated in the context condition.

Conclusions: Providing contextual information to non-patients about the relative severity of a given health condition increases their discrimination among these conditions, but does not eliminate the gap between patients’ and non-patients’ ratings. Contextual information does not equalize patients’ and non-patients’ perceptions of employments and thus cannot readily account for the happiness gap.

POSTER SESSION - UTILIT Y THEOR Y; HEALTH ECONOMICS; PATIENT & PHYSICIAN PREFERENCES; SIMULATION; TECHNOLOGY ASSESSMENT

COMPARISONS AMONG PATIENT-DERIVED STANDARD GAMBLE VALUES AND SF-6D AND EQ-5D VALUES DERIVED FROM THE SF-12 IN URBAN PATIENTS WITH POOR HEALTH

Schneider R and Teixeira P
Weill Medical College of Cornell University, New York, NY

Purpose: Community-derived utility values are recommended for cost-effectiveness analysis and often result in lower health state valuations than patient-derived values. Algorithms have been developed for determining community-derived values, using the SF-6D and EQ-5D, from the SF-12. We assessed within-subject agreement among these community-derived scores and between these scores and patient-derived standard gamble (SG) values for current health.

Methods: 92 patients with chronic hepatitis C infection recruited at urban HIV, methadone, and gastroenterology clinics to participate in a survey study completed a computer-assisted standard gamble assessment of current health and the SF-12. Score distributions were compared for one SF-6D algorithm and two EQ-5D algorithms by scatterplot, Pearson correlation coefficient, bivariate linear regression, and intra-class correlation coefficient (ICC). Differences from patient-derived standard gamble values were compared by paired t-test.

Results: Score intervals for SF-6D and the 2 EQ-5D algorithms were (0.35-0.86), (0.15-0.98), and (0.71-0.93). The correlations between the SF-6D and each of the EQ-5D scores were 0.79 and 0.85, agreements by ICC were 0.62 and 0.38. The correlation between the EQ-5D scores was 0.92 and agreement by ICC was 0.73. The mean (SD) differences from patient-derived SG score -0.33 (0.33) for the SF-6D, and -0.14 (0.43) and -0.11 (0.28) for the 2 EQ-5D algorithms. The SF-6D had a significantly greater mean decrement from SG than both EQ-5D algorithms.

Conclusions: Because lower boundaries differ between SF-6D and EQ-5D, there is not strong agreement between these scores for poor health states. SF-6D scores indicate greater differences between how community members and patients value health than EQ-5D scores.

POSTER SESSION - UTILIT Y THEOR Y; HEALTH ECONOMICS; PATIENT & PHYSICIAN PREFERENCES; SIMULATION; TECHNOLOGY ASSESSMENT

POOR HEALTH

Schneider R and Teixeira P
Weill Medical College of Cornell University, New York, NY

Purpose: Community-derived utility values are recommended for cost-effectiveness analysis and often result in lower health state valuations than patient-derived values. Algorithms have been developed for determining community-derived values, using the SF-6D and EQ-5D, from the SF-12. We assessed within-subject agreement among these community-derived scores and between these scores and patient-derived standard gamble (SG) values for current health.

Methods: 92 patients with chronic hepatitis C infection recruited at urban HIV, methadone, and gastroenterology clinics to participate in a survey study completed a computer-assisted standard gamble assessment of current health and the SF-12. Score distributions were compared for one SF-6D algorithm and two EQ-5D algorithms by scatterplot, Pearson correlation coefficient, bivariate linear regression, and intra-class correlation coefficient (ICC). Differences from patient-derived standard gamble values were compared by paired t-test.

Results: Score intervals for SF-6D and the 2 EQ-5D algorithms were (0.35-0.86), (0.15-0.98), and (0.71-0.93). The correlations between the SF-6D and each of the EQ-5D scores were 0.79 and 0.85, agreements by ICC were 0.62 and 0.38. The correlation between the EQ-5D scores was 0.92 and agreement by ICC was 0.73. The mean (SD) differences from patient-derived SG score -0.33 (0.33) for the SF-6D, and -0.14 (0.43) and -0.11 (0.28) for the 2 EQ-5D algorithms. The SF-6D had a significantly greater mean decrement from SG than both EQ-5D algorithms.

Conclusions: Because lower boundaries differ between SF-6D and EQ-5D, there is not strong agreement between these scores for poor health states. SF-6D scores indicate greater differences between how community members and patients value health than EQ-5D scores.
POSTER SESSION - UTILITY THEORY; HEALTH ECONOMICS; PATIENT & PHYSICIAN PREFERENCES; SIMULATION; TECHNOLOGY ASSESSMENT

A FOUR-YEAR UNDERGRADUATE SHARED MEDICAL DECISION MAKING CURRICULUM: EVALUATION OF YEAR 1

Schwartz A, Wątkaus C, Baker E, Carter E and McNutt R

1University of Illinois at Chicago, Chicago, IL; 2 Rush-Presbyterian-St. Luke's Medical Center, Chicago, IL

Purpose: During 2002-2003, we implemented the first year of a four-year longitudinal shared medical decision making (SMDM) curriculum. We report on the curriculum and evaluate the impact of this course on participants.

