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

Med Decis Making 2005; 25; 1
DOI: 10.1177/0272989X04274093

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 A[1], Elwyn G[2], Stacey D[3], Barratt A[4], Barry M[5], Coulter A[6], Holmes-Rovner M[7], Llewellyn-Thomas H[8], Moumjid N[9], Thomson R[10], Whelan T[11] and Collaboration I[1]

[1]University of Ottawa, Ottawa, ON, Canada; [2]University of Wales Swansea, Swansea, United Kingdom; [3]Ottawa Health Research Institute, Ottawa, ON, Canada; [4]University of Sydney, Sydney, Australia; [5]Harvard University, Boston, MA; [6]Picker Institute Europe, Oxford, United Kingdom; [7]Michigan State University, East Lansing, MI; [8]Dartmouth Medical School, Hanover, NH; [9]GRESAC - UMR 5823 CNRS, Lyon, France; [10]University of Newcastle upon Tyne, Newcastle upon Tyne, United Kingdom; [11]Hamilton 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 subcommittee 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. Members of a Shared Decision Making list-serve (N = 170) were presented with the broad classes of quality criteria and asked to provide additional suggestions. The methods sub-committee established the final set of voting items after testing their clarity with potential voters. The stakeholder sub-committee identified potential voters (e.g., patients, practitioners, researchers, policy makers) to review the quality criteria summaries and vote on the importance and feasibility of each quality criterion using a modified Delphi process.

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; balancing the presentation of options; using plain language; 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 summaries 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 K[1], Owens D[2] and Paltiel A[1]

[1]Yale University School of Medicine, New Haven, CT; [2]VA 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 homosexual 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-naïve population, (3) exclusive male negotiation of condom use, and (4) contact rates of 0-3 partners/yr and per-partner infectivity rates of 0.054-0.228, 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

OPENING PLENARY SESSION

A CAUTIONARY NOTE ON DATA SOURCES FOR EVIDENCE-BASED CLINICAL DECISIONS: WARFARIN AND STROKE PREVENTION

Thomson R, Martin E and Ruth W

University of Newcastle upon Tyne, Newcastle upon Tyne, United Kingdom

Purpose: To demonstrate whether different methods of risk assessment have different implications for treatment decisions.

Methods: Stroke risk due to non-valvular atrial fibrillation (NVAF) can be reduced by warfarin or aspirin; warfarin being more effective, with a higher risk profile. Choice of therapy requires assessment of both risks and benefits. Guidelines and risk stratification schemes incorporate stroke risk assessment for individual patients. However, the derivation of risk varies across schemes; from community based cohorts to randomised controlled trials (RCTs). We compared the risks and guidance in widely-promulgated approaches using 193 patients from a population prevalence study of AF. Stroke risk was calculated using the Framingham equation, an AF-specific Framingham equation, the CHADS2 score (derived from a US Medicare dataset) and Scottish (SIGN) guidelines. We compared treatment guidance using SIGN and a decision analytical approach using the Framingham equation. In the latter we classified some patients as "risk too low to benefit from warfarin (RTL)," if the risk of haemorrhagic stroke caused by warfarin approximated to, or exceeded, risk reduction of thromboembolic stroke.

Results: The schemes produce markedly different risk estimates. Both Framingham equations produce significantly lower risks for men and women than SIGN. The CHADS2 score provides an intermediate level of risk, significantly lower than SIGN for men and women, significantly higher than Framingham for men. 24/148 (16%) NVAF patients without past history of stroke/TIA would be advised warfarin using the decision analysis tool; 62 (42%) using SIGN ($p < 0.01$). 73 (49%) would be classified as RTL using the Framingham equation, of which 54 (74%) would be advised to take warfarin using SIGN guidelines.

Discussion: Stroke risk in NVAF patients differs markedly using different schemes. Community-based cohorts give lower risk estimates than CHADS2; both are lower than risks derived from RCT control groups. Using community-derived risks would lead to far fewer patients treated with warfarin than guidance derived from RCT control groups. Using the latter may lead to many low risk patients being treated with high risk therapy. This emphasises the importance of deriving risk from good epidemiological studies and raises the debate about appropriate source of data for risk stratification of patients to support effective clinical decisions. RCTs are excellent for effectiveness data, but may be severely limited for risk assessment.

OPENING PLENARY SESSION

HIGHLIGHTING "ADDITIONAL RISK" YIELDS MORE CONSISTENT INTERPRETATIONS OF SIDE EFFECT RISK COMMUNICATIONS

Zikmund-Fisher B[1], Roberts T[2], Fagerlin A[2], Derry H[2] and Ubel P[2]

[1]VA Ann Arbor Healthcare System, Ann Arbor, MI; [2]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 14% risk of an MI). This approach, however, forces people to do mental subtraction to tell how much risk the 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 in 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 denominator of all risk statistics ("out 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 took 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 most risk existed 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 additional risk specifically highlighted, yields the most consistent pattern of risk perceptions.

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 medical decision 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 Neuman-Morganstern utility functions that account for life duration, time-modulated goals, and extrinsic goals. We extend the preference assumptions underlying the QALY model, such as the zero condition and standard gamble independence (Miyamoto 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 accounted 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 Neuman-Morganstern 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

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

Stout N, Rosenberg M, Trentham-Dietz A, Smith M, Robinson S and Fryback 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 60 hypothetical screening scenarios implemented for 1990-2000 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 the frequency of mammograms. Screening as it actually occurred during this time 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 1990-2000 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. Scenarios consisting of fixed screening patterns dominated the actual screening patterns. Screening women aged 55-75 every three years accrued a similar number of QALYs at a cost of approximately $600 million less than the actual screening patterns. Alternatively, screening women age 50-75 every two years accrued 8,000 more QALYs at a similar cost compared with the actual screening patterns.

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 1990. 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 CONCURRRENT SESSION A - PUBLIC HEALTH 1

MODELING COHORT-SPECIFIC TRENDS IN BODY MASS INDEX FOR THE US POPULATION

Wang Y[1], Colditz G[2] and Kuntz K[1]

[1]Harvard School of Public Health, Boston, MA; [2]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.

Methods: We compiled 45,343 adult subjects from four National Health and Nutrition Examination Surveys (NHANES I, 1971-1974; NHANES II, 1976-1980; NHANES III 1988-1994; NHANES1999-2002) and adjusted for the sample structure. We fit multivariate generalized estimating equations (GEE) regression models separately for four race and sex groups. Based on the best-fitted models, we calculated cohort-specific average changes in BMI over the period of 1970-2010.

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, 30% 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 CISNET model estimates that approximately 36,000 (1%) colorectal cancer cases and 9,000 (0.6%) 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 CONCURRRENT SESSION A - PUBLIC HEALTH 1

POPULATION-BASED ASSESSMENT OF FUTURE BURDEN OF DISEASE AND COSTS RELATED TO CHRONIC HEPATITIS C AND ASSOCIATED DISEASES IN GERMANY

Siebert U[1], Sroczynski G[1], Aidelsburger P[2], Esteban E[3], Wasem J[2], Rossol S[4] and Wong J[5]

[1]Massachusetts General Hospital, Harvard Medical School, Boston, MA; [2]University of Duisburg-Essen, Essen, Germany; [3]University of Munich, Munich, Germany; [4]University of Heidelberg, Rüsselsheim, Germany; [5]Tufts-New England Medical Center, Boston, MA

Purpose: Antiviral treatment (AVT) has been shown to be efficacious and cost-effective in clinical trial-based populations with chronic hepatitis C (CHC). However, most of these decision-analyses ignore co-morbid illness associated with CHC, leading to overly optimistic results. The objectives of this study were (1) to predict population-based clinical and economic burden of CHC and associated diseases for the next 20 years in Germany and (2) to examine the potential impact of antiviral treatments (AVT).

Methods: We linked the German Hepatitis C Model, a validated and published Markov model of CHC, to German CHC prevalence and incidence data to project the 20-year and lifetime morbidity, mortality and costs for all treatable patients with known CHC and elevated transaminases in Germany. The model considered HCV genotype and CHC-related diseases (e.g., HIV co-infection, hemophilia, extra-hepatic manifestations) and evaluated the following policies: (1) no AVT, (2) interferon monotherapy, (3) interferon plus ribavirin, and (4) pegylated interferon plus ribavirin. We used pooled treatment efficacy data from meta-analyses of randomized clinical trials, results from the German Hepatitis C Quality of Life Study (n = 428), actual variable costs and reimbursement costs in the German health care system, and literature-based epidemiologic data on co-morbidities. For each policy, we calculated the incidence of clinical complications, CHC-related deaths, population life years (LY), quality-adjusted life years (QALY), costs, and incremental cost-effectiveness ratios (ICER). We adopted the societal perspective and used a 3% annual discount rate.

Results: In the absence of AVT during the next 20 years, HCV would cause more than 16,000 CHC-related deaths and 29,000 cases of liver cirrhosis leading to 1,200 liver transplantations. Peginterferon plus ribavirin would prevent about half of these complications and would add about 53,000 LY (or 49,000 QALYs) at a total cost of 1.3 billion EUR (undiscounted values). In the discounted lifetime analysis, peginterferon plus ribavirin was the most effective strategy with an ICER of 23,000 EUR/QALY compared with interferon mono-therapy (next non-dominated strategy). Peginterferon plus ribavirin remained cost-effective in sensitivity analyses across a broad rage of parameters including mortality, quality of life and costs of co-morbidities.

Conclusion: Incorporating co-morbid illnesses associated with CHC increased the ICER of peginterferon plus ribavirin, but treatment remained cost-effective when compared to other well-accepted medical technologies and would halve the burden of disease.

ORAL CONCURRENT SESSION B - HEALTH SERVICES RESEARCH

GENDER DISPARITIES IN PERCUTANEOUS CORONARY INTERVENTIONS IN PENNSYLVANIA

Hollenbeak C[1], Weisman C[1], Rossi M[2] and Ettinger S[1]

[1]Penn State College of Medicine, Hershey, PA; [2]Lehigh Valley Hospital, Allentown, PA

Purpose. The purpose of this research was to determine whether there were gender disparities in the use of percutaneous coronary interventions (PCI) in the treatment of acute myocardial infarction (AMI) in Pennsylvania and, if so, whether outcomes were affected.

Methods. Data were provided by the Pennsylvania Health Care Cost Containment Council (PHC4) and included all AMI patients at all acute care hospitals in the state of Pennsylvania during the year 2000. The population was stratified by gender to identify disparities in treatment and outcomes. Multivariate analyses were performed using binary logistic regression. Retrospective matching on propensity scores was performed using a "greedy" matching algorithm.

Results. During the year 2000, 10,170 (32.4%) AMI patients were treated with PCI and 21,181 (67.6%) patients were medically managed in Pennsylvania hospitals. Across all patients, women were significantly less likely to get PCI than men (23.9% vs. 40%, $p < 0.0001$) and were also more likely to die in the hospital (12.7% vs. 9.7%, $p < 0.0001$). Restricting the sample to patients treated at hospitals where PCI was available, women were significantly less likely than men to get receive the intervention (34.4% vs. 65.4%, $p < 0.0001$). They were also more likely to die than men when treated at hospitals offering PCI (10.1% vs. 7.3%, $p < 0.0001$). These disparities in treatment and outcomes were confirmed in multivariate analyses. After controlling for age, race, severity at admission, type of infarct, and source of admission, women still had a 25% lower odds than men of getting PCI ($p < 0.0001$) regardless of the availability of PCI. Finally, we used propensity score methods to match 3,022 women who received PCI to 3,022 women who did not. Results showed that women who received PCI were significantly less likely to die (2.4% vs. 10.7%, $p < 0.0001$).

Conclusions. In Pennsylvania, women appear to be less likely to be treated for AMI with PCI. Furthermore, PCI is strongly associated with better outcomes. Although there were factors we could not control, including location of infarct, time from symptom onset to treatment, and patient preferences, these results suggest that the morbidity and mortality associated with AMI in women could be reduced by increased used of PCI.

ORAL CONCURRENT SESSION B - HEALTH SERVICES RESEARCH

TRENDS IN THE UTILIZATION OF DIABETES-RELATED TESTS

Huang E, Basu A and Manning W

University of Chicago, Chicago, IL

Purpose: Since 1997 diabetes practice guidelines have called for lower risk factor targets, intensified medication regimens, and more frequent laboratory tests than in previous guidelines. We assess recent trends in test ordering for diabetes patients enrolled in a large managed care organization.

Methods: We utilized claims data for patients enrolled in health plans affiliated with UnitedHealthCare (UCHP) for the period 1996 to 2000. We identified non-pregnant adults with diabetes using the HEDIS and Hebert criteria. Separate cohorts of patients entering the health plan in each calendar year were followed for a period of one year. We studied trends in the 1) proportion of patients receiving a test in their first year, 2) the number of tests ordered in the first year, and 3) the time interval between ordered tests for glycosylated hemoglobin (HbA_{1C}), urine microalbumin, and serum cholesterol (both lipid panels or specific components of the cholesterol) tests. We compared trends for 31 participating UCHP care plans.

Results: We found a steady and progressive rise in the proportion of patients receiving HbA_{1C} or microalbumin tests over the years (Table 1). However, the overall proportion of patients receiving the microalbumin test remained low. A less prominent rise was observed for cholesterol testing.

	1996	1997	1998	1999	2000
N	11040	19680	28432	28122	38441
HbA_{1C} (%)	28	36	38	43	46
Urine Microalbumin (%)	2	3	4	6	8
Cholesterol test of any kind (%)	38	46	46	48	47

Among patients who received a test, the average annual frequency of HbA_{1C} (1.9), urine microalbumin (1.2), and cholesterol (1.8) testing did not change significantly. The median time between the first and second tests also did not change significantly. These findings were consistent across HMOs.

Conclusions: In this national sample of managed care patients, the proportion of diabetes patients receiving the recommended laboratory tests increased significantly, especially for glycosylated hemoglobin. However, only 50% of patients received a blood glucose or serum cholesterol test when all patients with diabetes should be receiving these tests on an annual basis. Diabetes care appears to be slowly intensifying as recommended in practice guidelines but significant care improvement is still necessary.

ORAL CONCURRRENT 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 emphasised population health improvement and reduction of health inequalities whilst also embracing the challenge 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: I. 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. II. 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: I. 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. II. 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. III. 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 incongruities 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 A - PUBLIC HEALTH 1

COST EFFECTIVENESS OF PARTIALLY EFFECTIVE HIV VACCINES IN THE ERA OF HIGHLY ACTIVE ANTIRETROVIRAL THERAPY

Owens D[1], Cavallaro J[1], Edwards D[2] and Shachter R[1]

[1]Stanford University, Stanford, CA; [2]Sandia National Laboratories, Livermore, CA

A vaccine remains the best hope for prevention of HIV worldwide. Difficulties in vaccine development have led many authorities to conclude that vaccines are likely to be only partially effective. We analyzed the costs and benefits of partially effective HIV vaccines in a population of men who have sex with men (MSM) in the era of highly active antiretroviral therapy (HAART).

We developed a dynamic transmission model to evaluate the total costs and benefits of preventive and therapeutic vaccines in a population of MSM in San Francisco, California (MSM population size 46,800), from the societal perspective. We used published and unpublished data on prevalence, incidence, and sexual risk behavior from San Francisco. We report findings for an early-stage epidemic in which prevalence was 10% and increasing. We used data on infectivity during the course of HIV disease, along with behavioral data to estimate transmission. We assumed that patients with symptomatic disease or AIDS would receive HAART. We used published data on costs of therapy, and evaluated vaccines of various duration and efficacy. Our base case preventive vaccine had 75% efficacy and lasted 10 years; the base case therapeutic vaccine added 5 years to length of life. We assumed a vaccine cost of $1000.

Over 20 years, the base-case preventive vaccine averted 5917 infections, added 6299 quality-adjusted life years (QALYs), and saved $123 million. As long as vaccine efficacy was greater than 10%, the vaccine cost less than $50,000 per QALY gained. Vaccines that had greater than 25% efficacy were cost saving. The effect of a therapeutic vaccine on the number of infections depends on the reduction in infectivity. A therapeutic vaccine that reduced infectivity by 50% averted 1,089 infections; if infectivity was reduced by 90%, almost twice as many infections were averted (1,965). Therapeutic vaccines were cost saving as long as they added at least 1 year to length of life, and risk behavior did not increase.

HIV vaccines of very modest efficacy provide enormous health benefit and are economically efficient. Vaccines with more than modest efficacy are cost saving over a wide variety of assumptions, even in the era of HAART. These findings imply that it is reasonable to initiate phase III efficacy trials for vaccines with modest expected efficacy.

ORAL CONCURRENT SESSION B - HEALTH SERVICES RESEARCH

THE IMPACT OF PROGNOSIS ON THE TREATMENT OF EARLY STAGE PROSTATE CANCER PATIENTS

Howard D

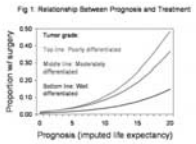
Emory University, Atlanta, GA

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: SEER-Medicare data were used for the study. The analysis has three steps. 1) A 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 = 44,880). 2) Parameter 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 = 32,798). 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 over-treatment.



ORAL CONCURRENT SESSION B - HEALTH SERVICES RESEARCH

POTENTIAL IMPACT OF PRESCRIPTION DRUG CO-PAYMENT INCREASE ON THE MEDICAID POPULATION

Bae S[1], Paltiel A[2], Fuhlbrigge A[3], Weiss S[3] and Kuntz K[4]

[1]Harvard University, Boston, MA; [2]Yale School of Medicine, New Haven, CT; [3] Brigham and womens' hospital, Harvard Medical School, Boston, MA; [4]Harvard School of Public Health, Boston, MA

Purpose: Last year, the State of Massachusetts increased the co-payment for prescription drugs by $1.50 for Medicaid beneficiaries. We sought to determine the likely health outcomes and cost shifts attributable to this co-payment increase using the example of inhaled corticosteroids (ICS) use among adult asthmatic Medicaid beneficiaries.

Methods: We compared the predicted costs and health outcomes projected over a 5-year time horizon with and without the increase in co-payment from four different perspectives (State government, hospitals, pharmacies and patients). We estimated that 10.8% patients on ICS therapy would not refill their prescriptions, based on the literature and inflation adjustment using the consumer price index. Predicted costs and health outcomes were based on a previously developed asthma model for evaluating ICS therapy.

Results: With an increased co-payment from 50¢ to $2, our analysis predicted that the State government would save on average $124 per asthmatic patient over 5 years, whereas hospitals would lose $18 per patient, pharmacies would lose $39 per patient, and patients would pay an additional $71 each over five years. We also projected an additional 87 urgent-care visits, 15 emergency room (ER) visits, and 7 hospitalizations over 5 years per 1,000 asthmatic patients with the co-payment increase. Projected government savings were a result of lower reimbursement to the pharmacies with decreased drug utilization and a decreased pharmacy reimbursement rate. Although we predicted that the State government would incur a cost increase associated with the increased number of acute exacerbations, this was more than offset by the savings from decreased pharmacy reimbursement. Financial losses attributed to hospitals were a result of the increased number of acute exacerbations, and the fact that the State government reimbursement rate is lower than the actual cost to hospitals. Pharmacies made less profit after the increase in co-payment because of the predicted decrease in the number of Medicaid patients who refill their prescription drugs.

Conclusion: In the example of ICS in asthmatic patients, an increase in prescription drug co-payment shifts the financial burden from the State government to the hospitals, pharmacies, and Medicaid patients. In addition, asthmatic patients will experience additional acute exacerbations that result in increased number of ER visits, urgent-care visits, and hospitalizations.

ORAL CONCURRENT SESSION A - PUBLIC HEALTH 1

THE COST EFFECTIVENESS OF ALTERNATIVE STRATEGIES FOR STOCKPILING AND DISPENSING MEDICAL AND PHARMACEUTICAL SUPPLIES FOR A RESPONSE TO ANTHRAX BIOTERRORISM

Bravata D[1], Zaric G[2], Holty J[1], Brandeau M[1], Wilhelm E[1], McDonald K[1] and Owens D[3]

[1]Stanford University, Stanford, CA; [2]University of Western Ontario, London, ON, Canada; [3]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, additional regional inventories would be available for dispensing 36 hours after attack detection, the local capacity to dispense prophylactic antibioics 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 annual probability of an attack increased from 0.0001 to 0.01.