Curriculum: The four-year curriculum has ten objectives for students, including shared decision-making approaches, probabilistic reasoning, clinical utility assessment, and application of expected-utility decision models. In the M1 year, 18 hours of curricular time were devoted to the course, including large-group didactic sessions, small-group case-based sessions, and preceptor-based experiences.

Evaluation: Tools included an objective test of cognitive skills, an attitude survey about the content and course, summative course evaluations, and two post-course focus groups. One hundred thirteen M1 students and 97 concurrent M2 students enrolled in the course in their completed the instruments. Analysis of data was performed using Student's t-test and Chi-square analysis.

Results: (a) M1 students significantly outperformed M2 students in 9 out of 12 cognitive skills items, and had a significantly higher total correct score. (b) Attitudes about course concepts loaded on three factors, interpreted as: "comfort with quantifying risk and value," "importance of patient decision participation," and "importance of physician decision participation." M1 students held both patient and physician decision participation significantly more important than M2 students. There was no difference in comfort with quantification. (c) Both M1 and M2 students rated anatomy as more useful to them as a clinician than behavioral science or SMDM. M1 students further rated behavioral science as more useful than SMDM. Both groups also rated anatomy as more useful to them as a patient than the other courses; there were no differences between ratings of behavioral science and shared decision making in this context. (d) Course evaluations and focus groups revealed dissatisfaction with course process and conflicting messages about decision making from other M1 courses.

Conclusions: Students were dissatisfaction with the first offering of the M1 course, and did not perceive it as useful. Nevertheless, students improved SMDM-relevant cognitive skills as compared to peers, and perceived greater roles for patient and physician in decision making. Changes are planned to improve student experiences while maintaining educational impact.

POSTER SESSION - UTILITY THEORY; HEALTH ECONOMICS; PATIENT & PHYSICIAN PREFERENCES; SIMULATION; TECHNOLOGY ASSESSMENT

IMPACT OF DISCUSSION ON UTILITY VALUES ELICITED IN A GROUP SETTING

Stein K, Grabb T, Braizer J, Ratcliffe J, Millie K and Round A

1Partnerships in Medical School, Exeter, United Kingdom; 2School of Health and Related Research, Sheffield, United Kingdom; 3University of Southampton, Southampton, United Kingdom

Purpose: To investigate the impact of group discussion following individual preference elicitation.

Methods: A group of 15 non-health professionals was established to measure preferences on a range of health states using the standard gamble method. The group met five times over six months. Health state scenarios were derived from disease specific outcome measures reported in clinical trials of health technologies. Preferences were initially elicited using the standard gamble (titration approach) without discussion in the group. Each scenario and the initial preferences expressed by participants in qualitative interviews.

Results: 40 health states relating to seven specific conditions were valued, giving 445 individual utility values. Mean utility values ranged from 0.27 to 0.98. Changes were made for 12 scenarios (40%) by seven individuals. Changes ranged from 0.03 to 0.45 (mean 0.04). The impact on summary values was limited. Mean utility was affected in only 7 of the 12 scenarios. The average mean change following discussion was 0.01 (range -0.01 to 0.07). Only three median values were affected by changes (range -0.05 to 0.03). Qualitative data revealed a range of perceived benefits from discussion.

Conclusions: Preferences elicited using the standard gamble are unaffected by brief discussion in a group. Changes have no significant impact on the summary measures but participants revised their initial preference following discussion. The number of changes made and their impact were highly variable.

POSTER SESSION - UTILITY THEORY; HEALTH ECONOMICS; PATIENT & PHYSICIAN PREFERENCES; SIMULATION; TECHNOLOGY ASSESSMENT

ACHIEVING OPTIMAL OUTCOMES AFTER RADICAL PROSTATECTOMY: ASSESSING POSTOPERATIVE TRAJECTORY USING A MARKOV MODEL

Ekiin E, Saranchuk J, Touijer A, Seiler P, Fox A and Kurtz M

1Memorial Sloan-Kettering Cancer Center, New York, NY

Purpose: To investigate the impact of group discussion following individual preference elicitation.

Methods: A group of 15 non-health professionals was established to measure preferences on a range of health states using the standard gamble method. The group met five times over six months. Health state scenarios were derived from disease specific outcome measures reported in clinical trials of health technologies. Preferences were initially elicited using the standard gamble (titration approach) without discussion in the group. Each scenario and the initial preferences expressed by participants in qualitative interviews.

Results: 40 health states relating to seven specific conditions were valued, giving 445 individual utility values. Mean utility values ranged from 0.27 to 0.98. Changes were made for 12 scenarios (40%) by seven individuals. Changes ranged from 0.03 to 0.45 (mean 0.04). The impact on summary values was limited. Mean utility was affected in only 7 of the 12 scenarios. The average mean change following discussion was 0.01 (range -0.01 to 0.07). Only three median values were affected by changes (range -0.05 to 0.03). Qualitative data revealed a range of perceived benefits from discussion.

Conclusions: Preferences elicited using the standard gamble are unaffected by brief discussion in a group. Changes have no significant impact on the summary measures but participants revised their initial preference following discussion. The number of changes made and their impact were highly variable.

POSTER SESSION - UTILITY THEORY; HEALTH ECONOMICS; PATIENT & PHYSICIAN PREFERENCES; SIMULATION; TECHNOLOGY ASSESSMENT

DIFFERENCES BETWEEN PATIENTS' AND PHYSICIAN'S PERCEPTIONS OF BENEFIT IN PHASE I CLINICAL TRIALS: METHODOLOGICAL AND BIOETHICAL CONSIDERATIONS

Weinfurt K, Depuy V, Castoll L, Sells D, Sulmary D, Schulman K and Menopel N

1Duke Clinical Research Institute, Durham, NC; 2St. Vincent's Hospital Manhattan, New York, NY; 3Fox Chase Cancer Center, Philadelphia, PA

Purpose: Patients' expectations of benefit from phase I trials often exceed their physicians' estimates. We hypothesized that patient-physician differences are related more to differences in the way patients and physicians use the quantitative response scale than to differences in each person's underlying qualitative sense of the relative likelihoods of various treatment outcomes.

Methods: Adult cancer patients considering participation in phase I trials along with their physicians, completed surveys before initiating therapy. Patients and their physicians rated the likelihood (0% to 100%) of benefit and harm from experimental and standard therapies, yielding a profile of 4 likelihood ratings for each person. We assessed each patient-physician pair's differences in use of the response scale by examining differences in the mean of each person's ratings ("level") and differences in the standard deviation of each person's ratings ("scatter"). In addition, we characterized the relative "ups and downs" of the ratings in each subject's profile as the "shape" of that profile. We assessed patient-physician differences in the qualitative sense of the relative likelihoods of treatment outcomes by correlating the shape of each patient's profile with the shape of the physician's profile. We then correlated discrepancies in expectation of benefit from experimental therapy with differences in level, scatter, and shape.

Results: On average, patient's profiles were 28.46 points higher in level and displayed greater scatter (difference in SD, 15.13) than their physicians' profiles. The average correlation between the shapes of patients' and physicians' profiles was 0.63, reflecting poor agreement in the qualitative sense of the likelihoods of the 4 treatment outcomes. Disagreement in the likelihood of benefit from experimental therapy was associated with disagreement in level (r=0.68; 95% CI, 0.56 to 0.80) and scatter (r=0.45; 95% CI, 0.32 to 0.59) but not shape (r=0.02; 95% CI, -0.09 to 0.13).

Conclusions: The magnitude of discrepancy between patients and physicians in expectation of benefit from experimental therapy was related to the way subjects used the response scale but was unrelated to how well patients and physicians agreed in their qualitative sense of the relative likelihoods of each treatment outcome. If treatment decisions are made on the basis of this qualitative sense, patients and physicians might disagree substantially about chance of benefit from experimental therapy yet still arrive at the same decision.
POSTER SESSION - UTILITY THEORY; HEALTH ECONOMICS; PATIENT & PHYSICIAN PREFERENCES; SIMULATION; TECHNOLOGY ASSESSMENT

CREATING A SYNTHETIC POPULATION OF INDIVIDUALS FROM PUBLIC AND DE-IDENTIFIED DATA USING A MODIFIED ITERATIVE PROPORTIONAL FITTING ALGORITHM

Fridsma D and Roberts M
University of Pittsburgh, Pittsburgh, PA

Purpose: The Health Insurance Portability and Accountability Act of 1996 (HIPAA) extended important privacy protections over an individual’s health information, but made it difficult to use that information for research purposes. In this report, we describe the application of an iterative proportional fitting algorithm (IPF) to census and hospital discharge data to create a synthetic population of individuals. Using only public and de-identified medical data, we created a synthetic population of individuals that was statistically equivalent to the real population of patients in Allegheny County and suitable for agent-based and micro-simulation studies.

Methods: We used three sources of data to create our synthetic population—the 1990 Census Public Use Microdata Sample (PUMS), the Census Summary Tape File 3A (STF-3A), and de-identified hospital discharge data from the MARS database at the University of Pittsburgh. For the datasets in which we only had summary statistical data and in which the cross-tabulations shared data elements with the synthetic population, we used a modified IPF algorithm to integrate this information and derive individual-level data (PUMS, STF-3A). For data in which we had individual-level but de-identified data (MARS datasets), we used a probabilistic matching algorithm to integrate this information into the synthetic population. Identifiers of age, gender, nationality, and zip code location, were used to match records in our synthetic population with records obtained from MARS. We made no assumptions as to the underlying joint distributions of the data fields. This assumption simplified the task of creating the synthetic population, and generated a solution (of many that are possible) that fit the data used for the synthetic population.

Results: Using both IPF and a probabilistic matching algorithm, we created a 1.3 million synthetic individual and household level records representative of Allegheny County. The dataset was constructed using only public census and de-identified data, yet contained detailed individual and household level data. We maintained the marginal statistics for the data, but filled in the cells of the tables with anonymized, but representative data.