Conclusions: Because of the rapid availability of regional inventories, the critical determinant of mortality following anthrax bioterrorism is local dispensing capacity. Bioterrorism preparedness efforts directed at improving local dispensing capacity may yield greater benefits than stockpiling and maintaining local inventories.

ORAL CONCURRENT SESSION A - PUBLIC HEALTH 1

ESTIMATING POPULATION EFFECTS FROM PREVENTION TRIALS: AN EXAMPLE USING THE PROSTATE CANCER PREVENTION TRIAL

Zeliadt S[1], Etzioni R[1], Penson D[2], Thompson I[3] and Ramsey S[1]

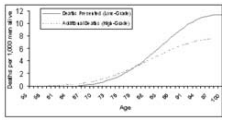
[1]Fred Hutchinson Cancer Research Center/ University of Washington, Seattle, WA; [2]University of Souther California, Los Angeles, CA; [3]Health Science Center at San Antonio University of Texas, San Antonio, TX

Purpose. This analysis estimates the lifetime cost-effectiveness of using finasteride to prevent prostate cancer based on the 7-year Prostate Cancer Prevention Trial (PCPT). Daily treatment with finasteride in this trial reduced prostate cancer prevalence by 25%; however, an increase in the number of high-grade tumors among the treatment group makes it challenging to translate the trial findings to community practice.

Methods. We use a Markov model to estimate the lifetime impact of finasteride on prostate cancer incidence and death, taking into account benign prostatic hyperplasia (BPH) and associated lifetime medical care costs. We translate the prevalence estimates from the trial into age-conditional probabilities of developing cancer (by grade). We conduct an extensive sensitivity analysis to evaluate the influence of multiple assumptions, including the impact of finasteride on high-grade tumors.

Results. The reduction in low-grade prostate cancer associated with finasteride does not translate into a survival benefit until many years after initiating treatment. Based on the prevalence estimates in the trial, we estimate that finasteride leads to an increase of 5.7 LYs and 46.2 QALYs per 1,000 men treated at an incremental cost of $1,700,000 per LY gained and $200,000 per QALY gained. Assuming finasteride does not increase the incidence of high-grade tumors, the incremental costs are $290,000 per LY gained and $130,000 per QALY gained. We estimate that annual medical care expenditures would increase by $2.2 billion for men age 55 to 64 and $1.2 billion for men age 65 and older, even after accounting for savings from the reduction in prostate cancer and BPH. These estimates assume 50% of eligible men over age 55 in the U.S. would begin daily preventive use of finasteride.

Conclusions. To achieve an incremental cost below $100,000 per QALY gained, finasteride must be shown to prevent high-grade as well as low-grade disease, and the price of the drug must be reduced by at least 50% from its current average wholesale price.



ORAL CONCURRENT SESSION B - HEALTH SERVICES RESEARCH

HOSPITAL VOLUME, OUTCOMES, AND COSTS OF STROKE CARE IN A REGIONAL MARKET

Votruba M and Cebul R

Case Western Reserve University, Hunting Valley, OH

Purpose: Few studies have investigated the volume-outcome relationship for non-surgical care, and none have examined the implications of seemingly justified volume-based referral policies in a local market. We estimate the effect of hospital volume on risk-adjusted 30-day mortality and in-hospital costs for patients with acute stroke in northeast Ohio (NEO).

Methods: This is a retrospective cohort of 13,096 patients admitted for acute stroke to one of 29 hospitals in NEO between 1991-97. Hospital volume was measured in 3 ways: 1) average annual number of stroke cases; 2) number of stroke cases in the 365 days preceding a given patient's admission; and 3) average daily census, including non-stroke patients. Risk adjustment included stroke severity (c-statistic 0.865), hospital teaching status, and patient socioeconomic markers. In-hospital costs were estimated from patient charges and hospital-specific cost-to-charge ratios; we also examined discharges to extended care and readmissions within 30 days.

Results: Median (interquartile range) annual stroke volume and costs were 148 (89-206) and $5214 ($3638-8010, 1997 dollars), respectively; 30-day mortality was 15.0%. Increasing volume was associated with decreasing risk-adjusted 30-day mortality (per 100-patient increase in volume, adjusted OR 0.897; 95% CI: 0.827,0.974); however, the two smallest volume hospitals had lower than predicted mortality, and the only plausible volume threshold would have excluded 26 of 29 hospitals. Interestingly, average daily census had at least as strong an effect size and pseudo-R2 as condition-specific volume measures. OLS regressions on log costs showed no significant effect of volume ($p = 0.338$), nor were there significant volume effects on discharge disposition ($p = 0.509$) or 30-day readmission rates ($p = 0.388$).

Discussion: Despite a significant inverse relationship of volume to stroke mortality and a favorable overall cost-benefit relationship, the volume-outcome relationship is imprecise and there is no threshold that clearly discriminates hospital performance. Further, the equivalence of the alternative volume measures in predicting outcome imply quite different quality improvement strategies. In this regional analysis, volume is an inadequate proxy for quality and a poor guide to selective referral.

ORAL CONCURRENT SESSION B - HEALTH SERVICES RESEARCH

A CLINICAL CENTER FOR SHARED DECISION MAKING: THE FIRST FIVE YEARS

Llewellyn-Thomas H[1], Clay K[2] and Mimnaugh D[1]

[1]Dartmouth Medical School, Hanover, NH; [2]Dartmouth Hitchcock Medical Center, Lebanon, NH

Purpose: The Center for Shared Decision Making (CSDM) at Dartmouth-Hitchcock Medical Center (DHMC) is the world's first hospital-based clinic designed to provide patients facing close-call, preference-sensitive health care choices with an individualized decision support service.

Methods: 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 + condition-specific decision aids (DAs; e.g. PSA screening, elective back surgery, breast or prostate cancer, etc.). These Level 3 patients complete evaluative questionnaires about this decision support service.

Results: To date, over 1500 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 preference (21%), compared with before (30%). Comprehension: Positive evaluations range from 84% (re. options' risks) to 92% (re. overall comparisons of the options). Values: Positive evaluations range from 74% (re. options' risks) to 78% (re. options' benefits). Making a Choice: 86-87% report enough support/free of social pressure; 68% 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 (90%); consider pros and cons (92%); identify questions to ask (90%); 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 BRIDGE FURTHER? EVIDENCE OF LIMITED EXTERNAL VALIDITY OF MODELS MAPPING SF-36 TO HEALTH UTILITY

Yu A, Yu 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 6923 Southern California Kaiser Permanente members who filled both SF-36 and HUI2 in year 1994-1995. Missing item responses were imputed by MCMC and propensity score method with a missing at random assumption. The linear regression framework mapping from SF-36 subscales to HUI2 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 characteristic 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 ($F = 1.57$, $df = 290$, $p < 0.0001$).

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 A - PATIENT AND PHYSICIAN BEHAVIOR/ PREFERENCES 1

PATIENT PREFERENCES REGARDING THE TIMING OF ANTIRETROVIRAL THERAPY INITIATION FOR CHRONIC HIV INFECTION

Locadia M, Van Grieken R, De Vries H, Sprangers M and Nieuwkerk P

Academic Medical Center / University of Amsterdam, Amsterdam, Netherlands

Purpose: The optimal timing of highly active antiretroviral therapy (HAART) initiation for chronic HIV infection is controversial. Earlier initiation, i.e. when CD4-cell counts are higher, reduces the risk of mortality but entails an earlier exposure to pill burden and potential toxicities as compared to later initiation. Therefore, the decision about when to initiate HAART involves a trade-off. We investigated preferences for earlier or later HAART initiation among HIV-negative patients, currently treated for an STD, who were considered to be at risk for acquiring HIV and therefore potential candidates for the decision about when to initiate HAART.

Methods: During an interview, patients were offered the choice between immediate HAART initiation versus delaying initiation for one year, when imagining having an asymptomatic HIV infection. Both options were presented as vignettes, including descriptions of pill burden, possible toxicities, and consequences for daily living. A difference in the three-year mortality risk between both options was also presented, and systematically varied between 0% and 10% to determine the threshold at which preference would switch to therapy initiation. We assumed a difference in risks between both options of approximately 1% for a CD4-cell count of 350, and 3% for a CD4-cell count of 200. Most clinical guidelines would recommend delay of HAART at 350 CD4-cells, and HAART initiation at 200 CD4-cells.

Results: Forty patients were interviewed (mean age 35 years, 58% males). One patient was unable to make a trade-off. Eleven patients (28%) would prefer HAART initiation even if mortality risks would be equal for both options. Most of these patients wanted to do anything they could against HIV, and would therefore initiate HAART. Eighteen patients (46%) switched their preference to HAART initiation when delay of HAART would result in a higher mortality risk. Ten patients (25%) always opted for delay of HAART even if this would result in a 10% greater mortality risk.

Conclusions: Our results show a large variation in patient preferences regarding the moment of HAART initiation. Some patients were more inclined to initiate HAART than would be recommended by current treatment guidelines, while others were more conservative than current guidelines. These findings emphasize the need for shared decision-making when deciding on the most optimal timing of HAART initiation among patients with a chronic asymptomatic HIV infection.

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 A[1], Dietz R[2], Osterziel K[2] and Siebert U[3]

[1] Charitè Campus Virchow Klinikum, Humboldt Medical School, Berlin, Germany; [2]Charitè Campus Virchow Klinikum, Humboldt Medical School, Berlin, Germany; [3]Massachusetts 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 40% 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 life long medical costs for DMPs.

Methods: Design: Cost and cost-effectiveness analysis using a 6 state Markov Model representing the number of prior hospitalizations (h=1 to h=4+) and death. Data sources: Pooled efficacy data from our meta-analyses of randomised clinical trials; SOLVD registry data for age-dependent hospitalizations and mortality rates adjusted for additional benefit from beta-blocker therapy and reimbursement 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, 33% on beta-blocker), our model yielded, on average, a remaining life expectancy of 3.24 years for conventional therapy 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. conventional 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 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 regarding 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 S[1], Anstrom K[1], Ludmer J[1], Glendening G[2] and Schulman K[1]

[1]Duke Clinical Research Institute, Durham, NC; [2]Novartis Pharmaceuticals Corporation, East Hanover, NJ

Purpose: The International Randomized Study of Interferon-α vs. STI571 (imatinib) (IRIS) demonstrated 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 hazards relationships estimated between age- and gender-matched individuals from the general population 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 15.30 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 $424,600 for imatinib-treated patients and $182,800 for IFN-α + LDAC-treated patients. With an annual discount rate of 3%, the incremental gain in survival among patients treated with imatinib was 3.93 LYS and 3.89 QALYs. The incremental cost-effectiveness ratios (ICERs) were estimated at $43,100/LYS (95%CI: 37,600 to 51,100) and $43,300/QALY (95%CI: 38,300 to 49,100). 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 survival estimation.

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

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

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

Araki S[1], Llewellyn-Thomas H[2], Tosteson A[3], Weinstein M[4], Hornstein M[5], Laufer M[5] and Kuntz K[4]

[1]Stanford University, Stanford, CA; [2]Dartmouth Medical School, Hanover, NH; [3]Dartmouth Medical School, Lebanon, NH; [4]Harvard School of Public Health, Boston, MA; [5]Brigham 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 who seeks relief from severe pelvic pain after first-line therapies have failed. Respondents rank-ordered eight treatments from the most to the least preferred and provided a rating on how likely they were to recommend each alternative (on a 5-point scale). We summarized the rank orderings and calculated the percentage of respondents who provided relatively high recommendation ratings ("probably" or "definitely" recommend) for each treatment. Logistic regression analyses (of the odds of a high recommendation rating) were 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 83.2% [65.5%]; laparoscopy 43.4% [23.0%]; progestin 34.9% [5.4%]; "other" 30.6% [9.1%]; alternative therapies 16.3% [4.0%]; danazol 12.3% [1.9%]; pain treatment center 11.9% [1.9%]; hysterectomy 2.1% [0.7%]. Specialists were significantly more likely than generalists to recommend laparoscopy (OR = 2.0; p = 0.01), alternative therapies (OR = 2.3; p = 0.03), and "other" treatments (12.2%; p = 0.03). Female physicians were more likely to recommend GnRH agonists (OR = 2.0; p = 0.08), progestin (OR = 1.7; p = 0.05), alternative therapies (OR = 2.0; p = 0.05), and "other" treatments (OR = 3.3; p = 0.08). Finally, younger physicians were significantly more likely to recommend laparoscopy (OR = 1.9; p = 0.04), and less likely to recommend danazol (OR = 0.3; p = 0.08) and alternative therapies (OR = 0.4; p = 0.09). There were no statistically significant associations between treatment recommendation 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 progestin to 0.76 for "other" treatment.

Conclusions: We found substantial variations in physicians' treatment recommendations for endometriosis pain. Furthermore, there was evidence of significant physician-related effects on recommendations, suggesting the possibility of unwarranted sources of variation.

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

LUNG TRANSPLANT CANDIDATES' ESTIMATION OF POST-TRANSPLANT UTILITIES

Singer L[1] and Gordon C[2]

[1]University of Toronto, Toronto, ON, Canada; [2]University Health Network, Toronto, ON, Canada

Purpose: To determine how accurately lung transplant candidates envisage transplant recipients' 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 51, 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. Recipients were separated into those with BOS (n = 23) and those without BOS (n = 76) 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 = 85). 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 0.37 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% ready, 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.

ORAL CONCURRENT SESSION B - CLINICAL EFFECTIVENESS AND QUALITY OF LIFE

ALCOHOL SCREENING AND INTERVENTION IN PRIMARY CARE EXTENDS QUALITY-ADJUSTED LIFE AND SAVES MONEY

Kraemer K[1], Roberts M[1], Freedner N[2], Palfai T[3] and Saitz R[2]

[1]University of Pittsburgh, Pittsburgh, PA; [2]Boston Medical Center, Boston, MA; [3]Boston University, Boston, MA

Purpose: Randomized controlled trials show that brief counseling can reduce hazardous alcohol use in primary care settings. However, the value and feasibility of screening all primary care patients for alcohol use and delivering these interventions have been questioned. The purpose of this study was to estimate the cost-effectiveness of alcohol screening and intervention in primary care.

Methods: We designed a Markov decision model to compare a strategy of alcohol screening and intervention to a strategy of no screening. The model tracked 6 alcohol-related health states (abstinence, safe drinking, at-risk drinking, alcohol abuse, alcohol dependence, and alcohol dependence in recovery). Model parameters were obtained from published values for: alcohol screening sensitivity/specificity, prevalence of alcohol problems in primary care, efficacy of brief intervention, transition between alcohol-related health states, mortality, costs for alcohol screening and intervention, and lifetime health care costs. Simplifying assumptions were made in cases where published data were not available. Standard gamble utility estimates for each alcohol-related health state were obtained from a clinic/community sample. We used separate models for men and women because transition probabilities between health states were substantially different by gender. We calculated the incremental cost-effectiveness ratio (cost per quality-adjusted life-year (QALY)) from the societal perspective and discounted costs and benefits at a rate of 3%.

Results: Under baseline conditions, the screening and intervention strategy dominated and was cost saving compared to the no screening strategy. In men, screening and intervention resulted in lifetime costs of $110,700 and 17.10 QALYs, and no screening resulted in costs of $111,000 and 17.05 QALYs. In women, screening and intervention resulted in lifetime costs of $110,200 and 17.15 QALYs, and no screening resulted in costs of $110,500 and 17.10 QALYs. Therefore, screening and intervention yielded a savings of $300 and a gain of 0.05 QALYs per man or woman screened. Results were robust to a range of alcohol use prevalence, intervention efficacy estimates, costs, utilities, and discount rates.

Conclusions: Screening and intervention for hazardous alcohol use in primary care extends quality-adjusted life and saves money. These results run counter to recent literature raising concerns about the value of alcohol screening and intervention in primary care and have the potential to inform clinical practice guidelines and policy decisions regarding the prevention of hazardous alcohol use.

ORAL CONCURRENT SESSION B - CLINICAL EFFECTIVENESS AND QUALITY OF LIFE

COMPETING DIAGNOSTIC TESTS FOR CORONARY ARTERY DISEASE: WHICH PARAMETER UNCERTAINTIES MATTER?

Groot Koerkamp B[1], Heijenbrok-Kal M[2], Kuntz K[1] and Hunink M[2]

[1]Harvard School of Public Health, Boston, MA; [2]Erasmus Medical Center, Rotterdam, Netherlands

Purpose: Various diagnostic imaging tests are available to diagnose coronary artery disease. The choice among these tests is difficult due to uncertainty about parameters, such as test characteristics and risk reductions of interventions. Our objective was to apply value of information (VOI) analysis to identify key parameter uncertainties—regarding the choice between imaging tests for coronary artery disease—for which future clinical 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.

Results: At a willingness-to-pay threshold of $50,000/QALY, CTA yielded the highest net health benefit, which was only 1 quality-adjusted life-day (QALD) greater than DSA—the next best strategy—and 40 QALD greater than not testing. The total EVPI was 2.5 QALD per patient. This implies that resolving all decision uncertainty is expected to improve net benefit with on average 2.5 QALD per patient. The population EVPI was $425 million. Uncertainty about the quality-of-life weights for varying severity of chest pain was responsible for 30% of the expected benefit. This is explained by the considerable uncertainty about these quality-of-life weights and their covariance structure, as well as their impact on false negative test results. Other important parameters were the prior probability of disease and the sensitivity and specificity of CTA. Resolving uncertainty about the probabilities of disease-related mortality and morbidity, other test characteristics, and health state transitions had a negligible expected benefit.

Conclusion: The total population EVPI is sufficiently large to justify gathering more evidence regarding the choice between these tests. Our results suggest that an observational study to obtain better estimates of quality-of-life weights, the prior probability of coronary artery disease, and test characteristics of CTA is the most useful next clinical study.

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

WOMEN'S DECISION MAKING ROLES REGARDING PROPHYLACTIC MASTECTOMY

Nekhlyudov L[1] and Geiger A[2]

[1]Harvard Medical School/Harvard Pilgrim Health Care, Boston, MA; [2]Southern California 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 future risk of breast cancer. Yet little is known about a woman's role in decision making regarding 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' mean age was 61 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, $p < 0.0001$) but were similar by race, age, breast cancer risk factors and concern 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 (83%, 88%, 72% respectively, $p=0.04$, adjusted for education). Decision making roles did not predict long-term satisfaction, 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 satisfaction 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/ PREFERENCES 1

UTILITY-BASED ASSESSMENTS OF QUALITY OF LIFE IN A RANDOMIZED TRIAL OF ANTIRETROVIRAL THERAPY IN ADVANCED HIV DISEASE

Joyce V[1], Araki S[2], Sundaram V[3], Hill A[1], Baker F[1], Munakata J[1], Kyriakides T[4], Sanders G[5], Griffin S[6], Holodniy M[3], Brown S[7], Nease RF[8], Sculpher M[6], Anis A[9], Yu W[1], Owens D[3] and Investigators t[4]

[1]VA Palo Alto Health Care System, Menlo Park, CA; [2]Stanford University, Stanford, CA; [3]VA Palo Alto Health Care System, Palo Alto, CA; [4]VA Connecticut Health Care System, West Haven, CT; [5]Duke, Durham, NC; [6]University of York, York, United Kingdom; [7]Bronx VA Medical Center, Bronx, NY; [8]Express Scripts, St. Louis, MO [9]University 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 Antiretrovirals) trial enrolls patients with plasma viral load (pVL) 2,500 copies/mL, a CD4 count 300/mm3, and who are currently on ART. We assessed US baseline quality of life assessments (n = 1231) for current health using time trade-off (TTO), standard gamble (SG), the Health Utilities Index 2 and 3 (HUI2, HUI3), the EuroQol (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 using 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 each 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 $r = 0.09$ (SG & HUI2) to $r = 0.87$ (HUI2 & HUI3). All were statistically significant (p 0.05) except for one pair (SG & HUI2). All mean utility scores between instruments were significantly different (p 0.05), except between SG & HUI2, SG & EQ-5D, and HUI2 & 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

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 unsettled 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 construct-irrelevant variables and utility responses and it finds that specific sub-types of risk are related with utility responses on the standard gamble and the time tradeoff instruments. This research highlights the need to study these sub-types of risk more closely.