Conclusions: This work showed that IPF is a suitable technique to generate household and individual level data set of patients from publicly available and de-identified data. It applied a number of well-tested mathematical processes that have been used for census data to medical datasets.

POSTER SESSION - UTILITY THEORY; HEALTH ECONOMICS; PATIENT & PHYSICIAN PREFERENCES; SIMULATION; TECHNOLOGY ASSESSMENT

GENETIC PROGRAMMING OR MULTIVARIABLE LOGISTIC REGRESSION IN DIAGNOSTIC RESEARCH: A CLINICAL EXAMPLE

Binheurvel C\(^1\), Siciliano I\(^1\), Grobbee D\(^1\) and Moons K\(^1\)
\(^1\)University Medical Center Utrecht, Utrecht, Netherlands; ‘KiQ Ltd.’, Amsterdam, Netherlands

Aim: Genetic programming is a search method that can be used to solve complex associations between large numbers of variables. It has been used for e.g. myoelectrical signal recognition, but its value for medical prediction as in diagnostic and prognostic settings, has not been documented. We compared genetic programming and the commonly used logistic regression technique in the development of a prediction model using empirical data from a study on diagnosis of pulmonary embolism.

Methods: Using part (67%) of the data, we developed and internally validated (using bootstrapping techniques) a diagnostic prediction model by genetic programming and by logistic regression, and compared both on their predictive ability in the remaining data (validation set).

Results: In the validation set, the area under the ROC curve of the genetic programming model was significantly larger (0.73 vs. 0.64-0.62) than that of the logistic regression model (0.68 vs. 0.59-0.77). The calibration of both models was similar, indicating similar amount of overestimation.

Conclusions: Although the interpretation of a genetic programming model is less intuitive and this is the first empirical study quantifying its value for medical prediction, genetic programming seems a promising technique to develop prediction rules for diagnostic and prognostic purposes.

POSTER SESSION - UTILITY THEORY; HEALTH ECONOMICS; PATIENT & PHYSICIAN PREFERENCES; SIMULATION; TECHNOLOGY ASSESSMENT

MEASURING CHANGE IN QUALITY OF LIFE USING PROSPECTIVE VERSUS RETROSPECTIVE MEASURES

Nievoorken F, Tollenaar M, Oort F and Sprangers M
Academic Medical Center / University of Amsterdam, Amsterdam, Netherlands

Purpose: The usual method to measure change in quality of life (QoL) is to compare baseline- and follow-up measures. However, patients may change the reference value by which they evaluate their QoL over time. This may invalidate longitudinal QoL measurements. A method that aims to measure this change in reference values is the comparison of a retrospective baseline-with a conventional baseline-measurement. Another method for evaluating change in QoL is the retrospective appraisal of the perceived direction and magnitude of change, i.e., transition questions. Our objective was to measure change in QoL according to conventional baseline- and follow-up measurements, retrospective baselines- and follow-up measurements, and retrospective transition questions, and to determine which measure of change in QoL yields strongest associations with clinical measures of change.

Methods: HIV infected patients (n = 268) completed four MOS-HIV scales at week 0 (baseline measurement) and after 36 weeks (follow-up measurement). At week 36, they additionally completed these scales as retrospective baseline measurement, and they answered four questions about change in QoL, (transition questions). Change from week 0 to week 36 in CD4+cell count, body mass index, and plasma HIV viral load served as external criterion measures of change. We compared Pearson correlation coefficients between the three measures of change in QoL and the three clinical criterion measures using an overall Chi-square test.

Results: Improvement in QoL, according to the retrospective baseline-follow-up measurement was significantly larger than according to conventional baseline-follow-up measurement, as patients evaluated their QoL to be worse on the retrospective, than on the prospective baseline measurement. This may indicate that patients’ reference by which they evaluate their QoL over time had changed, which could invalidate prospective assessments of change. The method for measuring change incorporating a retrospective baseline measurement yielded strongest associations with clinical measures of change. As the other retrospective method, i.e., transition questions, yielded associations similar to the prospective method, this was not likely an artefact of the retrospective method in general.

Conclusion: A method for measuring change in QoL that incorporated a retrospective baseline measurement showed strongest associations with change in clinical indicators of health status, suggesting a more valid measurement of change in QoL than a conventional prospective method.
POSTER SESSION - UTILITY THEORY; HEALTH ECONOMICS; PATIENT & PHYSICIAN PREFERENCES; SIMULATION; TECHNOLOGY ASSESSMENT

HEALTH AND MATH DON’T MIX: DIFFICULTY INTERPRETING NUMERICAL INFORMATION IN HEALTH CONTEXTS

University of Michigan, Ann Arbor, MI

Purpose: Many clinical decisions require patients to weigh the risks and benefits of available treatments. In order to facilitate patients’ ability to make good decisions about their health care, it is critical that clinicians communicate these risks and benefits so that patients can decide which alternative best fits their preferences. However, communicating detailed risk information can be challenging, because many people have difficulty understanding concepts like percentages and frequencies, a phenomenon known as innumeracy. Moreover, health care risks may be harder to think about and interpret than other types of probabilistic information. We conducted an experiment testing whether people’s numeracy is influenced by the context of the numerical quiz.