ORAL CONCURRENT SESSION B - CLINICAL EFFECTIVENESS AND QUALITY OF LIFE

REFUSALS TO MAKE TRADE-OFFS: TASK CONTEXT AND EMOTION MATTER

Damschroder L[1], Zikmund-Fisher B[1], Roberts T[2] and Ubel P[2]

[1]VA Ann Arbor Healthcare System, Ann Arbor, MI; [2]University of Michigan, Ann Arbor, MI

Purpose: Person trade-off (PTO) elicitations 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 tradeoffs would increase and outrage about the task would be lessened by eliminating the need for people to make rationing choices. We presented half of our subjects with a traditional rationing task: "Imagine you can fund only one of two treatment programs...which one would you choose to fund?" and half with a non-rationing task: "Imagine two groups received treatment...which group received the greatest benefit?" We described three health conditions (mild, moderate, and severe shortness of breath), and subjects made choices between curing these conditions, two at a time, in three PTO elicitations within the assigned context.

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 ($p = 0.002$). 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 numeracy and those who thought the survey was relatively easy had higher odds for refusing to make tradeoffs (p's < 0.004).

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

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

Reyna V[1], Metz K[2], Poirier K[2], Roberts J[2], Adam M[2] and Lecroy C[2]

[1]University of Texas at Arlington, Arlington, TX; [2]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: 255 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 abstinent) and two scenarios from the following factorial design: 2(male vs. female had been sexually active) X 2(protected vs. unprotected prior sex) X 2(one prior partner vs. many prior partners). A variety of responses were elicited, including ratings of risks and benefits as "none," "low," "medium," or "high" and numerical scale values.

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 A - JUDGMENT AND DECISION MAKING

FORMER PATIENTS GIVE LOWER UTILITY RATINGS FOR COLOSTOMY THAN DO CURRENT PATIENTS: EVIDENCE FOR A THEORY DRIVEN RECALL BIAS

Sherriff R, 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. Colostomy 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 (TTO). The TTO 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. 95 subjects who still had colostomies 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 > .10). However, in the TTO 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 their colostomies reversed were willing to give up 44 months on average ($p < .001$). In addition the former patients, as compared to the current patients, reported that "having normal bowel function" was much more important to them ($p < .001$). Higher values on this item predicted lower utility values ($p < .001$).

Conclusion: 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."

ORAL CONCURRENT SESSION B - METHODOLOGICAL ADVANCES

LOCAL CONTROL VERSUS COVARIATE ADJUSTMENT IN ASSESSING TREATMENT EFFECTS IN CLINICAL TRIAL DATA

Johnston J[1], Obenchain R[1], Rousculp M[1], Hornung R[2] and Eckman M[2]

[1]Eli Lilly, Indianapolis, IN; [2]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: Despite randomization, small treatment group differences on important baseline predictors of mortality were noted. CA using logistic regression reduced the estimated survival advantage of DAA from 5.2% (unadjusted p=0.024) to 4.6% (p=0.038.) In contrast, the LC approach supported a range of DAA survival advantages from 4.4% (60 clusters) to 7.3% (500 clusters). Higher estimates (with lower uncertainty) emerged as cluster size decreased even though more patients were forced into clusters non-informative about local treatment differences (i.e., containing only treated or only untreated patients). Whereas CA requires pre-specification of presumed relationships between predictors and outcome, LC allows for arbitrary, even discontinuous, relationships between predictors and outcome. As such, LC is an attractive choice for exploratory analyses and for assessing the robustness of intent-to-treat analyses. Conversely, taking advantage of available clinical knowledge in forming clusters and interpreting results can be problematic. Additional challenges include uncertainty around the best approach to weighting across clusters and in determining an "optimal" number of clusters. Finally, the LC approach requires analysis and graphical display capability not currently available in many statistical packages.

Conclusions: LC can be used to explore treatment effects and may be useful in assessing the robustness of baseline efficacy estimates from randomized clinical trials.

ORAL CONCURRENT SESSION B - METHODOLOGICAL ADVANCES

THE VALUE OF IMPLEMENTATION AND THE VALUE OF INFORMATION: COMBINED AND UNEVEN DEVELOPMENT

Fenwick E, Claxton K and Sculpher M

University of York, York, United Kingdom

Purpose: To develop a framework to inform the separate but linked policy decisions regarding investment in healthcare services, further research and strategies to ensure appropriate implementation of health technologies given existing evidence.

Methods: In a budget constrained healthcare system the decision to invest in implementation strategies must be made alongside those regarding investment in healthcare services and further research. We present a framework that examines the value of further information and the value of implementation strategies separately but simultaneously. We provide a measure of the maximum return to further research (expected value of perfect information: EVPI) and an upper bound on the value of adopting implementation strategies (expected value of perfect implementation: EVPImp). This framework is demonstrated using a series of health care technologies selected from those previously considered by the UK National Institute for Clinical Excellence (NICE) including: Orlistat for obesity, Zanamivir for influenza and prophylactic extraction of wisdom teeth. The information used for the case studies was taken from the NICE guidance and assessment reports.

Results: In the case of wisdom teeth the value of further research is low (EVPI = £0) but the value of adopting appropriate implementation strategies is substantial (EVPImp = £20m). In other circumstances, investment is worthwhile in both further research and implementation strategies, e.g., in the case of Zanamivir further clinical trials are worthwhile (EVPI = £6m) as are strategies to restrict use to high risk groups presenting within 24 hours on symptom onset (EVPImp = £3.5m). Similarly in the case of Orlistat, there is considerable value of adopting implementation strategies that ensure that only patients who maintain the required weight loss continue to receive the drug (EVPImp = £11m).

Conclusions: Previous methods for valuing implementation strategies have confused the value of research and the value of implementation. This framework demonstrates that the value of information and the value of implementation can be examined separately but simultaneously in a single framework. This can usefully inform policy decisions about investment in healthcare services, further research and adopting implementation strategies.

ORAL CONCURRENT SESSION A - JUDGMENT AND DECISION MAKING

DIFFERENTIAL EFFECT OF A BREAST CANCER DECISION AID BY MARITAL STATUS

Hess R[1], Barnato A[1] and Peele P[2]

[1]University of Pittsburgh School of Medicine, Pittsburgh, PA; [2]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 = 160 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, $\alpha = 0.933$), subjective knowledge of treatment risks and outcomes (5 items, $\alpha = 0.793$), and decision uncertainty (5 items, $\alpha = 0.790$) at the time of the initial consultation and treatment decision, and for decision regret 3 months later (5 items, $\alpha = 0.830$). 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 61.8 years, 81% were white, 37% single (widowed, divorced, or separated), and 65% had low severity breast cancer. 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 A - JUDGMENT AND DECISION MAKING

A DECISION ANALYSIS BALANCING SURVIVAL AND CISPLATIN OTOTOXICITY IN CHILDREN WITH GERM CELL TUMORS

Li Y[1], Womer R[2] and Silber J[2]

[1]University of Pennsylvania School of Medicine, Philadelphia, PA; [2]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 relapse, 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 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%, $p < 0.01$) suggested high-dose treatment. Further analyses of the SG utility scores revealed that this difference was due to parents' assessment of cancer survival with sensorineural hearing loss as manageable and not worth the risk of death (average SG utility 0.96). In contrast, children 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 = 0.92, significantly different from parents', $p < 0.05$). The differences in model-recommended decisions were not affected by the respondents' numeracy and/or their propensity to exhibit omission bias.

Conclusions: 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 may help 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

QUALITY-ADJUSTED YEARS OF LIFE GAINED FROM IMPLANTABLE DEFIBRILLATOR: COMPARISON OF A NEW METHOD WITH TRADITIONAL ANALYSES IN CLINICAL TRIALS

Noyes K[1], Dick A[1], Zhao H[1], Corona E[1], Hall W[1], Zwanziger J[2] and Moss A[1]

[1]University of Rochester School of Medicine, Rochester, NY; [2]University of Illinois at Chicago School of Public Health, Chicago, IL

Purpose: Traditionally results 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, or the Kaplan-Meier survival approach that accounts for the weakness of the previous method but does not recognize differences in the HRQOL of survivors. Here we examine a novel approach that incorporates the strengths of the traditional methods, use it to examine HRQOL during the additional years of life saved, and compare our results with the results from traditional methods.

Methods: We used the original data from the MADIT II study that examined mortality in patients with implantable cardiac defibrillators (ICD) (n=1089). To estimate changes in HRQOL, we used a general health preference instrument, Health Utility Index 3 (HUI3). First, we developed a multivariate fixed-effects model to impute the missing HUI3 data assuming missing at random mechanism. Then, using the imputed data, we calculated the total HRQOL gain over 3 years as the area under the normalized HRQOL profile over time, for each treatment arm. We also estimated the difference in HRQOL gains between treatments, as the area between the two treatment profiles. We bootstrapped all standard errors. These results were compared to the traditional estimates of changes in HRQOL from baseline and to results of survival analysis.

Results: Over the 3 years, subjects in the ICD arm experienced a better survival (&Delta Life Years Saved (ICD-control) = .122 years, $p = .03$) and had a similar change in the HUI3 score from baseline (ICD: -.056 (SE .032), control: -.039 (SE .035), &Delta (ICD-control) = -.017, $p = .73$) conditional on survival. When adjusted for survival and censoring, use of ICD resulted in a loss of -0.257 (SE 0.050) QALYs, while subjects in the control arm lost on average -0.311 (SE 0.011) QALYs, &Delta HRQOL (ICD-control) = .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 B - METHODOLOGICAL ADVANCES

VARIABLE CORRELATION SUBSTANTIALLY INFLUENCES UNCERTAINTY AND VALUE OF INFORMATION ESTIMATION IN PROBABILISTIC SENSITIVITY ANALYSES

Groeneveld P[1], Glick H[2] and Schwartz J[2]

[1]Philadelphia VA Medical Center, Philadelphia, PA; [2]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 mortality 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 B was "correct" at λ = \$100,000/QALY was only 0.04, and the EVPI was \$2075 per patient. 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 patient.

Conclusion: The underlying correlation structure among costs and outcomes in cost-effectiveness models 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

Wigton R[1], Darr C[2], Leeman-Castillo B[3] and Gonzales R[4]

[1]Univ of Nebraska Medical Center College of Medicine, Omaha, NE; [2]University of Colorado at Denver, Westminster, CO; [3]University of Colorado at Denver, Denver, CO; [4]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) would be able to distinguish their own decision policy for prescribing antibiotics in RI from that of other PCPs.

Methods: Each of 81community 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 cue 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 cue 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 whic 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, a 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 A - JUDGMENT AND DECISION MAKING

PROSPECT THEORY IN THE VALUATION OF ARTHRITIS HEALTH STATES

Suarez-Almazor M[1], Souchek J[1], Byrne M[2], Kelly P[1], Richardson M[1], Pak C[1] and Nelson H[1]

[1]Baylor College of Medicine/Michael E. DeBakey VA Medical Center, Houston, TX; [2]University 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 391 subjects (198 patients with OA and 193 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-5D 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-5D York tariffs.

RESULTS: In the public, 27.2% had worse health status than the vignette, 9.9% had similar health, and 68% had better health. For the patient group, the percentages were 50.5%, 16.7%, and 32.8%, respectively. The difference score between the scenario and current health was related to the valuations of the scenario as follows (Spearman correlation: VAS r = -0.06 (p = 0.22); SG r = 0.35 (p < 0.001); TTO r = 0.15 (p = 0.004); and WTP r = -0.10 (p = 0.10). As postulated in PT, 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 loss (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.

CONCLUSION: Our findings suggest that subjects' valuation of arthritis states follow PT principles when using SG, 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

BAYESIAN META-ANALYSIS OF PAPANICOLAU SMEAR ACCURACY

Cong X[1], Cox D[1] and Cantor S[2]

[1]Rice University, Houston, TX; [2]The University of Texas M. D. Anderson Cancer Center, Houston, TX

Purpose: The purpose of this study is to reanalyze a published meta-analysis of Papanicolau (Pap) smears (Fahey, et al. Am J Epidemiol 1995; 141:680-689) 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 (*?*) and specificity (*?*) 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, $\delta = \text{logit}(\beta) - \text{logit}(1 - \alpha)$, where *?* 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.

		Bayesian Model 1*	Bayesian Model 2*	SROC Curve**
Follow-up	Sensitivity	0.61 (0.52-0.70)	0.70 (0.58-0.80)	0.66 (0.58-0.73)
	Specificity	0.63 (0.53-0.72)	0.73 (0.63-0.81)	0.66 (0.58-0.73)
Screening	Sensitivity	0.55 (0.44-0.66)	0.61 (0.46-0.75)	0.58 (0.49-0.67)
	Specificity	0.65 (0.54-0.75)	0.76 (0.66-0.84)	0.69 (0.62-0.77)

*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 B - METHODOLOGICAL ADVANCES

IMPACT OF HEART FAILURE ON 1-YEAR COSTS AFTER MYOCARDIAL INFARCTION RELATED HOSPITALIZATION: AN APPLICATION USING A FLEXIBLE LINK AND VARIANCE FUNCTION MODEL

Basu A[1], Arondekar B[2] and Rathouz P[1]

[1]University of Chicago, Chicago, IL; [2]University of Illinois at Chicago, Philadelphia, PA

Background: Increased incidence of myocardial infraction (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 expenditure 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 hospitalization, 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 OLS 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: 15,116 patients were eligible for the study. Completed 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 $14700 (1135). 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 2

A NATIONAL CATALOGUE OF PREFERENCE WEIGHTS FOR CHRONIC CONDITIONS

Sullivan P[1] and Lawrence W[2]

[1]Pharmaceutical Outcomes Research Program, University of Colorado School of Pharmacy, Denver, CO; [2]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 Medical Expenditure 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 A - PATIENT AND PHYSICIAN BEHAVIOR/ PREFERENCES 2

ATTITUDES OF PATIENTS WITH INCURABLE CANCER TOWARDS MEDICAL TREATMENT

Voogt E[1], van der Heide A[1], van Leeuwen A[1], Rietjens J[1], Visser A[2] and van der Maas P[1]

[1]Erasmuc MC, University Medical Center Rotterdam, Rotterdam, Netherlands; [2]Helen Dowling Institute, Utrecht, Netherlands

When cancer has advanced to a stage in which cure becomes very unlikely, patients may have to consider whether they appreciate further life-prolonging treatment or not. We assessed the attitudes of cancer patients towards such treatment.

Patients who suffered from breast, lung, colorectal, prostate or ovary cancer, which had recently progressed into an incurable stage, were interviewed and asked to fill out a written questionnaire about their attitudes concerning life-prolonging treatment and about end-of-life decision-making (Quality Quantity Questionnaire, Stiggelbout, 1996).

122 patients (mean age 64 yr. (sd = 10.5), 53% women) participated in the study. Patient attitudes concerning trade-offs between quality of life and length of life could be categorized into three different profiles: striving for quality of life (33%), striving for length of life (38%), and no clear preference (29%). Older patient and patients who were more tired, or had less positive feelings, and patients who had discussed their wishes concerning medical treatment with their health care professionals or family members, or had filled out an advance directive were more inclined to strive for quality of life. In contrast, patients with a history of cancer of less than six months were more inclined to prefer life-prolongation than patients with a longer history of cancer.

We conclude that patients with incurable cancer vary in their attitudes concerning the application of life-prolonging treatment. These attitudes are related to patient and disease characteristics and to discussing and making decisions about end-of-life care. Being aware of these differences may help physicians in their communication with patients about end-of-life care.

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

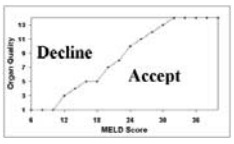
Alagoz O[1], Schaefer A[1], Maillart L[2] and Roberts M[1]

[1]University of Pittsburgh, Pittsburgh, PA; [2]Case 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 figure shows 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. Organs 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 B - METHODS IN JUDGMENT AND DECISION MAKING

VIEWERS OF SURVIVAL GRAPHS OFTEN IGNORE TIME AXIS LABELS, BIASING PERCEPTIONS OF TREATMENT EFFECTIVENESS

Zikmund-Fisher B1, Fagerlin A[2] and Ubel P[2]

[1]VA Ann Arbor Healthcare System, Ann Arbor, MI; [2]University of Michigan, Ann Arbor, MI

Purpose: Reports of randomized clinical trials often use survival curves to summarize clinical outcomes over time and graphically demonstrate evidence of treatment effectiveness. Survival curves can also be used in patient communications to display how health risks accumulate over time. However, for survival graphs to be an effective communication method, people must recognize how long a time period is shown and correct their interpretations accordingly. But, do people actually think about the time axis labels when they view these graphs?

Methods: In a randomized survey experiment, administered online, we tested whether people viewing survival curves appropriately adjust their risk perceptions according to the time period shown. Internet users (N = 864) were recruited from a demographically balanced U.S. panel. Participants read about a hypothetical disease and then viewed one of four survival graphs that displayed mortality risks with and without treatment. They were also informed that all risks were constant over time. Survival graphs showed either a visually large or visually small difference between treatments and were labeled to represent either 5 year or 15 year risk statistics. Participants then provided ratings of disease seriousness, as well as treatment effectiveness for each possible treatment.

Results: Variations in ratings corresponded more with the visual similarity of the graphics than with changes in the statistical risk exhibited, with participants perceiving greater disease seriousness and differences in treatment effectiveness in longer term (15 year) graphs. For example, when comparing two possible treatments, survey participants perceived a 1% difference in annual mortality risk to be much more significant when displayed over 15 years than when it was displayed for only 5 years ($p < 0.001$). Conversely, perceptions of the disease were statistically indistinguishable in a pair of visually similar graphs even when the 5 year graph showed an untreated mortality risk almost three times larger (e.g., 20.9% vs. 7.5%) than the 15 year graph ($p > 0.14$).

Conclusions: When people interpret survival curves, they do not pay adequate attention to the time frame represented in the curves and respond, instead, to the visual image presented in the graphic. This cognitive bias will impact the use of survival graphs as a patient communication tool and raises concerns about clinicians' abilities to accurately interpret published research results.

ORAL CONCURRENT SESSION A - PATIENT AND PHYSICIAN BEHAVIOR/ PREFERENCES 2

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

Kuppermann M[1], Gregorich S[1] and Nease R[2]

[1]University of California, San Francisco, San Francisco, CA; [2]Express Scripts, Inc., and Washington University St. Louis, 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 prior inclinations regarding prenatal testing in a randomized clinical trial.

Methods. 496 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 1-2 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. At follow-up, inclination to undergo diagnostic testing was lower among women who viewed PT Tool (24% versus 33%, $p < .01$), 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.

Use of invasive testing by baseline (pre-randomization) inclination, women < 35 years

Intervention	Low/moderate inclination	High inclination	All women
Control	29.2%	94.9%	66.4%
PT Tool	44.8%	84.3%	66.3%

Within the control group, baseline invasive testing inclinations were highly predictive of actual test behavior (r = .66), whereas among viewers of PT Tool this association was more moderate (r = .41). That is, women who viewed PT Tool were more likely to exhibit invasive testing utilization that was inconsistent with their baseline inclination, suggesting that PT Tool helped to change their minds (p = .03).

Conclusion. The significant interaction effect and lack of a main effect suggested that PT Tool is a useful and unbiased decision aid that helps women make informed choices about prenatal testing.

ORAL CONCURRENT SESSION A - PATIENT AND PHYSICIAN BEHAVIOR/ PREFERENCES 2

COMMUNICATING THE BENEFITS OF RISK REDUCING INTERVENTIONS - NUMBER NEEDED TO TREAT (NNT) OR POSTPONEMENT OF ADVERSE EVENTS? RANDOMIZED TRIAL OF LAY PEOPLES' OPINIONS

Halvorsen P[1], Kristiansen I[2] and Selmer R[3]

[1]University of Tromsø, Norway, 9516 ALTA, Norway; [2]University of Southern Denmark, Odense, N-0853 Oslo, Norway; [3]The Norwegian Institute of Public Health, N-0403 Oslo, Norway

Purpose: To explore whether people respond differently to NNT than to postponement of adverse events when considering risk reducing drug therapies.