Methods: This study reports data from a random subset of the Health and Retirement Study which surveys more than 22,000 Americans over the age of 50 every two years. 1019 participants answered four numerical questions. We randomized subjects to one of three contexts: 1) numeric context: e.g. “How many people out of 1000 would be expected to get the disease?”, or 2) retail context: e.g. “If you were to buy a television that was discounted 15%, how much money would you save if the original cost was $1000?”

Results: Numeracy was lower in health contexts than in the other two contexts, with 10%-20% fewer people being able to answer the health questions correctly. The mean proportion of people answering the health context questions correctly ranged from 23%-71%, compared to 36%-87% and 45%-62% in the numeric and retail contexts. Accuracy was lower in respondents over 65, women, non-Caucasians, and those with a high school education or less (all p’s<.0001). People who rated their health as fair or poor had significantly poorer numeracy than did those who rate their health as good or very good and this relationship remains after controlling for age, education, gender, and race (p<.001).

Conclusions: Many people have difficulty understanding risk information in health contexts, especially people in poor health. This is a potentially large barrier to successful risk communication. Future research should test whether comprehension is improved when numerical information is explained using more familiar contexts.

POSTER SESSION - UTILITY THEORY; HEALTH ECONOMICS; PATIENT & PHYSICIAN PREFERENCES; SIMULATION; TECHNOLOGY ASSESSMENT

LABOR MARKET EFFECTS OF INSULIN DEPENDENT AND NON-INSULIN DEPENDENT DIABETES AMONG CANADIAN LABOR FORCE

Farahbakhsh F
McMaster university, hamilton, ON, Canada

Objective: Many labor market analyses have established significant negative impact of diabetes in labor market behavior. This study investigates the impact of diabetes on the probability of working and weekly working hours for Canadian with insulin dependent diabetes (IDDM) and non-insulin dependent diabetes (NIDDM) and diabetes related comorbidities/complications, compared with healthy groups.

Method: Two-part model were performed to predict the weekly working hours (after adjusting for other chronic disorders and socio-demographic) of diabetes among respondent to the Canadian Community Health Survey 2001. Logitistics regression and multiple OLS analyses were used to predict the probability of working and weekly working hours. Sampling weights were used to be representative of the Canadian population.

Results: Women and men with IDDM had lower probability of having job by 18% and 5%, respectively compared with healthy groups. These probabilities were even less for women and men with diabetes related comorbidities/complications by 13% and 5% compared with IDDM without diabetes related comorbidities/complications. The probability of working for women and men with NIDDM also were 7% and 2% less than women and men without this disorder. Similarly, these probabilities were even less for women and men with NIDDM related comorbidities/complications by 2% compared with those without diabetes related comorbidities/complications (all at p-values < 0.05). The predicted weekly working hours for women and men without diabetes were 29 and 41 hours, with IDDM were 18 and 30, with IDDM and comorbidities/complications were 13 and 23, with NIDDM were 21 and 35 hours and finally with NIDDM related comorbidities/complications were 17 and 30, respectively.

Conclusions: This is the first study to estimate the impact of diabetes on Labour market outcome among Canadian Labour force. The effect of diabetes and its related comorbidities on the probability of unemployment and predicted weekly work hours for both men and women in Canada is substantial. The results of this study have implications for cost-effectiveness of diabetes control and may facilitate studies of the health burden of diabetes for the prevention and treatment of diabetes and thus increase the labor productivity.

POSTER SESSION - UTILITY THEORY; HEALTH ECONOMICS; PATIENT & PHYSICIAN PREFERENCES; SIMULATION; TECHNOLOGY ASSESSMENT

IMPACT OF GENERIC ENTRY ON COST-EFFECTIVENESS ANALYSES

Shih V, Han S and Cantor S
The University of Texas S M. D. Anderson Cancer Center, Houston, TX; Rice University, Houston, TX

Purpose: The market share of generic drugs has grown substantially since the passage of the Waxman-Hatch Act, increasing from 19% in 1984 to 50% in 2001. At entry, generics typically are priced lower than brand name drugs; thus, incorporating the impact of generic entry introduces an additional source of uncertainty in economic models as the timing of entry and the level of generic pricing are subject to variation. Our study explored the impact of generic entry on cost-effectiveness analyses.

Methods: We developed a mathematical model that incorporated the possibility of generic entry. We constructed two examples using simulated data on costs and effectiveness. In the first example, we compared a new treatment with an existing treatment for a severe acute condition to examine the short-term effect of generic entry. In the second example, we compared a new drug with an old drug for a chronic condition by introducing variations in drug prices and the timing of generic entry to analyze the long-term effect of generic entry. To better assess the modeling uncertainty, we employed both a deterministic analysis and a Bayesian probabilistic approach and presented the results using cost-effectiveness acceptability curves.

Results: The first example showed that the incremental cost-effectiveness ratio (ICER) increased as the price of the generic drug decreased. Corresponding to a societal willingness-to-pay of $50,000 per quality-adjusted life year (QALY), the probability that the new treatment was cost-effective was approximately 0.50 in the base case (i.e., no generic entry) and decreased as the annual cost of generics reduced, ranging from 0.45 at a price of $18,000 to 0.48 at $10,000. In the second example, the probability of an ICER less than $50,000 per QALY was 0.70 in the base case, but changed to 0.10 if the entry took place in year 3, 0.44 in year 5, and 0.64 in year 8 when the monthly drug cost was $600. If the monthly cost became $700, the probability was 0.18, 0.50, and 0.68 if entry occurred in year 3, year 5, and 8, respectively. Conclusions: Failure to incorporate the impact of generic entry would underestimate the ICER, and thus, overstate the economic benefit of the new drug. Incorporating generic entry into pharmaco-economic models would yield more accurate projections of the ICER and enhance decision making.
and Conrads-Frank A
Massachusetts General Hospital, Harvard Medical School, Boston, MA

Purpose: Remaining life expectancy (RLE) and quality-adjusted life expectancy (QALE) are standard measures of life expectancy. Our evaluation in decision theory is more straightforward. We sought to compare the performance of Gompertz approximation (GPA) and Declining Exponential Approximation of Life Expectancy (DEALE), using life table method as gold standard for calculating RLE and QALE.

Methods: All analyses were performed for two different underlying models for disease-specific mortality rates (DSM). For the multiplicative model, DSM was modeled as age-independent mortality rate ratio, multiplied by the age-specific background mortality. For the additive model, DSM was modeled as age-independent mortality rate difference, added to the background mortality. Background mortality was estimated from statistical life table data. In our basecase analysis, we set DSM being 2 x background mortality at age 45. This represents the mortality rate difference in the additive model and translates to a relative mortality rate ratio of 3 in the multiplicative model. We used (1) the closed formulas by Pollard based on the Gompertz function and (2) the closed DEALE formulas to model and translates to a relative mortality rate ratio of 3 in the multiplicative model. We used (1) the closed formulas by Pollard based on the Gompertz function and (2) the closed DEALE formulas to calculate undiscounted and discounted RLE and QALE. Results were compared to actual survival data.

Conclusions: Both approach methods underestimated the actual undiscounted RLE for both the additive and multiplicative baseline. Baseline results for men: for the multiplicative model, GPA (bias -4%) performed better than DEALE (-49%), whereas for the additive model, DEALE (-66%) was superior to GPA (-28%). Results for women showed similar patterns regarding magnitude and direction of bias. The use of time-independent disease-specific utility decrements yielded similar pattern to RLE and QALE. When using DSM in sensitivity analysis, bias was positively correlated with DSM, but bias direction (sign) and ranking of both methods did not change. Similarly, changing discount rates did not alter the bias pattern.

Based on our simple model, the Gompertz function should be preferred for multiplicative and the DEALE approach for additive models. The magnitude of the bias depends strongly on model parameters.

POSTER SESSION - UTILITY THEORY; HEALTH ECONOMICS; PATIENT & PHYSICIAN PREFERENCES; SIMULATION; TECHNOLOGY ASSESSMENT
A SUMMARY MEASURE OF POPULATION HEALTH
Steenow S, Woodward RC, and Cutler D
Harvard Interfaculty Program for Health Systems Improvement, Cambridge, MA; Harvard University and the National Bureau of Economic Research, Cambridge, MA

POSTER SESSION - UTILITY THEORY; HEALTH ECONOMICS; PATIENT & PHYSICIAN PREFERENCES; SIMULATION; TECHNOLOGY ASSESSMENT
PREDICTING DERMATOLOGICAL UTILITIES: AN EXPLORATION OF CENSORED UTILITY DATA PREDICTION MODELS
Chen S, Sternberg M, Bayoumi A, Olikin I, Soon S, Cruz P, Sexton S, McCull C, and Goldstein M
Emory University, Atlanta, GA; Emory University, School of Public Health, Atlanta, GA; St. Michael's Hospital, University of Toronto, Toronto, ON, Canada; Stanford University, Palo Alto, CA; University of Texas at Southwestern, Dallas, TX; Stanford University, Stanford, CA

Difficulties in obtaining public utilities have prompted researchers to develop regression models to predict utilities from health status measures. However, these models are based on general health instruments and may not be relevant for dermatological health status. Moreover, regression models based on linear regression (LR) are only valid when certain assumptions are met. Violations of these assumptions can lead to biased prediction models. If the population under study rarely records utilities of 1 and 0 it may be reasonable to disregard the ceiling or floor effect and use OLS. However, we have demonstrated that for most dermatological health status, utility scores will cluster close to the ceiling of 1.

PURPOSE: We explored several different regression methods to identify the best method to predict dermatological utilities from a validated skin-specific health status measure, Skindex.