Methods: A sample of attendees (n=1,754) to a general population health survey was mailed a questionnaire and asked about preferences for a hypothetical drug therapy to prevent heart attacks. Based on the 4S study, treatment effect after 5 years was presented in 3 formats: "For every heart attack that is avoided, 13 patients must take the therapy" (NNT) or "for all patients taking therapy, heart attack is postponed by 2 months" or "for one out of four patients taking therapy, heart attack is postponed by 8 months, while the others don't benefit." Respondents were randomly allocated to one of these formats. Another 1,000 attendees were asked about preferences for a drug therapy against hip fractures. Based on the FIT study, benefit from 5 years of drug therapy was presented either as "for every hip fracture that is avoided, 57 patients must take the therapy" (NNT) or "for all patients taking therapy, hip fracture is postponed by 16 days" or "for three out of 100 patients who take therapy, hip fracture is postponed by 16 months, while the others don't benefit." Again, allocation to one of the effect formats was random.

Results: The overall response rate was 81%. Presented with the heart attack scenario, 93% consented to the drug therapy in response to the NNT-format, 69% consented in response to "smaller postponement for everybody," while 82% consented in response to "greater postponement for a fraction" ($\chi 2 = 89.6$, $p < 0.001$). Corresponding figures for the hip fracture scenario were 74%, 34% and 56% respectively ($\chi 2 = 91.5$, $p < 0.001$). In multivariate logistic regression models, additional predictors for consent to drug therapy were poor education, desire to know one's risk status and perception of the risk reduction format as easy to understand. However, 40% of the respondents reported difficulties in understanding the effect format.

Conclusion: Treatment effects presented in terms of NNTs yield higher consent rates than postponement of adverse events. The results may reflect preferences for substantial but uncertain benefits over certain, but small benefits, or simply a response to prospects of "complete prevention," "great postponement" and "small postponement," disregarding probabilities. Finally, the results may reflect difficulties in understanding the effect measures.

ORAL CONCURRENT SESSION B - METHODS IN JUDGMENT AND DECISION MAKING

RANDOMISATION TURNS DIAGNOSTIC RESEARCH INTO INTERVENTION RESEARCH

Moons K, Biesheuvel C and Grobbee 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 that after establishing a test'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 randomisation 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 etiologic research. The nature of the diagnostic question 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 finally 3) if other (randomised) therapeutic studies have provided evidence on 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, Stiell 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 correctly.

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 largely 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 correct addresses. Overall response rate was 261/376, or 69.4%. 260 / 261 respondents reported being familiar with the OAR. Of those, most reported applying the rule consistently; 90% reported using the OAR always or most of the time in rule-appropriate situations. The majority reported not using the rule exclusively; 42% reported basing their decision primarily on the basis of the rule, while 42% reported considering a small number of other factors, and 15% said that the OAR was only one of many factors going into their decision. Only 31% of respondents correctly recalled all components of the OAR and excluded all foils.

Conclusions. 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

PATIENTS' VIEWS OF PRIVACY AND RESEARCH: DELIBERATIVE DEMOCRACY AT WORK

Damschroder L[1], Pritts J[2], Neblo M[3] and Hayward R[1]

[1]VA Ann Arbor Healthcare System, Ann Arbor, MI; [2]Georgetown University, Washington, DC; [3]The Ohio State University, Columbus, OH

Purpose: The HIPAA Privacy Rule went into effect April 2003, giving IRBs only vague guidelines about when researchers can 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 deliberated 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 40% who held these views related to pharmaceutical company researchers. Veterans who participated in the all-day deliberation became more willing ($p = 0.008$) 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 ($p=0.53$) at the time of the follow-up survey. 90% 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 be 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 A - PATIENT AND PHYSICIAN BEHAVIOR/ PREFERENCES 2

VALUE FOR THE FUTURE AND BREAST CANCER PREVENTIVE HEALTH BEHAVIOR

Gurmankin A[1], Micco E[2] and Armstrong K[2]

[1]Rutgers, the State University of New Jersey, New Brunswick, NJ; [2]University of Pennsylvania, Philadelphia, PA

Background: Many everyday health decisions involve intertemporal choice: trade-offs between immediate pleasure or convenience and a potentially larger, delayed health benefit. Time preference, or the extent to which people discount future benefits in favor of immediate benefits, might represent an important determinant of preventive health behavior. However, research to date on the relationship between time preference and health behavior has yielded mixed results.

Purpose: To examine the association between future time preference and utilization of genetic counseling for BRCA1/2 testing, annual mammography screening, and monthly self breast examination.

Methods: Because use of BRCA1/2 counseling is rare, a prospective health system based case-control study was used to assess the association between time preference and BRCA1/2 counseling. Because women who undergo BRCA1/2 counseling are not representative of the general population of women, a nested cross-sectional analysis of controls was used to assess the associations between time preference and mammography and self breast exam. Cases (n = 234) were adult women without breast or ovarian cancer who underwent BRCA1/2 counseling within the University of Pennsylvania health system. Controls (n = 566) were adult women without breast or ovarian cancer who saw a primary care physician in the same health system but did not seek genetic counseling. Subjects completed a questionnaire assessing their time preference (using the Revised Consideration of Future Consequences Scale, Cronbach's á = .71), annual mammography adherence, self breast exam frequency, family history, and sociodemographics.

Results: A stronger future time preference was seen among those with higher educational attainment ($p < .0001$) and income ($p < .0001$), and who were Caucasian ($p = .025$). The association between a future time preference and health behavior was strong and positive for utilization of genetic counseling (comparing cases and controls) (OR 2.6, 95% CI 1.8-3.6) and weaker but still significant for adherence to annual mammogram (OR 1.5, 95% CI 1.0-2.4), and absent for adherence to monthly self breast exam (OR 1.0, 95% CI 0.7-1.3). For all three behavioral outcomes, the odds ratios did not change when adjusting for degree of family history and sociodemographic characteristics.

Conclusions: Time preference is associated with health behavior, but the strength of the association varies with the behavior. The extent to which people value the future may be important for understanding and increasing the extent to which they perform preventive health behaviors.

ORAL CONCURRENT SESSION B - METHODS IN JUDGMENT AND DECISION MAKING

"IF I'M BETTER THAN YOU, THEN I'M OK": COMPARATIVE INFORMATION BIASES BELIEFS ABOUT RISK PREVENTION STRATEGIES

Fagerlin A[1], Zikmund-Fisher B[2] and Ubel P[1]

[1]University of Michigan, Ann Arbor, MI; [2]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 their risk of cancer would drop to 1.5% if they took the medication. Would this information change your preference for the medication? Should it change your preference? This study tested whether attitudes toward taking 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%. Three 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 a 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 women 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

PRESENTING PROBABILISTIC INFORMATION IN A RANDOMIZED CLINICAL TRIAL CONSENT DOCUMENT: DO FORMAT AND PREFERENCE MATTER?

MacMillan K[1] and Llewellyn-Thomas H[2]

[1]University of Toronto, Toronto, ON, Canada; [2]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) matches peoples' preferences. Positive or adverse effects on comprehension would be particularly relevant when seeking informed consent for RCT participation. An experimental design was used to test whether comprehension was affected by 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 questionnaires assessing b) 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 preferences. 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 46 subjects who preferred the verbal format, comprehension scores were lower, regardless of whether they received a "match" (n = 20; mean 3.80) or a "mismatch" (n = 26; mean 5.46). Among the 178 subjects who preferred the numeric format, comprehension scores did not differ significantly between those who received a "match" (n = 85, mean 6.83) and those who received a "mismatch" (n = 93, mean 6.78).

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.

ORAL CONCURRENT SESSION A - SIMULATION

THE CASE FOR LIMITED MALIGNANT POTENTIAL BREAST CANCER

Fryback D[1], Rosenberg M[1], Stout N[1], Trentham-Dietz A[1], Remington P[1], Kuruchittham V[1] and Chang P[2]

[1]University of Wisconsin, Madison, WI; [2]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 (LMP) 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 operating 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 focus (<.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 recede 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

Knudsen A, Wang Y, Weinstein M, Weeks J and Kuntz K

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 1978 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 1978 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 1978 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 67.3% is attributable to screening. The number of cancer deaths represents a 13.1% reduction from the estimated number that would have occurred in the absence of changes in risk factors, screening, or treatment (22.0% reduction in 2000). Advancements in treatment alone account for 59.9% of the reduction, while risk factors and screening account for 7.8% and 30.7% 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-LIMTED COUNTRY

Soorapanth S and Sansom S

Centers for Disease Control and Prevention, Atlanta, GA

PURPOSE: To assess the cost effectiveness of HIV rescreening during late pregnancy to prevent perinatal HIV transmission in South Africa, a country with high HIV prevalence and incidence among pregnant women.

METHODS: A decision-analysis model, from a healthcare system perspective, was used to compare the marginal costs and effectiveness of an HIV rescreening strategy during late pregnancy (at 36 weeks or at labor and delivery in addition to an initial test at 24 weeks) with an initial test only early in pregnancy. Because pediatric antiretroviral use is becoming more widely available in South Africa for HIV-infected children, scenarios in which pediatric antiretroviral therapy (ART) was and was not available were analyzed. Sensitivity analysis, based on statistical experimental design technique, is used to estimate the effects of input parameters on outcomes.

RESULTS: At a national average HIV prevalence of 26.5% and annual incidence of 2.7% among pregnant women in South Africa, HIV rescreening would prevent an additional 14 infant infections, and increase program costs by US$18,271 per 10,000 pregnant women tested. If pediatric ART is available, the rescreening would avert disease costs by US$42,768, resulting in a net savings of US$24,500 per 10,000 pregnant women. If pediatric ART is not available, the rescreening would avert pediatric disease treatment costs by US$4,541 per 10,000 pregnant women, resulting in a marginal cost of US$44 per life-year saved. The cost and effectiveness are sensitive to the costs of infant formula and the medical care of HIV-infected children, as well as the probabilities of test acceptance and adherence to peripartum antiretroviral prophylaxis.

CONCLUSIONS: Under the more likely scenario, where pediatric ART is available, HIV rescreening during late pregnancy is cost saving, and the result is robust over wide ranges of parameter values. Without the availability of pediatric ART, rescreening incurs an acceptable cost per infant life-year saved; however, the result is sensitive to variations in parameters. Overall, screening pregnant women twice for HIV in high HIV prevalence, resource-limited settings, appears to be a cost-effective strategy for reducing mother-to-child transmission.

ORAL CONCURRENT SESSION B - SCREENING IN CHRONIC DISEASE

COST EFFECTIVENESS OF HIV SCREENING IN THE ELDERLY

Sanders G[1], Sundaram V[2], Bayoumi A[3], Holodniy M[2] and Owens D[4]

[1]Duke, Durham, NC; [2]VA Palo Alto Health Care System, Palo Alto, CA; [3]St. Michael's Hospital, University of Toronto, Toronto, ON, Canada; [4]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 occurs in patients older than 55. We assessed the cost-effectiveness for screening 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 350cells/ L or their viral load exceeded 55,000 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 $13,700 per quality-adjusted life year (QALY) as compared to current practice. Because of the increased risk of age-specific mortality (unrelated 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, $ 90,800/QALY and $250,300/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

Bentley T, Weinstein M, Willett W and Kuntz 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 folic acid consumption following the 1998 mandate to fortify enriched grain products with folic 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: <201; 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 MIs, 74 NTDs, and 51 colon cancers; there is no risk of B-12 masking 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 10-year period, 85 individuals experienced neurological complications resulting from masked B-12 deficiency, but 9,462 MIs and 2,326 colon cancers were averted. In this age cohort, non-Hispanic whites received the greatest benefits for MIs and colon cancer, while also suffering the greatest burden from B-12 deficiency; Mexican-American males and non-Hispanic black females gained the least benefit from MI and colon cancer prevention.

Conclusions: The potential risks of a small number of cases of B-12 masking were weighed against a substantial predicted benefit of NTDs, MIs, 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.

ORAL CONCURRENT SESSION A - SIMULATION

DEVELOPMENT OF A NATURAL HISTORY MODEL OF LUNG CANCER

McMahon P[1], Kuntz K[2], Weinstein M[2], Weeks J[2], Zaslavsky A[3] and Gazelle G[4]

[1]Massachusetts General Hospital, Harvard University, Boston, MA; [2]Harvard School of Public Health, Boston, MA; [3]Harvard Medical School, Boston, MA; [4]Massachusetts General Hospital, Boston, MA

Purpose: To develop a natural history model of lung cancer to underlie a Lung Cancer Policy Model (LCPM). Unlike published models of lung cancer screening, the LCPM avoids reliance on observed stage shifts as the mechanism for survival gains, thereby allowing estimation of screening effectiveness in populations not represented in trials.

Methods: The natural history model is a state-transition Monte Carlo model. Transitions among the states (general population free from diagnoses, clinical staging, follow-up of indeterminate lesions, and dead) are functions of underlying cancerous or benign lesions, detection via symptoms or incidental imaging exams, test characteristics, and risk factors. Input parameters are based on national survey and vital statistics data, literature estimates of lung cancer doubling times, and incidence of benign lesions observed in CT screening trials. Tumor growth is modeled with Gompertz functions, and metastasis depends on tumor characteristics and individual variation. Parameters for unobservable events (e.g., metastasis) are derived by calibration to multiple endpoints from tumor registries, trials, and the literature. For each simulated individual, characteristics include: demographics; smoking history (cigarettes/day, years of smoking, years since quitting, secondhand smoke exposure); and disease characteristics (locations, sizes, doubling times of ≤3 lung cancers and ≤3 benign lesions, metastases, and nodal involvement). Effects of demographics and smoking on competing risks were estimated from a Bayesian evidence synthesis of individual survey and population data with published cohort studies: coefficients incorporate 1st and 2nd order uncertainty and preserve correlations between predictors.

Results: The first phase model (onset of cancer only) produces results consistent with SEER estimates (incidence by stage, histological type, and size), autopsy studies, screening trials, clinical experience, and cohort studies (e.g., lung cancer in lifelong non-smokers). Age-specific incidence estimates respond in predictable ways to changes in inputs (e.g., higher smoking rates lead to more lung cancers). Calibration to SEER incidence estimates suggests there is a large reservoir of undetected lung cancers.

Conclusions: Our model integrates multiple data sources and is calibrated to tumor registry data, and provides insights into the natural history of lung cancer. Unlike published models, our model design will yield estimates of lead, length, and overdiagnosis biases.

ORAL CONCURRENT SESSION B - SCREENING IN CHRONIC DISEASE

IMPLEMENT OR ANALYZE? A VALUE OF INFORMATION ANALYSIS OF ULTRASOUND FOR BREAST CANCER SCREENING

Schleinitz M[1], Blume J[2] and Berg W[3]

[1]Brown University and Rhode Island Hospital, Providence, RI; [2]Brown University, Providence, RI; [3]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 if 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 and 25% lifetime risk of breast cancer. We determined the value 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 5,000 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 520,000 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 efficacy of US had a maximum value of $121 million, perfect information on the cost of US and on utilities each had a value of $98 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 dollars. This value is concentrated in three areas, the incremental efficacy of US, the cost of US and women's utilities for breast cancer related health states. Further research on these topics should be undertaken before a decision is made on adding US to a breast cancer screening regimen.

ORAL CONCURRENT SESSION B - SCREENING IN CHRONIC DISEASE

A QUALITATIVE STUDY OF AFRICAN-AMERICAN'S DECISION TO SCREEN FOR PROSTATE CANCER: BEST NOT TO KNOW?

Goldberg J

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.

Methods: A triangulation methodology was used involving two qualitative *Methods:* 4 focus groups each with 6-8 low-income, 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 domain 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 better the cognitive and emotional processes underlying this decision. From an applied standpoint, this work points us towards unexplored avenues to improve management decisions for early-stage prostate cancer detection for A-A men at greater risk for the disease.

ORAL CONCURRENT SESSION A - SIMULATION

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

Ivy J

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 or imperfect 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 C[1], Freedberg K[1], Losina E[2] and Walensky R[1]

[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/ul: 30% >500 cells/ul, 30% 201-500 cells/ul, 30% 51-200 cells/ul, and 10% 50 cells/ul. 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 80% and 87% at 3 months, and 68% and 80% at 6 months. We performed a wide range of sensitivity analyses on ART efficacy, prevalence of baseline nevirapine resistance and other parameters.

	Projected life expectancy (months)		
ART regimen	**Overall**	**NVP-resistance**	**No NVP-resistance**
NVP-based	94.3	92.7	94.8
PI-based	95.9	95.8	96.3

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.

ORAL CONCURRENT SESSION B - SCREENING IN CHRONIC DISEASE

COST-EFFECTIVENESS OF CERVICAL CANCER SCREENING IN KENYA, INDIA, PERU, SOUTH AFRICA, AND THAILAND

Goldie S[1], Goldhaber-Fiebert J[2] and Kim J[2]

[1]Harvard School of Public Health, Boston, MA; [2]Harvard University, Boston, MA

Purpose: To assess the cost-effectiveness of alternative cervical cancer screening programs in five regions of the world where organized cytology screening has thus far not been sustainable.

Methods: We calibrated a series of computer-based models that simulate the natural history of cervical cancer to country-specific data from South Africa, Kenya, Thailand, India and Peru. Outcomes include lifetime risk of cancer, years of life saved (YLS), and lifetime costs (international dollars). Strategies differ by initial test - visual inspection with acetic acid (VIA), cervical cytology, and HPV DNA testing (Hybrid Capture II); target age; number of clinic visits; and follow-up protocols. Micro-costing methods were used to assess direct medical, time and programmatic costs. Assumptions were made in collective format with representation from each member of the Alliance for Cervical Cancer Prevention.

Results: In all 5 countries, lifetime cancer risk was reduced by 25-35% with a single lifetime screen (followed by cryotherapy for women with positive results) using either 1-visit VIA or 2-visit HPV, targeted between 35 and 40 years. Although cost components differed considerably between countries, a single-lifetime strategy was identified in each that would cost less than $500 per YLS. Results were most sensitive to targeted screening age (which varied with HIV prevalence), strategy-specific loss to follow-up, and the relative programmatic costs (administration, quality control, training) associated with VIA, HPV, and cytology.

Conclusions: Using country-specific data in South Africa, Kenya, Thailand, India and Peru, cervical cancer screening strategies were identified that would be effective, feasible, and cost-effective.

ORAL CONCURRENT SESSION B - SCREENING IN CHRONIC DISEASE

CHLAMYDIA TRACHOMATIS SCREENING IN U.S. WOMEN: COST-EFFECTIVENESS OF RECENTLY PROPOSED STRATEGIES

Hu D[1], Hook III E[2] and Goldie S[1]

[1]Harvard School of Public Health, Boston, MA; [2]University of Alabama at Birmingham, Birmingham, AL

PURPOSE: Clinical guidelines have traditionally advised annual Chlamydia trachomatis (CT) screening for women < 25 years age. However, recent data indicating a high rate of reinfection within six months of an initial infection has led to a reexamination of these guidelines, including the optimal screening strategy and target age range. The purpose of this study was to assess the cost-effectiveness of recently proposed strategies for chlamydia screening.

METHODS: We developed a mathematical model of the natural history of CT to simulate screening with urine-based nucleic acid amplification tests, diagnosis, and treatment in a cohort of 100,000 U.S. adolescent girls. The following strategies targeted to specific age groups (e.g., 15-19, 15-24, 15-29 years) were compared: (1) no screening, (2) annual screening for all women, (3) annual screening for all women followed by a single repeat test within 3 to 6 months of a positive test result, and (4) annual screening for all women followed by selective semiannual screening for women with a history of previous infection. In addition to the base case, we also evaluated the impact of including the indirect transmission effects of screening programs on the yearly incidence of CT infection and conducted probabilistic sensitivity analyses. Data were from prospective cohort studies, national databases, and published literature.

RESULTS: Screening for chlamydial infection prevented between 11% to 42% of all pelvic inflammatory disease and its sequelae. The most efficient and cost-effective strategy was screening all women ages annually and selectively targeting those with documented CT infection for semiannual surveillance. Relative to annual screening alone, this strategy targeted to women ages 15-24 years had an incremental cost-effectiveness ratio of $2,830 per QALY. In comparison, the same strategy extended to women ages 15-29 years was even more effective and cost $7,490 per QALY. When the indirect effects of screening on the yearly incidence of CT infection were considered, all strategies became more cost-effective. In probabilistic sensitivity analysis, annual screening in women ages 15-29 years followed by semiannual screening for those with documented infection had an incremental cost-effectiveness ratio less than $50,000 per QALY in 99% of simulations.

CONCLUSIONS: Screening all women age 15-29 annually for CT and selectively targeting those with a history of infection for semiannual screening is very cost-effective compared to other well-accepted clinical interventions.

ORAL CONCURRENT SESSION A - COST EFFECTIVE ANALYSIS: METHODS

A PILOT STUDY OF VALUE OF INFORAMTION ANALYSIS TO SUPPORT RESEARCH RECOMMENDATIONS FOR THE NATIONAL INSTITUTE FOR CLINICAL EXCELLENCE

Claxton K[1], Eggington S[2], Ginnelly L[1], Griffin S[1], McCabe C[2], Philips Z[3], Tappenden P[2] and Wailoo A[2]

[1]University of York, York, United Kingdom; [2]University of Sheffield, Sheffield, United Kingdom; [3]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 (GPAs); screening for age related macular degeneration (AMD); neuroaminidase inhibitors (NIs) for influenza; liquid based cytology (LBC); and beta interferons for multiple sclerosis (BIs). 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 reanalysis 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 ranged from £2.8m to £865m). 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 low (£2.8m). In other cases it indicated that additional research should be commissioned, e.g., the EVPI surrounding CD for stroke patients, and GPAs was high (£865m and £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 conditional 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

Epstein D[1], Chalabi Z[2], Claxton K[1] and Sculpher M[1]

[1]University of York, York, United Kingdom; [2]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 pharmaceuticals 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 analysis 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 B - PUBLIC HEALTH 2

OPTIMAL PACKAGE OF WOMEN'S HEALTH SERVICES DELIVERED IN A SINGLE HEALTH CARE VISIT

Kim J[1], Salomon J[2], Weinstein M[1] and Goldie S[1]

[1]Harvard School of Public Health, Boston, MA; [2]Harvard School of Public Health, Cambridge, MA

Purpose: Provision of primary care services for older women in resource-poor settings is a public health priority. Motivated by the momentum of the impending widespread introduction of single lifetime cervical cancer screening in poor countries, we sought to identify potential interventions that might be included in a package of personal health services targeted to women during this single health care visit between the ages of 35 and 45.

Methods: We developed an integer programming (IP) model to maximize disability-adjusted life-years (DALYs) averted from health interventions, subject to budget and human resource constraints in four resource-poor regions. In addition to cervical cancer, interventions for six other diseases were considered: breast cancer, colorectal cancer, cardiovascular disease, depression, iron deficiency, and sexually-transmitted diseases. Inputs to the IP model were calculated using Markov models, which estimated DALYs averted and costs per woman for each intervention. Human resource constraints were expressed as limits to available staff-time (distinguishing clinic and laboratory personnel) for each intervention. Data were obtained from regional reports and surveys published by the World Health Organization, international databases, the published literature, and expert opinion.

Results: If only a budget constraint ($100/woman) is applied, most of the programs would be funded in all regions at a total cost of $41 to $99 per woman, and total benefits of 0.1369 to 0.2613 DALYs averted per woman. With the addition of a staff-time constraint equal to one-half of what is required to implement all interventions, the more staff-intensive interventions, such as screening for breast and colorectal cancer, would be excluded from the package. A more typical scenario in poor countries is that technical laboratory staff is more scarce than clinic staff; we therefore examined a scenario that included the budget constraint and limits of 75% of the clinic staff-time, but only 25% of the lab staff-time. For this analysis, only interventions for cervical cancer, depression, and iron deficiency were included as part of the package.

Conclusions: While only focusing on a select group of diseases and interventions, this analysis demonstrates the advantages of broadening one's analytic scope from assessing costs and benefits associated with a single disease-intervention pair, to using methods of decision sciences to design a package of services for multiple diseases, explicitly taking into account several real-world constraints.

ORAL CONCURRENT SESSION B - PUBLIC HEALTH 2

DESIGN OF AN ULTRASOUND SCREENING PROGRAM FOR ABDOMINAL AORTIC ANEURYSM USING COST-EFFECTIVENESS ANALYSIS

Silverstein M[1], Chaikof E[2], Pitts S[2] and Ballard D[3]

[1]Health Care Research, West Roxbury, MA; [2]Emory University School of Medicine, Atlanta, GA; [3]Baylor Health Care System, Dallas, TX

Purpose: Randomized controlled trials of ultrasound (USG) screening for abdominal aortic aneurysm (AAA) have shown USG effective in reducing AAA related mortality but not total mortality in older men. Recent adoption of endovascular surgery (EVAR) in place of open surgical repair (OSR), patient refusal of USG, incidental discovery and early surgery for AAA, and refusal or being unfit for elective surgery could influence the cost-effectiveness of USG in clinical practice settings.

Methods: A 17-state Markov model was developed to compare USG screening for AAA to usual care by evaluating the incremental cost-effectiveness ratio (ICER), number of ruptured AAAs and AAA deaths prevented, relative risk (RR) of ruptured AAA and AAA death, and number needed to be invited (NNI) or screened (NNS) to prevent an AAA rupture or AAA death. Outcomes of a single USG screening examination of white men at age 65 with follow-up at 3, 6, 9 or 12 months for intermediate sized AAA (4.5-5.5 cm) and either EVAR or OSR for large AAA >5.5 cm) over 20 years were analyzed for a cohort of 100,000 men. Monte Carlo sensitivity analysis was performed and acceptability curves were constructed.

Results: USG screening would have prevented 764 AAA ruptures (RR = 0.69), with a NNI = 131 and NNS = 105. USG would reduce total mortality by ~ 1% (764/61408). Net QALYs accrue in USG screening after 6 years and continue to occur through year 20. USG screening would have an ICER of $15,722 per QALY gained, (95% CI $14,318 - $18,177) even with follow-up of intermediate size AAA every 3 months. 95% of ICERs were < $17,625. The optimal age to initiate USG is 61; the ICER is < $20,000 per QALY for screening as early as age 55. As EVAR replaces OSR for elective AAA surgery the ICER decreases to $14,138. The ICER is not sensitive to likely ranges for refusal of screening, refusal of elective surgery, incidental discovery of AAA, or early surgery for small or intermediate size AAA.

Conclusion: One-time ultrasound screening for AAA in older men at age 65, or possibly as early as age 55, can be recommended as cost-effective in reducing risk of AAA-related mortality but would have a small impact in reducing total mortality of older white men.

ORAL CONCURRENT SESSION A - COST EFFECTIVE ANALYSIS: METHODS

LIMITATIONS OF ACCEPTABILITY CURVES FOR PRESENTING UNCERTAINTY IN COST-EFFECTIVENESS ANALYSES

Groot Koerkamp B[1], Hunink M[2], Stijnen T[2], Hammitt J[1], Kuntz K[1] and Weinstein M[1]

[1]Harvard School of Public Health, Boston, MA; [2]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 correlations 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)Maximizing 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 low 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 A - COST EFFECTIVE ANALYSIS: METHODS

MICROCOSTING METHODS FOR LEVERAGING LIMITED DATA IN FIVE DEVELOPING COUNTRIES

Goldhaber-Fiebert J and Goldie S

Harvard School of Public Health and the Alliance for Cervical Cancer Prevention, Boston, MA

Purpose: To estimate direct medical, direct non-medical, and patient time costs for several cervical cancer screening and treatment modalities in India, Kenya, Peru, South Africa, and Thailand, using limited primary data.

Method(s): Direct medical costs for several cervical screening modalities including specimen transport and laboratory services, diagnostic work-up, pre-cancer treatment, and cancer care were estimated based on literature reviews, national health system and civil service data, hospital charge sheets, primary micro-costing studies, World Bank and World Health Organization data, simplified spatial models, and laboratory productivity models. Direct non-medical costs for patient time and transport were estimated, accounting for inadequate rural transport and women's larger role in informal or unpaid labor, using primary survey data as well as World Bank and International Labour Organization data. A set of standardizing assumptions was developed in consultation with clinical experts and program sites in each country. All cost estimates were standardized to year 2000 international dollars using purchasing power parity conversion factors.

Results: Three screening modalities with different resource requirements for level of provider training, test kits, equipment, laboratory facilities were assessed. Location of service delivery was accounted for by country which led to differences in transport cost and time for patients receiving care as well as transport of laboratory specimens. Our results demonstrate that patient time and transport costs in countries with inadequate infrastructure and rural populations can equal or supersede the direct medical costs of screening. Other factors impacting total costs such as recurrent programmatic costs to maintain quality and efficiency as well as the potentially non-linear implications of population coverage levels were also explored. We produced ranges of plausible cost estimates for each country which when compared to study results and estimates from other sources had a high degree of concordance, providing general face validity.

Conclusions: Our cost estimation methodologies rely primarily on publicly available international dataset inputs and required only a small number of country-specific inputs. These methods, as applied to cervical cancer screening in five different regions of the world, provide one approach for conducting cost-effectiveness analyses when primary data are limited and/or incomplete or programs have not yet been implemented.

ORAL CONCURRENT SESSION B - PUBLIC HEALTH 2

PREDICTING OPTIMAL ANTHRAX RESPONSE PARAMETERS: IMPACT OF INCIDENCE

Hupert N[1], Neukermans C[1], Cuomo J[2], Koshy M[1] and Callahan M[1]

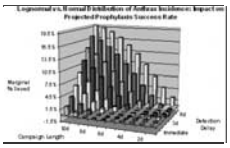
[1]Weill Medical College, Cornell University, New York, NY; [2]RAND, Santa Monica, CA

Purpose: The incubation period and consequent epidemic curve caused by a large-scale anthrax exposure in 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 ≥97.5% 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 whose illness is prevented by timely antibiotic prophylaxis. Model assumptions include random mixing, medication efficacy (i.e., not antibiotic resistant) and patient compliance. We used two basic incidence curves for inhalational anthrax: a lognormal distribution of hospitalized cases derived from the 1979 Sverdlovsk outbreak (mean 2.398 (SD 0.713) with peak cases on day 7, from Brookmeyer, et al., Biostatistics, 2001) and a normal distribution estimated from combining the first two weeks of Sverdlovsk with the 11 inhalational anthrax cases in the 2001 U.S. outbreak (mean and peak day 8.3 (SD 3.0) using @Risk for Excel). We ran multiple sensitivity analyses varying the day of peak cases +/- 3 days for each curve.

Results: Prevention of symptomatic inhalational anthrax in ≥97.5% of exposed persons is attainable only with extremely rapid agent detection and prophylaxis response (≤3d using lognormal incidence, ≤2d using normal incidence) and rapid completion of prophylaxis campaigns (≤5d for both distributions). Within this range of optimal response parameters, we found little difference in projected outcomes between the baseline lognormal and normal incidence distributions (Graph). However, substantial (>5%) divergence of outcomes occurs if anthrax incidence is more rapid than the baseline distributions predict (peak cases ≤day 6), with the lognormal distribution leading to consistently better prophylaxis outcomes.

Conclusions: Prophylaxis campaign parameters and projected patient outcomes were similar using a previously published lognormal distribution and a newly-calculated normal distribution estimating the incidence of inhalational anthrax after an outdoor exposure. These results provide an evidence base for bioterrorism preparedness planning.



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[1], Bauch C[2], Foty R[1], Pham B[1] and Krahn M[3]

[1]Glaxo SmithKline Canada, Oakville, ON, Canada; [2]University of Guelph, Guelph, ON, Canada; [3]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 (adjusted), factors for underreporting, and change in incidence over time; and 4) deriving estimates of the latter factors from the fitted model.

MEDLINE and EMBASE were searched ["hepatitis" and "Canada"(n = 365)] to identify studies that reported HA epidemiological data. Prevalence rates from 13 studies (a total of 28 seroprevalence studies from 55 epidemiological ones) involving persons with no known risk factors were used (Step-1). HA reported incidence data were obtained from the National Notifiable Disease Registry System (Step-2). Average reported incidence (1980-2000) 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).

Results: Approximately 1/3 of the Canadian-born population had serologic evidence of past infection, with a prevalence rate ranging from < 1% in ages < 10, 3% in 10-19, 1%-10% in 20-29, 7%-15% in 30-39, and above 30% in ages 40+.

The average annual reported incidence rate of HA was 6.2/100,000 (range 4.3, 9.5) and 6.8 (range 2.9, 10.8) from 1980-1989 and 1990-2000, respectively.

The model estimated an annual average of 11,244 (95% CI [8,148 - 13,591]) HA cases in Canada, 6.61 [4.79 - 7.99] times the average annual reported incidence of 1,701 cases from 1980-2000.

For a typical birth cohort (e.g., n = 403,434 in 1990), the model predicts 3,114 HA cases by age 1; 16,707 by age 20; and 29,930 by age 30 (i.e., a prevalence of 7.4% in 2020).

Conclusion: Reliable estimates of true incidence are important for estimating the health benefits and costs of immunization programs. In cases when there is discordance between true and reported incidence of disease, catalytic modeling provides a useful framework for synthesizing fragmented epidemiological data to derive true incidence.

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

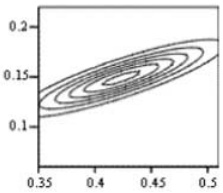
Hazen G

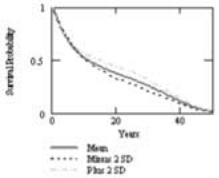
Northwestern University, Evanston, IL

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 m for those not permanently cured. Analysts can choose a combination (p, μ) to accurately fit cause-specific survival data from the SEER cancer registry. However, if a model sensitivity analysis on (p, μ) is desired, it is not obvious what neighborhood of the estimated (p, μ) should be explored. We address this question by using Bayesian analysis to derive the approximate posterior distribution of (p, μ) given survival data.

Methods: Bayesian theory states that the large-sample posterior distribution of (p, μ) 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, μ) given SEER survival data is approximately bivariate normal with mean/SD equal to 0.423/0.048 for p and 0.149/0.020 for μ, and correlation 0.859. For sensitivity analysis, a representative one-dimensional (p, μ)-neighborhood 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 50-year-old white female is shown below. We present like results for other stages of ovarian cancer.





ORAL CONCURRENT SESSION A - COST EFFECTIVE ANALYSIS: METHODS

$50,000 PER QALY: INERTIA, INDIFFERENCE, OR IRRATIONALITY?

Braithwaite 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 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 1276 interventions, and US government surveys. We examined thresholds of $20,000, $50,000, $100,000, and unlimited $/QALY. All costs were in 2002 US dollars.

Results: 44% of health care services cost less than $20,000/QALY, and 78% cost less than $100,000/QALY. In the base case analysis (allocated resources = 14.9% 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 GNP, LE was highest using a threshold of $100,000/QALY (73.0 years) because resources were partially unused with lower thresholds (e.g., 33% of resources unused with $20,000/QALY, LE 70.5 years) and were used inefficiently with higher thresholds (unlimited $/QALY, LE 71.2 years). When allocated resources were decreased to 5%, LE was highest (70.2 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

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 disability 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 impact 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 $2100 [8 mo] to $34,300 [1 mo] per 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 [7 mo lead-time] to $46,700 [1 mo]), while decreasing PID cases 14.4% [1 mo] to 60.6% [12 mo]. When the incidence of infection decreases to 5%/yr, costs/QALY for 6 month compared to 12 month screening are > $75,000 for all lead-times (range $75,800 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 B - PUBLIC HEALTH 2

EVALUATING THE U.S. FOLIC ACID FORTIFICATION POLICY: DID WE SUCCEED?

Bentley T, Willett W, Weinstein M and Kuntz 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 to 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; 1% (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.

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

Liu Y, Sander B and Maetzel A

University Health Network, Toronto, ON, Canada

Background: Surgical fusion is routinely proposed to patients with rheumatoid involvement of the cervial 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 surgery as 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. PROPH would be outperformed by NEURO only if natural progression to RII/RIII is less than 10% or if more than 40% of prophylactic surgeries result in RII/RIII or death.

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 RII or RIII. High-quality, prospective, observational studies are needed to corroborate these findings. Cost-effectiveness and quality of life assessment would provide added value.

ORAL CONCURRENT SESSION A - CLINICAL STRATEGIES AND GUIDELINES

EVALUATING CLINICAL GUIDELINES AND PREDICTIVE INSTRUMENTS FOR ACUTE CARDIAC ISCHEMIA

Mills B[1], Estrada S[1], Reyna V[1],Stuckler D[2], Baker J[1], Burnham G [1], Tesoro J[1], Lloyd F [3]

[1]University of Texas at Arlington, Arlington, TX; [2] Yale University, New Haven, CT; [3] Mayo 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 Insensitive Predictive Instrument (ACI-TIPI), 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); overlap was 19%. Ethnic breakdown was 72% 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/or 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. Sequential logistic regression analyses with occurrence of a cardiac procedure as a criterion variable were conducted, 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 ACI-TIPI 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, ACI-TIPI scores, AHCPR risk levels, and newer guidelines' AMI risk level predicted procedures—and each 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 had lower predictive validity for the occurrence of subsequent cardiac diagnoses. Paradoxically, 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.

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, Kabeto 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 of the American people has hardly changed. Nevertheless, financial status may play a larger role under some circumstances. Using data from the Health and Retirement Study (HRS), we examined the hypothesis that the relationship between financial status and subjective well-being would be moderated by health status. Specifically we predicted that financial status would matter more in predicting well-being for people in poorer health, or for people who had recently experienced the onset of a physical disability.

Methods: The HRS is a longitudinal study of over 16,000 individuals aged 51 to 61, surveyed every two years since 1992. The HRS uses measures of financial status—including savings, pensions, investments, and real estate—to create an index of overall net worth. In addition, it contains many measures of health status and disability, and a well-being measure. Using this data, we determined whether participants had recently experienced the onset of a significant physical disability. The longitudinal design allowed us to test whether financial resources prior to the onset of disability affected later well-being.

Results: Using hierarchical linear modeling in a combination of cross-sectional and longitudinal analyses, we found evidence that:

1) Wealth was a stronger predictor of well-being for people with lower subjective health ($p < .01$), for people who have cancer ($p < .01$), and for people who currently have a physical disability ($p < .01$).

2) Wealth was a much stronger predictor for people who report a new disability, accounting for 3 times as much variance in well-being as for people who were not disabled ($p < .01$).

3) Prior wealth buffered well-being after the onset of a disability ($p < .05$).

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 B - HEALTH ECONOMICS

EXTRAPOLATING UTILITY AND COST DATA IN TERMINALLY ILL PATIENTS: A PRAGMATIC MODELING APPROACH

van den Hout W[1], Van Houwelingen J[1], Noordijk E[1], Kramer G[2] and Leer J[2]

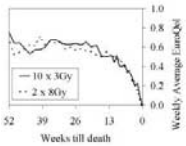
[1]Leiden University Medical Center, Leiden, Netherlands; [2]Joint Centre of Oncology RADIAN, Arnhem, Netherlands

Purpose: In a trial comparing two radiotherapy schedules for palliation of non-small-cell lung cancer, the longer schedule unexpectedly provided significantly better survival. A model was constructed to estimate lifelong costs and quality adjusted life years.

Methods: Patients were randomized to receive a 10 x 3Gy or a 2 x 8Gy radiotherapy schedule (n = 297, median age 69 years, tumor stage IIIA/B or IV). They were followed for at most one year, with symptom palliation as primary outcome measure. Utility was measured using the EuroQol classification system. Societal costs were measured only during the initial 12 weeks, using questionnaires in a subset of 56 patients. There were three types of missing data: 10% of the patients were alive at study closure, about 20% of the questionnaires were not returned, and cost questionnaires were returned for at most 12 weeks. Neglecting missing data would lead to bias, because at the end of follow-up there were significant differences in survival (13% versus 7%, $p = 0.05$) and utility (0.28 versus 0.17, $p = 0.03$). The observed relationships between utility, remaining lifetime and costs were used to model censured survival times and missing utility and costs data.