METHODS: We interviewed 254 consecutive patients from general dermatology clinics at Stanford Medical Center (Stanford, CA), Grady Hospital (Atlanta, GA), and Parkland Hospital (Dallas, TX). Subjects completed both Skindex and a time-tradeoff utility assessment. We randomly divided our data into derivation and validation sets. The derivation data were analyzed using LR, Tobit regression (TR), Least Absolute Deviations (LAD), Least Trimmed Squares (LTS), Least Median Squares (LMS), and Censored LAD (CLAD). Candidate predictor variables included both Skindex and demographic parameters. We chose the same predictor variables for all models. We used the validation dataset to determine the median absolute prediction error (MAPE) (difference between the predicted and actual utility) and interquartile range (IQR) for each model.

RESULTS: In our sample, 48.2% report a health utility of 1, with very little dispersion around the median of .9998 (IQR: 0.1 - 94.7%). The MAPE and IQR were as follows: LR: 0.042 (0.073); TR: 0.073 (0.1); LAD: 0.027 (0.083); LTS: 0.039 (0.096); LMS: 0.039 (0.096); CLAD: 0.053 (0.11).

CONCLUSIONS: Although the LR model is not appropriate to estimate utilities when censoring is present, alternative regression models are available to alleviate the problems of bias inherent in ignoring ceiling effects. We have found for our dermatology-based population that the LAD regression gave the least difference in predicted and actual utilities. CLAD did not improve on prediction ability. Other investigators creating prediction models in the face of censored utility data should explore these regression methods.
POSTER SESSION - UTILITY THEORY; HEALTH ECONOMICS; PATIENT & PHYSICIAN PREFERENCES; SIMULATION; TECHNOLOGY ASSESSMENT

HEALTH VALUES OF PATIENTS INFECTED WITH HIV, HEPATITIS C, OR BOTH: ARE TWO VIRUSES WORSE THAN ONE?

Mena J, Sherman K, Leonard A, Sherman S, Mandell K and Devat J
Cincinnati VA Medical Center and University of Cincinnati, Cincinnati, OH; University of Cincinnati, Cincinnati, OH

Purpose: To compare health values of patients coinfected with HIV/hepatitis C (HCV) with those of patients singly infected with HIV or HCV and to assess the relationship of clinical and non-health-related factors to health values.

Methods: We interviewed subjects with HIV and/or HCV from Cincinnati and Pittsburgh in 2003 and 2004. We assessed rating scale (RS), time trade-off (TTO), and standard gamble (SG) utilities using U-Maker (each scaled from 0-1.0), and we explored univariate (Pearson and Spearman correlations) and multivariable (linear regressions) associations of the health values with the Mental Component Summary (MCS, 0-100) and Physical Component Summary (PCS, 0-100) of the SF-12, the number of bothersome symptoms (BS, range 0-15) from the Justice scale, spirituality (SPIR, 0-92) from the FACT-SpiX as well as a number of demographic, clinical, and psychosocial characteristics collected by interview and chart review.

Results: Of the 203 subjects, 59 (29%) had HIV monoinfection, 69 (34%) had HCV monoinfection, and 75 (37%) were coinfected. The mean (SD) age for the cohort was 45.7 (8.3) years; 77% were male; 58% were white; and 43% had a history of injection drug use. Selected results (mean [SD]) by infection type are shown in the table. In multivariable models, RS was significantly associated with sexual orientation, PCS, MCS, BS and SPIR (R² = 0.62), TTO with BS and SPIR (R² = 0.24); and SG with infection type (HCV monoinfection better), PCS, and BS (R² = 0.25).

Conclusions: Health values of patients with HIV, HCV, or both appear to be driven more by symptoms, function, and spirituality than by infection type or number of infections. Our findings have implications for both Markov modeling of these disease states and for designing possible interventions to improve health values.

POSTER SESSION - UTILITY THEORY; HEALTH ECONOMICS; PATIENT & PHYSICIAN PREFERENCES; SIMULATION; TECHNOLOGY ASSESSMENT

THE GAP EFFECT: DISCONTINUITIES OF PREFERENCES AROUND DEAD

Stalmeier P, Busschbach J, Lammers L and Krabbe P
University MC Nijmegen, Nijmegen, Netherlands; Erasmus Medical Center, Rotterdam, Netherlands; Erasmus MC Rotterdam, Rotterdam, Netherlands

Background: The assessment of negative values for health states considered to be worse than death is a controversial issue. Objective: To investigate how gate health states are valued when they are close to death.

Methods: A secondary analysis of the EuroQol EQ-5D data of the Measurement of Valuation and Health (MVH) study was made. Visual Analog Scale (VAS) and Time Trade-Off (TTO) values for 43 hypothetical health states were included per study, 2 for the meta-analysis and 3 for the pooled analysis.

Results: Complete data were obtained in 2997 respondents. The differences between the ordered VAS scores decreased gradually as the health states deteriorated. In contrast, significant gaps around dead were found for the positive as well as the negative TTO scores.

Discussion: These results are interpreted in light of a descriptive QALY model. This model was expanded to include utilities worse than dead. The VAS task does not pick up that bad states become intolerable, i.e. worse than dead, when they last too long, but the TTO task does. The current QALY model seems to lack descriptive validity for states valued worse than dead.