Results: For the 10 x 3Gy and 2 x 8Gy schedules respectively, average life expectancy was estimated at 35 and 26 weeks ($p = 0.02$) and quality adjusted at 19 and 12 weeks ($p = 0.03$). The differences were larger and more significant than for the non-extrapolated data. Radiotherapy costs were estimated at Ευρο Σανσ∀3.700 and Ευρο Σανσ∀2.200 ($p > 0.001$) and non-radiotherapy costs at Ευρο Σανσ∀5.900 and Ευρο Σανσ∀5.000 ($p = 0.17$), with estimated cost-utility ratio 19.000 Ευρο Σανσ∀/QALY. Without taking costs during the additional lifetime into account, the cost-utility ratio would have been 12.000 Ευρο Σανσ∀/QALY.

Conclusions: The increase in life expectancy (36%) for the longer radiotherapy schedule led to a less than proportional increase in non-radiotherapy costs (19%), because part of the costs were associated with the deteriorating health preceding death. Including the costs during the additional lifetime led to a higher but still acceptable cost-utility ratio.



ORAL CONCURRENT SESSION A - CLINICAL STRATEGIES AND GUIDELINES

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

Janssens A[1], Steyerberg E[1], Van Duijn C[1] and Criswell L[2]

[1]Erasmus MC, Rotterdam, Netherlands; [2]University of California, San Francisco, San Francisco, CA

Objective: Genetic factors influence disease progression, but may also predispose to other, non-genetic, predictors of prognosis such as age at onset and severity of early symptoms. The additional value of genetic information to the prediction of prognosis may therefore differ in the presence or absence of these 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 180 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 0.85 and specificity 0.40. The SE was a significant predictor of radiographic damage (multivariate OR 5.1 [95% CI 1.9, 14.2]), next to Health Assessment Questionnaire (HAQ) score (OR 4.6 per 1.0 point [1.9, 11.0]), rheumatoid factor (RF) positivity (OR 5.4 [1.7, 16.8]) disease duration (OR 1.5 per 5 years [1.1, 1.9]) and family income (OR 0.8 per 1000 US$ [0.6, 1.0]). Of these predictors, HAQ score ($p = 0.03$) and RF positivity ($p < 0.001$) were also associated to the SE. At the individual patient level, taking into account the risk profiles of individual women, the LR+ varied from 1.0 to 2.3, the LR- from 0.2 to 0.5, the sensitivity from 0.60 to 1.00 and the specificity from 0.00 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 were RF negative, these figures were 1.9, 0.4, 0.78 and 0.57.

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 A - CLINICAL STRATEGIES AND GUIDELINES

THE CLINICAL AND ECONOMIC IMPACT OF RAPID ADOPTION OF DRUG-ELUTING STENTS INTO A TERTIARY CARE PRACTICE

Long K, McMurtry E, Lennon R, Bresnahan J, Mathew V, Ting H, Holmes D and Rihal C

Mayo Clinic College of Medicine, Rochester, MN

Background: Randomized trials demonstrate reduced restenosis and need for repeat revascularization with drug-eluting stents (DES) compared with bare-metal stents (BMS). Increased supply costs are of serious economic concern to providers and healthcare organizations, and the full impact of commercial release of DES on providers is unknown.

Purpose: We assessed the clinical and economic outcomes of patients who underwent percutaneous coronary intervention (PCI) with either DES (sirolimus-eluting) or BMS by physician preference following commercial release of DES in the United States.

Methods: We considered all patients who received 2.5 to 3.5 mm stents between 4/22/03 and 11/30/03 in analysis. Clinical, angiographic, and follow-up data were prospectively recorded in the Mayo PCI Registry. Administrative data was used to estimate total procedural costs (hospital and physician) and length of stay (LOS). A standardized, year 2003 constant-dollar cost estimate was used to value health care services. Logistic regression and Cox proportional models were used to estimate risk of adverse events; generalized linear modeling was used to estimate costs adjusting for clinical and angiographic characteristics.

Results: 569 PCI patients received DES (mean 1.51 stents) and 204 patients received BMS (mean 1.61 stents). BMS patients were more likely to have urgent PCI (31% vs 11%), recent MI (37% vs 11%), shock (10% vs 1%), old MI (60% vs 44%), B2 or C lesion (83% vs 75%), thrombus (40% vs 19%), and stents in native arteries (97.1% vs 95.6%). We observed fewer in-hospital complications among DES-treated patients (1.9% versus 7.4%; $p < 0.001$) but this difference was not significant in adjusted models (OR 0.66; 95% CI: 0.26, 1.70). The adjusted hazard ratio for major adverse cardiovascular events during a median follow-up of 176 days was 0.69 ($p=0.14$). Cost prediction models suggest DES treated patients incur $1,328 more, on average, in cath lab expenses ($4,561 vs $3,233). Adjusted mean total costs were $654 higher for DES treated patients (95% CI of cost difference: $299, $1,009). Adjusted LOS was similar at 1.9 days.

Conclusions: Adoption of DES was extremely rapid following commercial release. BMS continue to be used among higher risk patients. Hospital costs are significantly higher with DES. Without adequate technology reimbursement, healthcare organizations may face an ethical dilemma: controlled access to what may be optimal patient care in order to maintain financial stability.

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

Tunceli K[1], Bradley C[2], Elston Lafata J[1] and Nerenz D[1]

[1]Henry Ford Health System, Detroit, MI; [2]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 leave workforce than diabetic men with complications. This study provides partial evidence that diabetes negatively effect work productivity of employees who survive and remain working. Although diabetes does 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

= UBRADLEY C[1], Neumark D[2], Oberst K[1], Brennan S[3] and Schenk M[3]

[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 = 496), 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 are 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 a 90% subject retention rate, that reflects the natural course of cancer's ill effects within the first years 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 6 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 study. The methods used to estimate productivity loss and other economic outcomes 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

Sox C[1], Koepsell T[1], Doctor J[1] and Christakis D[2]

[1]University of Washington, Seattle, WA; [2]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 1502 pediatricians practicing in the U.S. randomly selected from the AMA master list, and sent up to 4 times to non-responders. The questionnaire presented a vignette of a 5-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 that 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's sensitivity and specificity, and a non-technical explanation of these terms. Subjects were asked to estimate the likelihood that this patient had pertussis. We conducted t tests 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%. 81% 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 (45% vs. 38%, $p < 0.001$). Subjects who received technical or non-technical support were more likely than controls to chose 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 A - CLINICAL STRATEGIES AND GUIDELINES

MEDICATION ERRORS IN THE INTENSIVE CARE UNIT: COMPARISON BETWEEN COMPUTERIZED VERSUS PAPER-BASED PHYSICIAN ORDER ENTRY

Colpaert K[1], CLaus B[2], Somers A[2], Robays H[2], Vandewoude K[1], Oeyen S[1] and Decruyenaere J[1]

[1]University Hospital Ghent, Ghent, Belgium; [2]Ghent University Hospital , Belgium, Ghent, Belgium

Purpose: Medication errors (ME) in the Intensive Care Unit (ICU) are frequent and lead to attributable patient morbidity and mortality, increased length of ICU stay and substantial extra costs. We investigated if the introduction of a computerized ICU system (Deio Clinisoft, General Electric) reduces the incidence and severity of ME.

Methods: A one-month (March 2004) prospective trial was conducted comparing ME in a computerized unit (CU) versus a paper-based unit (PBU) in a university hospital. Every medication entry was registered and evaluated for ME, Medication Prescription Errors (MPE) and Rule Violations (RV) by an experienced and unit-independent clinical pharmacist. An independent panel evaluated the severity class of every ME according the NCC MERP guidelines.

Results: 160 patient-days resulting in 2662 medication prescriptions were evaluated. There were no differences in the CU-patients compared with the PBU-patients regarding age, admission reason, severity of illness (APACHE, SAPS, SOFA score), renal failure and number of administered medications. The incidence of ME, MPE, RV was significantly lower in the CU compared with the PBU (5,1% vs. 24,7%, 0,9% vs. 15,0% and 0,17% vs. 7,90%, respectively; all $P < 0.001$). The most significant reduction were seen in dosing errors (1,6% vs. 7,4%), dose errors in renal failure (0,4% vs. 4,5%) and wrong name errors (0,42% vs. 8,87%)(all $P < 0.001$). The most frequent drug classes involved were cardiovascular medication and antibiotics in both groups, but more errors in sedative, analgesic and gastro-intestinal medication prescription were seen in the PBU. The severity of ME was mainly class A (36% vs. 52%) and C (54% vs. 44%) according to the NCC MERP classification, but in the PBU more class D errors occurred (1,4% vs. 2,5%, $P = 0.025$). No fatal errors were seen. Allergy status was recorded in 69% of CU patients vs. only 2% in PBU. Patient weight was recorded in 60% of CU patients vs. 18% of PBU patients.

Conclusions: The ICU computerization, including the medication order entry, resulted in a significant decrease in the occurrence and severity of medication errors in the ICU. The main reasons are the more adequate recording of allergy status and patient weight and the drug dose recommendation by the ICU computerized system.

ORAL CONCURRENT SESSION B - HEALTH ECONOMICS

COST SHARING POLICIES AND HEALTH CARE DEMAND: LESSONS FROM CANADA

Li X[1], Guh D[2], Marra C[3], Lacaille D[1], Esdaile J[1] and Anis A[1]

[1]University of British Columbia, Vancouver, BC, Canada; [2]St. Paul's Hospital, Vancouver, BC, Canada; [3]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 (percentage 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, seniors paid 100% of their dispensing fee costs 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(Premium Assistance) whereby seniors paid a maximum of $25 and $10 per prescription to an annual maximum of $275 and $200, 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 elasticites. 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 driven cost containment concepts such a patient cost-sharing might have the unintended impact of increasing overall utilization.

	Exceed deductible				Not exceed deductible			
	2001 Plan A	2002 Plan A	2001 Plan A	2002 Plan A1	2001 Plan A	2002 Plan A	2001 Plan A	2002 PlanA1
N	1,170		675		1,813		1,569	
Rx's filled	-0.13		-0.20		-0.21		-0.29	
Physician visits	0.50		0.89		0.08		0.30	

ORAL CONCURRENT SESSION B - HEALTH ECONOMICS

A MULTI-STATE STOCHASTIC MODEL FOR ASSESSING COVARIATE EFFECTS ON SURVIVAL AND COST

GARDINER J, LUO Z, SIRBU C, BRADLEY C and GIVEN C

Michigan State University, East Lansing, MI

PURPOSE: We describe a longitudinal multi-state regression model to assess covariate effects on patient outcomes and cost, addressing issues of heteroscedasticity, skewness and censoring.

METHODS: Event histories of patients are modeled by finite-state Markov processes in continuous time. Cost incur at transitions between health states and in sojourn in health states. These expenditure streams over a specified period are combined to form net present values. First, we model the impact of treatments, patient-specific demographic and clinical characteristics on survival through the transition intensities using a Cox regression model. Second, we use a mixed model for transition and sojourn costs with transition times as random effects and patient characteristics as fixed effects. Censoring of patient costs and outcomes, discounting of costs and benefits are all incorporated into the model.

APPLICATION: We used data on 605 incident cases of lung, prostate, colon and breast cancer recruited from Michigan community hospitals during 1994-1997. Charge data were obtained from Medicare claim files which included reimbursements for inpatient and outpatient care, physician services, home health care/skilled nursing facilities, laboratory tests and treatments. A patient's total charge was used as a proxy for cost. Complete cost was deemed known if a patient had at least 2 years of follow up, or if death occurred within 2 years of diagnosis. Otherwise, a patient's cost was regarded as incomplete. Approximately 31% of the cost data were censored. Physical function as assessed by the SF-36 at baseline and subsequently at 4 waves was categorized as "good" or "fair." During follow up at least one transition in physical function occurred in 50.4% of the patients, and 22.7% remained in fair physical function throughout. Our regression model revealed that treatments (surgery, chemotherapy/radiation) cancer site and comorbidity were significantly associated with cost. Transitions were influenced by these same factors as well as a count of symptoms. Mortality was 10%. Survival was best among breast cancer patients and worst among lung cancer patients.

CONCLUSIONS: The joint regression model provides a rigorous and flexible statistical approach to assessing the influence of patient variables on both cost and health outcome. While incorporating explanatory variables, the model accommodates heteroscedasticity, skewness and censoring. It also provides a unified framework for inference on summary measures used in cost-effectiveness analysis.

POSTER SESSION - PUBLIC HEALTH; METHODOLOGICAL ADVANCES

MARKOV MODEL OF A PARTIALLY-EFFECTIVE HIV VACCINE ON ADOLESCENT WOMEN IN SOUTH AFRICA

Amirfar S, Hollenberg J and Abdool Karim S

Weill Medical College of Cornell University, New York, NY

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 birth and breastfeeding. At the end of 10 years, the number of HIV infections and death 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 (cohort of 500,000): If in four years South Africa were to adopt the use of prophylactic antiretrovirals to prevent mother-to-child transmission in the peripartum period, an additional 31% of infant infections are prevented at 1 year. If the majority of AIDS patients have access to HAART by 2008, 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 important role in HIV prevention of adolescents and infants in South Africa no matter the domestic policy implemented.

Assumes no use of antiretrovirals as perinatal prophylaxis or as HAART			
Vaccine efficacy (%)	Adolescent infections	Infant infections	Adolescent HIV deaths
No vaccine	435,480	122,350	42,160
30	-7%	-12%	-12 %
50	-15%	-22%	-21%

POSTER SESSION - PUBLIC HEALTH; METHODOLOGICAL ADVANCES

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

Gohler A[1], Dietz R[2], Osterziel K[2] and Siebert U[3]

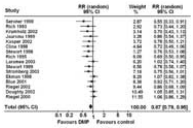
[1] Charitè Campus Virchow Klinikum, Humboldt Medical School, Berlin, Germany; [2]Charitè Campus Virchow Klinikum, Humboldt Medical School, Berlin, Germany; [3]Massachusetts General Hospital, Harvard Medical School, Boston, MA

Purpose: Hospital admission for CHF is an important public health problem. Although several randomized controlled trials (RCT) have successfully linked DMPs to improved outcomes and reduced readmission rates, most effects are statistically not significant and vary regarding their magnitude. We sought to (1) systematically combine the evidence on efficacy of disease management programs (DMP) in the treatment of congestive heart failure (CHF), (2) to identify reasons for the existing heterogeneity and (3) identify publication bias.

Methods: We performed a systematic MEDLINE research on RCTs investigating DMPs for CHF treatment from 1966-05/2004. We included all studies that were performed randomized, included the core curriculum of a DMP (i.e., patient education, medication optimization, follow-up after discharge), and reported mortality and hospitalization as outcomes. We performed a meta-analysis using random or fixed effects models depending on the statistical heterogeneity of effects and estimated the pooled relative risk (RR) with 95% confidence intervals (95%CI). We assessed effect heterogeneity using meta-regressions to identify the impact of covariates on the DMP effect size. Publication bias was assessed by inspection of funnel plots.

Results: Our analysis included 16 studies from 5 different countries with data from 2868 patients. A random effects model which compared DMPs vs. control groups yielded a pooled RR of 0.79 (95%CI 0.65-0.97) for mortality and of 0.87 (95%CI 0.79-0.95) for rehospitalization during DMP (see figure). Meta-regression analysis identified mean age, severity of disease (NYHA distribution), and duration of intervention as statistically significant variables explaining the heterogeneity. Funnel plot was asymmetric indicating a bias towards positive studies.

Conclusions: DMPs in the CHF treatment lead to a clinically relevant and statistically significant reduction of mortality and rehospitalization. Heterogeneity across studies could largely be explained by age, severity of disease, and duration of DMP. Our analysis may overestimate the true DMP effect, because of a potential publication bias.



Forest plot for DMP effect on hospitalization, sorted by study weight

POSTER SESSION - PUBLIC HEALTH; METHODOLOGICAL ADVANCES

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

Ekwueme D[1], Hutchinson A[1], Dean H[2] and Kim A[3]

[1]Centers of Disease Control and Prevention, Atlanta, GA; [2]Centers for Disease Control and Prevention, Atlanta, GA; [3]Business Computer Application, Inc./Northrop Grumman Mission Systems, Atlanta, GA

PURPOSE: The Centers for Disease Control and Prevention's "Guidance for HIV Partner Counseling and Referral Services"(PCRS) defined 3 different strategies to notify persons who may have been exposed to HIV infection: (1) provider—the health provider notifies partners; (2) patient—the index patient notifies partners and (3) mixed—the index patient, assisted by the provider, notifies partners. Despite the fact that PCRS has been found to be cost saving to society, there is still concern about the costs, particularly for provider PCRS, among public health decision makers. We therefore compared the costs and effectiveness of the 3 strategies.

METHODS: We developed a probabilistic decision model to calculate costs and effectiveness. To the determine effectiveness of the PCRS strategies, we conducted a systematic review of the literature from 1990-2003. Cost and resource utilization data were obtained from health departments, national wage data and the literature. The analysis was conducted from the societal and program perspectives. We performed univariate and multivariate sensitivity analyses using Monte Carlo simulation to ascertain the robustness of the base-case results.

RESULTS: Data from 8 studies were pooled for provider PCRS effectiveness parameters and 2 studies provided data on mixed and patient PCRS strategies. Notification and testing rates were 68% and 63% (provider), 62% and 48% (mixed) and 32% and 50% (patient). The newly identified HIV seropositivity rate was approximately 20% for all strategies. The mean per person cost of PCRS ranged from $152 (patient) to $253 (provider) from the societal perspective and $125 (patient) to $212 (provider) from the program perspective. Provider PCRS was most cost-effective and patient PCRS was least cost-effective from both perspectives. The mixed PCRS strategy was eliminated through extended dominance. The incremental cost per newly identified HIV-infected person with provider notification was $1,750 (95% CI: $1737-$1764) from the societal perspective and $1,515 (95% CI: $1503-$1527) from the program perspective. From the societal perspective, sensitivity analysis did not appreciably alter results; however, changing some of the effectiveness parameters to their lowest values for sensitivity analysis made the mixed strategy more cost-effective from the program 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

EFFECTS OF BEHAVIORAL INTERVENTIONS CONDUCTED IN U.S. STD CLINICS ON THE REDUCTION OF UNSAFE SEXUAL BEHAVIORS AND STD REINFECTION: META-ANALYSIS

Kim A[1], Griffin T[2], Marrero-Gonzalez P[1], Rama S[1] and Crepaz N[3]

[1]Business Computer Application, Inc./Northrop Grumman Mission Systems, Atlanta, GA; [2]Northrop Grumman Mission Systems, Program Evaluation Research Branch, Atlanta, GA; [3]Division of HIV/AIDS Prevention, NCHSTP, CDC, Atlanta, GA

Background: Patients attending STD clinics for testing and medical treatment are at elevated risk for HIV and STD infections. Interventions conducted in this setting provide a great opportunity to modify high-risk behaviors, thus preventing HIV and STD transmission. We performed a systematic review and meta-analysis of the HIV/STD behavioral prevention literature to identify interventions designed to reduce risky sexual behavior among U.S. STD clinic patients.

Methods: Systematic searches of 5 electronic databases were conducted to identify HIV/STD behavioral intervention evaluations published from 1988-2004 targeting STD clinic patients in the U.S. Studies were included if they had control/comparison groups and evaluated changes in biologic outcomes (STD reinfection assessed by medical chart review) and/or sex-risk behaviors (self-reported condom use or unprotected sex). Effect sizes (ESs) were estimated with odds ratios (ORs) and random-effects models were used to aggregate data.

Results: Thirteen intervention studies (11 RCTs, 2 quasi-experimental studies) met the inclusion criteria. The majority (ranging from 55% to 100%) of the study samples were ethnic minority groups, primarily African Americans and Hispanics. Overall, behavioral interventions had a significant impact on reduction of sexual risk behavior (OR = 0.55; 95% CI = 0.33, 0.91, 10 ESs). However, the intervention effects on preventing new STD reinfection fell short of significance (OR = 0.88; 95% CI = 0.70, 1.09; 14 ESs). Interventions provided skills training on condom use and negotiating safer sex with partners were effective in reducing sexual risk behaviors.

Conclusions: Behavioral interventions successfully modify high-risk behaviors of STD clinic patients, primarily African Americans and Hispanics—two populations which are greatly affected by HIV/STD epidemics. Although it was not significant, a 12% reduction of STD was observed. These different findings may be due to the sensitivity of outcome measures to detect changes or data collection and outcome classification issues, therefore further research is required. Nevertheless, interventions that promoted technical and interpersonal skills were effective in reducing sexual risk behaviors in STD clinics.

POSTER SESSION - PUBLIC HEALTH; METHODOLOGICAL ADVANCES

100 YEARS OF INHALATIONAL ANTHRAX: A SYSTEMATIC REVIEW OF CASES FROM 1900 TO 2001

Holty J[1], Bravata D[1], Liu H[1], Wilhelm E[1], Zaric G[2], McDonald K[1] and Owens D[1]

[1]Stanford University, Stanford, CA; [2]University of Western Ontario, London, ON, Canada

Purpose: The mortality rate form inhalational anthrax during the U.S. 2001 attack was significantly lower than had been observed historically. Current multi-drug therapy and intensive care may have improved survival. Our goal was to systematically evaluate the predictors of disease progression and mortality on which to base treatment recommendations and to compare recent cases to historical controls.

Methods: We searched MEDLINE from 1964 to November 2003 and the bibliographies of all retrieved articles for case reports of inhalational anthrax presenting between 1900 and 2003. We considered articles in any language eligible for inclusion if the authors established a definitive diagnosis of inhalational anthrax. We estimated mortality as a function of duration between symptom onset and antibiotic initiation using a Weibull model of disease progression and observed mortality rates. We used a logistic regression and Kaplan-Meier curves to evaluate the effects of patient characteristics and alternative treatment regimens on mortality and disease progression.

Results: We found 73 reports of 60 cases of inhalational anthrax presenting between 1900 and 2001. Mortality was not significantly associated with the type of therapy given (e.g. single or multi-drug regimens or anthrax anti-serum), but was significantly reduced among U.S. 2001 cases (OR 0.05; 95% CI: 0.004-0.6) and those patients who had antibiotics or serum initiated during the prodromal phase of disease (OR 0.05; 95% CI: 0.008-0.3). Advancing age significantly increased the observed mortality (OR 1.08; 95%CI: 1.01-1.15). Analyzing all cases, the untreated prodromal phase mean duration (4.2 days) was longer than historical accounts. This duration was not significantly affected by type of anthrax exposure (e.g. occupational verses bioterrorism) or when therapy was initiated. Patients who progressed to the fulminant phase had a mortality rate of 96% (regardless of the treatment they received). Most surviving patients (10/13) had pleural effusions requiring drainage. The observed mean time from symptom onset to antibiotic initiation during the U.S. 2001 attack was 3.9 days. Initiation of antibiotics on day two rather than day four reduces mortality by an estimated 56%.

Conclusions: Rapid initiation of antibiotics or anti-serum during the prodromal phase of inhalational anthrax dramatically reduces mortality. Despites advances in supportive care, fulminant phase inhalational anthrax is usually fatal. Efforts to improve early diagnosis and timely initiation of appropriate antibiotics are critical in reducing mortality.

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 P

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 form 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 limitation. 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 54 U.S. states and territories (see Table 1). Physician 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, HMO 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

	Unhealthy Days			
	Physical	Mental	Physical or mental	Activity limitation
Mean (days)	3.9	1.3	4.0	2.1
Standard Errors	0.9	0.7	0.9	0.7

POSTER SESSION - PUBLIC HEALTH; METHODOLOGICAL ADVANCES

PROGNOSIS AND THE VALUE OF CANCER SCREENING

Howard D

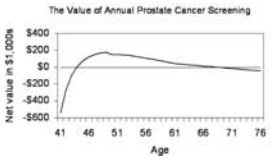
Emory University, Atlanta, GA

Purpose. Aggressive screening for asymptomatic disease is a bedrock of public health policy in the US. Cancer screening guidelines are generally of the form "screen starting at age x." However, the benefits of early detection and treatment decline sharply with age because patients are more likely to die from comorbid conditions. The purpose of this study was to assess the relationship between age and the benefits of screening.

Methods. An analytical model expresses the net benefit from screening as a function of age. The benefit function has two components. The first is the survival advantage from early detection (valued in dollars). This component, which is declining in patient age, accounts for the risk of death from competing causes. The second component is the cost per detected case, which is a function of the incidence rate and is increasing in age. Parameters from the literature were used to compute the value function for the case of prostate cancer screening.

Results. The benefit from screening is not monotonically increasing in age, as implied by current screening policies. There will be cases when net benefits are positive in an interval (x,y) and negative otherwise. For annual prostate cancer screening, x=45 and y=68.

Conclusion. In some cases, it makes sense to stop screening after a certain age. Screening recommendations appear to be influenced more by the likelihood of detection than the benefit of treatment once detected.



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 Behavior 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 the 20 states that used an optional module consisting of 11 questions about QOL. The analysis was restricted to the 365 people who reported vision problems (VP) to be their primary health concern, and the 14,812 individuals that self-reported one disease out of a list of 13 (e.g. hearing problems, arthritis, cardiovascular disease, 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 dependant 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 (0.21), p<0.0001), and of not enough rest (7.37 (0.87) vs. 11.54 (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 (0.15), p = 0.01) and a higher percent needing assistance with routine needs (47% (5.2%) vs. 30% (0.8%), 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 (0.88) vs. 7.13 (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

PREVALENCE AND COST IMPLICATIONS OF CHANGING DEFINITIONS FOR CHD RISK FACTORS

Kaplan R and Ganiats T

UCSD School of Medicine, 9500 Gilman Drive, La Jolla, CA

Purpose: Disease is often characterized as binary. A "diagnosis" is either present or absent. For most conditions, however, there is a spectrum of illness, and the definition of disease is based on a value (threshold) along a continuum. Sometimes the patient gets a very abnormal test value early in the disease, however for many diseases the patient slowly progresses from 'normal' to 'borderline' to abnormal. In order to give patients the benefit of early treatment, there has been a trend toward identification of disease earlier in the process by moving thresholds downward. We evaluated recent guidelines that lowered thresholds by defining prehypertension (SBP >120 mmHg), impaired fasting glucose (plasma glucose 100 mg/dl) and mild hypercholesterolemia (LDL >100 mg/dl)

Methods: Data from the third National Health and Nutrition Examination Survey (NHANES-III) were used to estimate the prevalence of prehypertenstion, impaired fasting glucose, and mild hypercholesterolemia in the US population age 50 years or older.

Results: We estimate that 37% of the 62 million Americans age 50 or older have fasting glucose levels greater than 100 mg/dl. Further, nearly 40% of the population have systolic blood pressures greater than 120 mmHg. Nearly three quarters of the population (73%) have LDL cholesterol levels greater than 100 mg/dl. According to this preliminary analysis, 97% of American adults age 50 or older have at least one of the three conditions.

Conclusions: Changes in diagnostic thresholds significantly expand the market for health care. Our analysis suggests that virtually the entire population of adults older than age 50 are eligible for a diagnosis (and potential intervention) under the new definitions of just three conditions. These changes in diagnostic thresholds are likely to have profound impacts on the costs of health care, but their effects upon population health have not been comprehensively evaluated.

POSTER SESSION - PUBLIC HEALTH; METHODOLOGICAL ADVANCES

COST-EFFECTIVENESS OF TYPE 2 DIABETES PREVENTION IN THE GENERAL POPULATION IN A US SETTING

Minshall M[1], Palmer A[2], Roze S[2] and Valentine W[2]

[1]CORE-USA, Indianapolis, IN; [2]CORE Center for Outcomes Research, Binningen, Switzerland

PURPOSE: An epidemic increase in the prevalence of type 2 diabetes (T2D) is predicted as the population ages, and western lifestyles become increasingly unhealthy. A number of recent clinical studies have demonstrated that the onset of T2D is delayed or prevented in high risk groups by lifestyle changes and/or medications. T2D patients have more than double the mortality rates and higher annual treatment costs of equally-matched non-diabetes patients. A simulation model was developed to assess acceptable cost limits for a broad-based general population-targeted program aimed at reducing the incidence of T2D by (mean±standard deviation) 30±10%.

METHODS: A Markov/second order Monte Carlo model used age- and gender-specific incidence of—and increased direct medical costs and mortality associated with—T2D in a US setting to forecast life expectancy (LE) and costs with or without intervention. Data were derived from published sources. The maximum costs/person a payer could outlay to achieve a 30±10%.reduction in T2D incidence a) without increasing the healthcare budget, and b) remaining within an attractive incremental cost-effectiveness ratio (ICER) limit (i.e. <$50,000/life year gained versus no intervention). Costs and LE were discounted at 3% annually. A health insurance perspective was taken. Sensitivity analysis was performed to identify parameters with important impacts on outcomes.

RESULTS: A hypothetical diabetes prevention intervention aimed at a general population with mean age 51 years would improve undiscounted LE by 0.33±0.16 years/person (0.14±0.07 discounted 3% p.a.). Up to a cost of $68/year/person, the program be cost saving overall. The ICER versus no intervention would be <$50,000 at a cost of $520/person/year. Sensitivity analysis revealed that the age of the target population, effectiveness of the program, incidence of diabetes, and increase in mortality associated with diabetes have large impacts on the results.

CONCLUSIONS: Diabetes prevention programs aimed at a general population could be cost saving or cost-effective if the costs of the program do not exceed the limits identified in this analysis. In higher-risk populations where the incidence of diabetes and effects of intervention are greater, these cost limits may be higher.

POSTER SESSION - PUBLIC HEALTH; METHODOLOGICAL ADVANCES

SHOULD WE TREAT PEOPLE WITH OCULAR HYPERTENSION TO PREVENT GLAUCOMA?

Kymes 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 ?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 simulation.

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 $30,950/QALY. Treatment of all persons with OH was not cost-effective with an ICER of $332,303/QALY. In sensitivity analysis, the cost of medication and the incidence of POAG were those factors that most influenced the cost-effectiveness of "Treat 2%." However, an extreme change in the value of these variables was required to change the decision. Monte Carlo simulation found the cost-effectiveness of the "Treat 2%" strategy to be robust to model assumptions, and the strategy remained the most cost-effective where the cost-effectiveness threshold was $26,000/QALY or greater. Below this level, the "Treat 5%" was the most cost-effective.

Conclusions: While medical treatment of all people with OH to prevent POAG is not cost-effective, it is cost-effective to treat people with OH and an annual risk of POAG of at least 2%.

ICER Strategy (QALYs)	Total Cost* (Cost/QALYs)	Total Effectiveness* (QALYs)	Incremental Cost	Incremental Effectiveness
Treat 5%	$ 5,724	12.44		
Treat no one	5,749	12.40	$ 26	-0.0300	Dominated
Treat 2%	7,129	12.48	1,406	0.0454	$ 30,950
Treat all	14,553	12.50	7,424	0.0223	332,303

POSTER SESSION - PUBLIC HEALTH; METHODOLOGICAL ADVANCES

HEALTH AND ECONOMIC EFFECTS OF A PROPOSED CALIFORNIA CIGARETTE EXCISE TAX INCREASE

Ong M[1], Alamar B[2] and Glantz S[2]

[1]VA Palo Alto and Stanford University, Stanford, CA; [2]University of California, San Francisco, San Francisco, CA

Purpose: To determine the health and economic effects of a new cigarette excise tax increase for the state of California.

Methods: We used a simulation model to estimate the effects of an additional tax increase on smoking behavior and consumption, 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 300 heart attacks, 150 strokes, 250 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 cents of this tax increase for the state tobacco control program would provide an additional $200 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 annually 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 cents 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

Robinson R[1], Hayes J[2], Baker S[3], Casavant M[3] and Mahan J[4]

[1]The Ohio State University, Mount Gilead, OH; [2]Childrens Research Institute, Columbus, OH; [3]Children's Hospital, Columbus, OH; [4]The Ohio State University, Columbus, OH

Purpose: As the therapeutic use of a medication increases, risk of unintentional overdose, medication errors and intentional overdose increases. The objective of this study is to identify the characteristic 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 socioeconomic 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 = 131,199), exposures (e.g., ingestion, inhalation, dermal contact) (n = 116,898) and information calls (n = 14,301) 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 B[1], Neukermans C[1], Narette S[2], Joseph P[2] and Fitzgerald D[3]

[1]Weill Medical College of Cornell University, New York, NY; [2]GHESKIO Center, Port-Au-Prince, Haiti; [3]Weill Medical College of Cornell Univeristy, New York, NY

Purpose: An estimated 10,500 newborns each year in Haiti are born to women infected with syphilis and at risk of developing congenital syphilis. We determined the incremental benefits of testing programs that screen pregnant women and deliver treatment to prevent congenital syphilis as part of comprehensive HIV VCT programs. We compared strategies that use the current RPR test, with results delivered at 1-week follow-up, with strategies that use a new rapid test, with results delivered immediately.

Methods: A decision analytic model simulated a cohort of pregnant women with access to prenatal care. Outcomes include disability adjusted life years lost (DALYs) of newborns and number of perinatal transmissions. Strategies include treatment based on screening by (1) rapid test only; (2) RPR test only; (3) both rapid and RPR test; (4) diagnosis of primary/secondary syphilis by physical exam only; (5) no treatment. Empiric treatment was also considered. A complete course of treatment is 3 rounds of 2.4 million units benzathine penicillin at one-week intervals. Data on stillbirths and congenital syphilis with and without treatment were from US historical data. Data from Haiti included national maternal syphilis seroprevalence; prenatal visit return rates; and test performance under field and optimal laboratory conditions from a WHO-TDR sponsored trial conducted at the GHESKIO Center in Port-au-Prince.

Results: Rapid testing with results delivered immediately minimizes DALYs and avoids the most transmissions. Under field conditions, the incremental benefit of rapid testing compared to the next best strategy (RPR test only) is 0.12 DALYs per woman tested and 347 transmissions avoided per 100,000 women tested. Rapid testing provides a benefit of 0.34 DALYs and 975 transmissions avoided versus physical exam, the current standard of care for women with access to prenatal care but not testing. Empiric treatment would avoid 99 additional transmissions, but has other costs not considered in this model. For women with access to optimal laboratory-based testing, the incremental benefit of rapid testing is 0.15 DALYs and 446 transmissions avoided compared to RPR testing only. Results were sensitive to rapid test performance and return rate after first visit.

Conclusions: Addition of rapid syphilis testing and treatment protocols to HIV VCT programs in Haiti for women with access to prenatal care will prevent perinatal syphilis transmissions and increase quality-adjusted life expectancies of newborns.

POSTER SESSION - PUBLIC HEALTH; METHODOLOGICAL ADVANCES

HEALTH CARE UTILIZATION IN PERINATALLY HIV-INFECTED U.S. CHILDREN: 2001 V. 1995

Sansom S, Anderson J, Sukalac T, Soorapanth S, Clark J, Bertolli J and Dominguez K

Centers for Disease Control and Prevention, Atlanta, GA

Purpose: Current measures of health care demanded by children perinatally infected with HIV are useful to assess the cost-effectiveness of prevention programs. Assessing health care utilization over time demonstrates the impact of new therapies.

Methods: The Pediatric Spectrum of Disease (PSD) study collects longitudinal data on HIV-infected children in six sites in the U.S. from the time of HIV diagnosis or exposure through death. We assessed health care utilization including hospitalizations, and HIV-related prescription drugs and laboratory tests, during the years 1995 and 2001 for all perinatally HIV infected children PSD. We assessed statistical significance between the two years using t-tests of difference in proportions and means.

Results: In 2001, 1,903 perinatally infected children were enrolled in PSD, with a median age of 9.0 years, compared with 2,246 children enrolled in 1995 with a median age of 4.9 years. During 2001, 13.9% of enrollees were hospitalized at least once (854 total hospitalizations) compared with 31.3% hospitalized at least once (2,796 total hospitalizations) in 1995 ($p < .05$). The rate of hospitalization was .2 per child-year in 2001 compared with .7 per child-year in 1995 ($p < .05$). Average length of hospital stay was 9.5 days in 2001, compared with 8.5 days in 1995 for an average 2.0 hospital days/child-year in 2001 and 5.1 hospital days/child-year in 1995 ($p < .05$). On average, children were prescribed 3.5 antiretroviral drugs in 2001 compared with 2.1 in 1995 ($p < .05$), and an average of 17.5 HIV-related lab tests were ordered for each child during 2001 compared with 12.8 during 1995 ($p < .05$).

Conclusions: Hospitalization among perinatally infected children in the United States decreased significantly from 1995 to 2001. The number of prescribed antiretroviral drugs and ordered laboratory tests increased. Future analysis should include the costs associated with health care utilization to determine if resource use overall has declined.

POSTER SESSION - PUBLIC HEALTH; METHODOLOGICAL ADVANCES

RISKS AND BENEFITS OF THE ORAQUICK HIV-1 RAPID TEST IN MULTIPAROUS WOMEN

Shroff S and Wong J

Tufts-New England Medical Center, Boston, MA

Background: 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 9-13%. The OraQuick rapid HIV test has a reported 99.6% sensitivity and 100% specificity, however multiparous women can have false positive results.

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

Methods: 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, 85.3% of pregnant women not receiving prenatal care were multiparous. The CDC reported HIV prevalence in women of childbearing age varies from 0.8% to 4.0%. For this prevalence range, of 33,103 multiparous women not receiving prenatal care, 267-1,335 women will be HIV-positive. OraQuick would identify 266-1329 of those HIV-infected, but also 2,136-2,207 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 $132 (ZDV = $41.15, OraQuick = $39, Western Blot = $52) per positive test for a total of $317,064-$466,752 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 seroprevalence data should be obtained.

POSTER SESSION - PUBLIC HEALTH; METHODOLOGICAL ADVANCES

MEASURING HOW PATIENTS USE INTERNET-BASED DECISION AIDS

Weber G[1], Fortin J[1], Landau C[2], Cyr M[3] and Col N[1]

[1]Harvard Medical School, Boston, MA; [2]Rhode Island Hospital, Providence, RI; [3]Brown 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. Tiered 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 (68%) completed all aspects of WISDOM; of these, 88% chose to send summary reports to their clinicians. Patients spent an average of 31 minutes (m) using WISDOM, with 11m allocated to inputting data, 4m interacting with the risk report, 3m with the symptom/treatment chart, and 1.5m with prevention options. Users were most interested initially in learning about diet and vitamins (43%), followed by lifestyle changes (33%), herbals (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 38s each on breast cancer and hip fracture, and 24s 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

DATA-DRIVEN BAYESIAN BELIEF NETWORK FOR FORECASTING EMPLOYMENT ONE YEAR AFTER TRAUMATIC BRAIN INJURY

Castro 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 DTREG.

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.

Sensitivity, specificity, and overall classification rates were calculated and a receiver operating characteristic analysis was done for each model.

RESULTS: The complex Bayesian network (CBN) had the highest sensitivity, specificity, and overall classification rate at 88%, 72%, and 80% respectively. It also had the best performance in the ROC analysis with an area under the curve of 0.86.

The independent belief network had a sensitivity of 78%, specificity of 67%, and overall classification rate of 72%. The logistic regression model had a sensitivity of 74%, specificity of 70%, and overall classification rate of 72%. The CART model had a sensitivity of 75%, specificity of 72%, and overall classification rate of 73%. The areas under the ROC curve for these models were 0.79 for the independent belief network, 0.80 for logistic regression, and 0.77 for the CART. Statistical tests showed the CBN performed significantly better than the other approaches.

CONCLUSION: The results demonstrate the accuracy of a data-driven complex Bayesian network in determining chance of employment one year after traumatic brain injury. Automated decision support systems should consider this approach when making forecasting judgments.