POSTER SESSION - UTILITY THEORY; HEALTH ECONOMICS; PATIENT & PHYSICIAN PREFERENCES; SIMULATION; TECHNOLOGY ASSESSMENT

THE EFFECT OF AGE, RACE AND GENDER ON UTILITY VALUES FOR HYPOTHETICAL HEALTH STATES

Wittgenstein E, Divi N, Halpern E, Araki S, Prosser L and Weeks J
Massachusetts General Hospital, Boston, MA; Stanford University, Stanford, CA; Harvard Medical School and Harvard Pilgrim Health Care, Boston, MA; Harvard School of Public Health, Boston, MA

Purpose: Health states are assumed to have a unique quality of life value, expressed as a utility, consistent across people with the same level of experience with the state. This research examines this assumption by exploring the effect on utilities for hypothetical health states of the age, race and gender of the individuals from whom the values are elicited.

Methods: Two parallel analyses were conducted: (1) a meta-analysis of the published literature (1976-2002), and (2) a pooled analysis of 4 primary data sets. For each analysis, a linear model was built by creating a rank-ordered outcome variable for utility and a rank-ordered predictor variable for health state severity, which assured a linear relationship within which to measure the effect of age, race and gender. The meta-analysis was conducted at the study level, and the pooled analysis at the individual level. Models also controlled for method of utility elicitation (meta-analysis) and study (pooled analysis). Health states were ranked through a modified Delphi approach with healthy individuals and physicians. An arbitrary subset of health states were included per study, 2 for the meta-analysis and 3 for the pooled analysis.

Results: The meta-analysis yielded 328 studies, 9 of which reported community values for hypothetical health states of the age, race and gender of the individuals from whom the values are elicited.

Conclusions: Our findings show significant differences between ethnic/gender groups in the valuation of health, with AA reporting less difference between the mild OA state and the severe OA state than by both the VAS, TTO and SG methods. The difference between mild and severe states was smaller for H/HV than for the SG method. For the severe OA state the odds that AA had scores 0.80 relative to W was 2.22 using the TTO method. Preferences for the mild OA state were not different among ethnic groups. Using the SG method, the odds that the scores were 0.80 in the public cohort vs. the patient cohort were greater than 1 for severe OA and mild OA. The public gave the severe OA state a higher preference score than patients did using the SG method. Education and age had significant, independent effects on utility scores. Age increased the SG utility scores, and the difference between severe and mild health states was less for VAS for older individuals. Education ameliorated the effects of other variables on TTO and SG scores.

Community utilities are recommended for use in societal perspective cost-effectiveness analyses (CEA), yet the composition of the sample from which such values are elicited is rarely reported. Individuals’ race may affect their valuation of health states, which implies that race must be considered in selecting samples to value states. More research is needed on variations in utilities due to individuals’ characteristics, and the validity of community values used in CEA.
DIAGNOSTIC DECISIONS ARE LOADED: THE EFFECT OF WORKLOAD ON RACIAL AND GENDER STEREOTYPING
Muroff J, Jackson J, Mowbray C and Himle J
University of Michigan, Ann Arbor, MI

Purpose: Despite having formal diagnostic criteria, psychiatric diagnoses can still be subjective, permitting non-psychiatric considerations, such as biases and stereotypes, to impact diagnostic decisions. This is especially problematic, since these decisions significantly influence immediate treatment, and may have long-term consequences on the course of the problem or illness. This study investigated whether workload and busyness influence psychiatric diagnosis and disposition decision-making in an emergency room setting.

Methods: Patterns of decisionmaking were examined through retrospective review of records of 1300 psychiatric patients, treated by 75 clinicians in an urban Psychiatric Emergency Service in the Mid-west, along with interviews of a random sample of these clinicians. Patient records were randomly sampled, according to the clinician’s level of cognitive load, based on the average number of patients typically seen during that shift and the actual volume of patients seen by the particular clinician during that shift. Cognitive load is the magnitude of demands imposed within specific time constraints and other limitations that restrict the amount of available cognitive resources to be dedicated to some task.

Results: Multinomial logistic regression analyses revealed that beyond clinical symptoms, nonpsychiatric factors such as race, gender, and cognitive load influence PES diagnostic and disposition decisionmaking. For example, when clinicians are under heavy load, they are significantly less likely (p<0.04) to assign a bipolar diagnosis than a psychotic diagnosis. In addition, Black patients were significantly less likely (p<0.001) to receive a bipolar diagnosis than a psychotic diagnosis, when compared to White patients. Furthermore, male patients were significantly less likely (p<0.026) to receive a bipolar diagnosis than a psychotic diagnosis, when compared to female patients. Additional analyses revealed other clinician and patient characteristics that significantly influence PES decisionmaking.

Conclusions: When clinicians are loaded down and busy, bias and stereotypes are more likely to influence their decisions, affecting specific sub-groups more than others. This could lead to misdiagnosis and/or inappropriate clinical decisions, which could have negative long-term effects on patients. These findings support the need for further evaluation and examination of these factors in order to help guide healthcare policy aimed to protect specific subgroups of client populations, which is critical to the delivery of quality mental health care.