POSTER SESSION - PUBLIC HEALTH; METHODOLOGICAL ADVANCES

EVSI FOR SURVIVAL TRIALS

BRENNAN A, CHILCOTT J, O'HAGAN A and KHARROUBI S

University of Sheffield, SHEFFIELD, United Kingdom

Purpose: To develop methods to undertake 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 (bivariate lognormal to characterise uncertain Weibull parameters) and a lognormal distribution for prior relative risk. Monte Carlo simulation of notional matched pairs of individual level trial participants is undertaken given this prior uncertainty (sample size = n). Data are censored for a pre-specified duration of follow-up d. For each simulated trial dataset, Bayesian updating of the Weibull and lognormal relative risk probability distributions enables probabilistic sensitivity analysis. EVSI are undertaken using Laplace approximation. Five alternative follow-up durations and sample sizes are examined.

Results: The illustrative model results shows that for a sample size of n = 100, and a duration of follow-up of 1 year, EVSI = 3.2m. For n = 100 and 2 years follow-up, EVSI = 3.9m. For n = 1,000 and 4 years follow-up, EVSI = 4.9m. Other combinations show the trade-offs between sample size and follow-up duration.

Conclusions: The methods described provide an approach to undertaking EVSI calculation to 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

MODEL PERFORMANCE TRACKING AND RISK ADJUSTMENT USING STATISTICAL PROCESS CONTROL METHODS

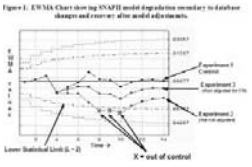
Chuo J[1], Zupancic J[2] and Ohno-Machado L[1]

[1]Brigham and Women's Hospital, Brookline, MA; [2]Beth Israel Deaconess Medical Center, Boston, MA

Purpose. Use Statistical Process Control Methods to identify when a clinical model needs to adapt to environmental variations in order to maintain its value.

Methods. SPC was used to scrutinize the performance of the Score of Neonatal Physiology II model (SNAPII), a validated logistic regression model that predicts in-hospital mortality for neonates in the intensive care unit. The SNAPII model was applied to a database of 3437 newborns admitted to 7 neonatal units in New England from 1994 to 1996. To accommodate temporal analysis, cases were chronologically arranged and grouped into 14 sequential periods containing equal number of cases. The c-index (equivalent to area under the ROC curve) was used to quantify model performance. To identify periods in which model performance fell below SPC-determined limits, we graphed the exponentially weighted moving average (EWMA) for each period. The EWMA approach can manage non-normally distributed data such as the c-index and detect small sustained deviations in its trend. The same analysis (experiment #2) was repeated on the SNAPII model for different clinical scenarios. These scenarios were artificially constructed by varying the patient data. The analysis of experiment #3 applied a risk-adjusted SNAPII model that accounted for the variations made in experiment #2.

Results. The C-indices resulting from applying the SNAPII model to the 14 periods were within the SPC limits of their overall mean, 0.88 (remained in statistical control); thereby reaffirming the model's validity. The model's performance shifted out of statistical control when applied to data containing unanticipated predictor-outcome relationships. The performance loss was partially recovered by risk-adjusting the model for the new relations. Conclusion. SPC framework can detect changes that could remain unnoticed using the current quality control measures, thereby helping to determine when and how a model may need to adapt to a new environment in order to remain useful. Such identification may improve understanding of how patients, diseases, and the environment interact in a dynamic clinical setting.



POSTER SESSION - PUBLIC HEALTH; METHODOLOGICAL ADVANCES

COMPARISON OF METHODS FOR ESTIMATING PARTIAL EVPI IN ECONOMIC ANALYSIS

Coyle D

Ottawa Health Research Institute, Ottawa, ON, Canada

Purpose: Value of information analysis provides a framework for the analysis of uncertainty within economic analysis, by focussing on the value of obtaining further information to reduce uncertainty. The mathematical definition of the expected value of perfect information (EVPI) is fixed though there are different methods in the literature for its estimation. In this study these methods are explored and compared.

Methods: Analysis was conducting using a disease model for Parkinson's disease. Five methods for estimating partial EVPIs (EVPPIs) were used: a single Monte Carlo simulation (MCS) method, the unit normal loss integral (UNLI) method, a two stage method using MCS, a two stage method using MCS and quadrature and a difference method requiring two MCS. The single MCS method is appropriate only when the complement of the data set is multi linear in net benefit, the UNLI method is appropriate when data parameters are linear in net benefit and are normally distributed. The other three methods are general methods in that they are argued to be appropriate for all data parameters. EVPPI was estimated for each individual parameter in the model as well as for three groups of parameters (transition probabilities, costs and utilities). In each case, EVPPI was estimated by the three general methods and either of the other methods if appropriate. Analysis compared estimates of EVPPI based on alternate numbers of replications as well as the complexity of analysis required for estimation.

Results: Using 5000 replications, four methods returned similar results for EVPPIs. 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 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 EVPPI. Thus, where appropriate EVPPI should be estimated using either the single MCS or UNLI method. However, as often in cases where neither of these methods is appropriate, the two stage MCS and quadrature methods should be used.

POSTER SESSION - PUBLIC HEALTH; METHODOLOGICAL ADVANCES

A COMPARISON OF THE CASE-CONTROL AND CASE-CROSSOVER DESIGNS FOR ESTIMATING COSTS OF FALL-RELATED INJURIES AMONG MEDICARE BENEFICIARIES 65 AND OLDER

Finkelstein E and Chen H

Research Triangle Institute, Research Triangle Park, NC

PURPOSE: Although the case-crossover design has been used widely in epidemiological studies as an alternative to the case-control design, it is rarely applied to cost-of-illness studies. In this study, unit costs for a series of hospitalized and non-hospitalized injuries are computed using the two approaches to allow for a direct comparison of the results.

METHODS: We use claims data from the Medicare fee-for-service 5% Standard Analytical Files. The claims contain detailed payment information for all covered services for hospital inpatient, hospital outpatient, skilled nursing, home health, hospice, physicians/supplier services and durable medical equipment. For the case-control design, fallers are tracked for 1-year after their first fall, and costs are compared to annual costs for a comparison sample of non-fallers. A generalized linear regression model with a logged dependent variable and a gamma distribution is used to estimate costs while controlling for covariates (including differences in demographics and comorbidities between fallers and non-fallers). The case-crossover sample compares monthly costs of fallers pre- and post-fall and is again estimated using a generalized linear regression model with a logged dependent variable and gamma distribution. This model also uses a cluster option to account for clustering within individuals across months.

RESULTS: We present unit costs for falls requiring 1) a hospitalization resulting in a live discharge, 2) an emergency department visit not resulting in an admission, and 3) falls requiring office-based or hospital outpatient visits only. Using the case-control design these costs are $22,260; $3,890; and $5,040 respectively. Using the case-crossover design, these estimates are reduced to $20,920; $3,230; and $4,200.

CONCLUSION: Estimates from the case-control design are between 6% and 20% greater than those from the case-crossover approach. These differences likely result from our inability to control for comorbidity differences between fallers and non-fallers in the case-control design. The case-crossover design, although computationally more intensive, avoids this pitfall and allows for producing more accurate estimates.

POSTER SESSION - PUBLIC HEALTH; METHODOLOGICAL ADVANCES

APPLICATION OF A BAYESIAN MARKOV MODEL TO DECISION MAKING

Ding M[1], Xing Y[2], Cox D[1] and Cormier J[2]

[1]Rice University, Houston, TX; [2]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 long-term health outcomes of patients with high-risk (stage III) melanoma under alternative 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 4 age cohorts: < 40, 40-49, 50-59, and > 59 years old. The survival time in each state was assumed to have an exponential distribution. Transition probabilities between health states were estimated according to Bayes Rule: Posterior distribution is proportional to Prior distribution *Likelihood. An informative gamma prior was used based on previously published studies. A sensitivity analysis was performed using different priors with the same mean and different variances. Health state utilities were incorporated from published data. The models were implemented in Splus.

Results: For high-risk melanoma patients, the model predicted that treatment with IFN- resulted in 6.94 QALYs compared to 5.12 QALYs for patients in the control group. The incremental effectiveness for patients treated with IFN- was 1.82 QALYs with a 95% credible interval of 0.48 to 3.37. Table 1 summarizes the effectiveness (QALYs) and incremental effectiveness for each age cohort. The results of the sensitivity analysis were very robust to the selection of prior distributions.

Table 1 QALYs for Various Age Cohorts

Age	Control	IFN-	Incremental Effectiveness	95% Credible Interval
<40	7	10.40	1.93	-3.60, 8.52
40-49	5.42	11.21	5.79	1.58, 12.22
50-59	6.54	5.25	-1.29	–4.01, 1.50
>59	4.17	4.68	0.50	–1.08, 2.67

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

Heckendorn J[1], Riis J[2], Ubel P[1] and Smith D[1]

[1]University of Michigan, Ann Arbor, MI; [2]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., "$10,000,000,000.00"), 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. 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 = 208). 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.

Results: Study 1, which directly compared the two WTP measures, indicated that the newly devised 'percentage' WTP measure showed fewer outliers ($p < .01$), had better overall distributional properties, and was not correlated with income ($r = .01$; the traditional WTP was strongly correlated with income, $r = .40$, $p < .001$). 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 ($p < .01$).

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 distributional properties, greater sensitivity to severity of health condition and was not correlated with income. This new method could be useful in many research applications.

POSTER SESSION - PUBLIC HEALTH; METHODOLOGICAL ADVANCES

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

LaJoie A[1], McCabe S[1] and Edgell S[2]

[1]School of Public Health and Information Sciences, University of Louisville, Louisville, KY; [2]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 Tinel's Sign, the Phalen's test, and the nerve-conduction velocity test. Results It was found that the Tinel's Sign and the Phalen's test are both highly sensitive, .97 and .92 respectively and specific .91 and .88 respectively.

Conclusions: The estimates 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 H, Barshay A, Grimm L, Robinson K and Sheffield C

Johns 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 statisticians or for 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's minimally clinically important difference for the absolute difference in percentage is elicited. Then, the user chooses the appropriate statistical model, depending on which parameters 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 models for this purpose.

Conclusion: 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

PROBABILISTIC MODELING OF MEDICAL ERRORS IN RADIOTHERAPY

Lee R[1], Ekaette E[1], Kelly K[2], Newcomb C[2], Craighead P[2] and Dunscombe P[2]

[1]University of Calgary, Calgary, AB, Canada; [2]Tom 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 positive net benefit.

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 B, Ubel P and Smith D

University 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, 152 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 "ping pong" set 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 time as the number of months since their amputation increased ($p < 0.05$). 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

USING COMBINED PATIENT CHARACTERISTICS TO PREDICT DIRECT AND INDIRECT COSTS IN PARKINSON'S DISEASE

Siebert U[1], Bornschein B[2], Spottke A[3], Berger K[4], Oertel W[5] and Dodel R[3]

[1]Massachusetts General Hospital, Harvard Medical School, Boston, MA; [2]Ludwig Maximilians University Munich, Germany, Munich, Germany; [3]University of Bonn, Bonn, Germany; [4]MERG Medical Economics Research Group, Munich, Germany; [5]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 $p <0.001$) and PDQ-39 ($p = 0.03$). The probability for the presence of indirect costs was dependent on age ($p < 0.001$), UPDRS ($p = 0.03$), PDQ-39 ($p = 0.04$), presence of depression ($p = 0.02$), and falls ($p = 0.006$). The magnitude of indirect costs was a function of clinical state ($p = 0.003$) and falls ($p = 0.007$). Variance explained by the models (adjusted R-square) ranged from 24% to 28%.

Conclusions: We identified UPDRS and quality-of-life as 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

THIEBAUD P[1], Nichol M[2] and PATEL B[3]

[1]University of Southern California, SAN DIEGO, CA; [2]University of Southern California, Los Angeles, CA; [3]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 379 type II diabetes patients. The estimation procedure uses a classification and regression tree (C&RT) 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, C&RT 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 C&RT, and the results regarding the effect of insulin treatment are compared.

Results: Logistic regression models created with C&RT 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 C&RT-based models, however, show a larger effect for insulin treatment on health care utilization, particularly by patients over the age of 53. This may indicate that C&RT-based models can better correct for selection bias.

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

POSTER SESSION - PUBLIC HEALTH; METHODOLOGICAL ADVANCES

USING INVERSE DECISION THEORY TO DETERMINE A C/B RATIO FOR SCREENING AND DIAGNOSIS OF CERVICAL INTRAEPITHELIAL NEOPLASIA (CIN)

Swartz R[1], Cox D[2], Cantor S[1], Davies K[2] and Follen M[1]

[1]The University of Texas M. D. Anderson Cancer Center, Houston, TX; [2]Rice University, Houston, TX

Purpose: We propose a new method, Inverse Decision Theory, to determine bounds on relative costs and losses for medical decisions. These bounds can be used to compare different treatment strategies and to determine a range for the cost-benefit (C/B) ratio (Metz CE. *SEM. NUC. MED.*: 1978: 8: 283-298). *Methods:* We used Bayesian sequential decision theory to model the sequence of tests used to screen and diagnose cervical intraepithelial neoplasia (CIN) (It includes a Papanicolaou smear followed by a colposcopic exam, followed by a biopsy. Positive results indicate progression to the next test, with positive biopsy indicating treatment.) We assumed this current standard of care (SOC) for CIN was optimal. We identified operating characteristics for the tests from the literature and a sample of 624 women. We then worked backwards through the sequential decision problem and solved for the costs (both monetary and patient outcomes associated with performing a test) and losses (both monetary and patient outcomes associated with a treatment decision) associated with the SOC. This yielded bounds on each of the costs and losses considered. Furthermore, the bounds identified conditions for situations when competing strategies would perform better than the current SOC and these boundaries were linear constraints on the costs and losses. We used linear programming to determine bounds on the C/B ratio.

Results: We found that treating someone based on a positive colposcopic result (forgoing the biopsy) is better than the SOC when the net loss for treating non-diseased women is less than 1.4 times the incremental cost of biopsy. The C/B ratio was found to have a lower bound of 0.0862.

Conclusions: The lower bound suggests that the C/B ratio for screening and diagnosing CIN is higher than previously reported and that the current standard of care has a high degree of benefit relative to its cost. IDT is an effective method for characterizing the costs and losses associated with established decision rules such as a screening and diagnosis care setting.

This research was supported in part by The National Cancer Institute grant 2PO1-CA82710 and in part by a fellowship supported by the National Cancer Institute grant R25 CA57730, Robert M. Chamberlain, Ph.D., Principal Investigator.

POSTER SESSION - PUBLIC HEALTH; METHODOLOGICAL ADVANCES

IS THERE A SPECIAL AWARD FOR THE EXPERIMENTAL DRUG IN RANDOMIZED CONTROLLED TRIALS? THE CASE OF INTERFERON FOR CHRONIC HEPATITIS C

Tinè F[1], Pagliaro L[1] and Attanasio M[2]

[1]V. Cervello Hospital, Palermo, Italy; [2]University of Palermo, Palermo, Italy

Purpose: an association between the experimental drug and benefit has been reported from RCTs of Fluoxetine. We searched for the same relationship 69 RCTs of Interferon alpha for Chronic Hepatitis C (CHC) that used the same dose of 9 MU/week at least in one of the arms.

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%, cirrhotics 21%, genotype-1 61%, yearly schedule 28%, high quality score (Jadad > 3) 28%, multicenter RCT 52%, unblind RCT 72%, oriental country 17%, large trial (> 200 pts) 19%, funding for profit 51%, no-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 4936 patients were sustained responders (benefit 17%). Only four variables maintained statistical significance after forward selection: mono vs multicentric trial (OR 1.9, 95%CI 1.5-2.5), 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 1.4, 95%CI 1.0-1.9). 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 (SROC) CURVE VERSUS BAYESIAN HIERARCHICAL MODELS FOR ASSESSMENT OF DIAGNOSTIC TESTS

Xing Y[1], Cong X[2], Foy M[2], Ding M[2], Cox D[2], Hunt K[1] and Cormier J[1]

[1]M.D. Anderson Cancer center, Houston, TX; [2]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 preoperative chemotherapy in patients with breast cancer.

Methods: A systematic review was conducted of studies which examined the results of SNB following preoperative chemotherapy. Inclusion criteria required completion axillary lymph node dissection as a test "gold standard." Robust resistant regression method was used to construct SROC curves and compared to test 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%. The adjusted parameters using the SROC curve revealed a global sensitivity of 87% (95% confidence interval, 82-93). In the Bayesian analysis, the beta model resulted in a posterior estimate of sensitivity of 83% (95% credible interval, 72-91), while 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 estimate of sensitivity for SNB following preoperative 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 hierarchical models 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

COMPARISON OF SF-6D, HUI, AND EQ-5D MEAN UTILITIES: A SYSTEMATIC REVIEW

Bosch J, Erasmus MC, Rotterdam, Netherlands

Purpose: In cost-effectiveness analyses performed from the societal perspective, SF-6D, Health Utilities Index (HUI), and/or EuroQol-5D (EQ-5D) utilities are frequently included to calculate quality-adjusted life years. An increasing number of studies compare these utilities and report on the disagreement between these measures. The objective of our study was to review the published studies that compare the utilities of SF-6D, HUI3, and EQ-5D and investigate whether mean utilities differ systematically.

Methods: A Medline-search of the English literature was performed and references of the derived articles were reviewed. Studies were included if patients assessed their current health and if mean utilities were reported of at least two of the three measures SF-6D, HUI3, and EQ-5D. We used random effects models to generate pooled results.

Results: Seven studies published between 2000 and 2004 met the inclusion criteria. In all studies SF-6D was assessed, in 3 studies HUI3, and in 6 studies EQ-5D. In total 762 patients were included with various diseases (e.g., liver transplant, arthritis, cardiovascular disease). All three measures were completed by 249 patients, among the remaining patients 246 patients completed the SF-6D and HUI3 and 267 patients the SF-6D and EQ-5D. The mean age of the patients was was 57 years and 56% was male. The mean utilities across all measures and studies ranged from 0.43 to 0.70. The pooled difference in mean utilities was 0.02 (95%CI: - 0.001-0.05) between SF-6D and HUI3, 0.10 (95%CI: 0.07-0.12) between SF-6D and EQ-5D, and 0.05 (95%CI: 0.002-0.09) between HUI3 and EQ-5D.

Conclusion: In conclusion, the range of mean utilities in studies that compare the utility measures SF-6D, HUI3, and/or EQ-5D was small, even though they were performed across various diseases. The results demonstrated that mean utilities of the SF-6D and HUI3 did not differ systematically, whereas mean utilities of the EQ-5D were systematically lower.

POSTER SESSION - PUBLIC HEALTH; METHODOLOGICAL ADVANCES

REGRESSION ARTIFACTS IN LINEAR MAPPING HEALTH STATUS MEASURE TO UTILITY: EVIDENCE FROM MONTE CARLO SIMULATION

Yu A, Yu Y and Nichol M

University of Southern California, Los Angeles, CA

Purpose: Correct measurement of change in utility is paramount in cost-effectiveness analysis of health intervention. Health status measures, such as the SF-36 are ordinal scales with demonstrated ceiling and floor effects, whereas utility has presumed interval characteristics. Scaling inconsistencies may create regression artifacts (regression to the mean) and underestimate the magnitude of utility change. The purpose of this study was to estimate the extent of such regression artifacts.

Methods: Monte Carlo simulation methods were used to determine the presence of regression artifacts and estimate their extent under various conditions. Utility was assumed to be linearly associated with latent trait of each SF-36 domain, according to linear regression models (Nichol, 2001). In this study, continuous latent trait scale used population norm t-scores (mean = 50, sd = 10), which were then linked to discrete domain scores. Population heterogeneous distribution and SF-36 measurement error were realized by random numbers of normal or uniform distributions. Parameters of interest were baseline utility (from 0 to 1), utility change (0.01, 0.05, 0.1), and measurement error (corresponding reliability 100%, 99%, 90%).

Results: Changes in utility scores translated from SF-36 strictly underestimated the true utility change in every scenario examined. Even with 100% measurement reliability, underestimation was 22% for 0.01 utility change, approximately 14% for both 0.05 and 0.1 utility change. With increasing measurement error, underestimation increased. For example, at 90% measurement reliability level, when true utility changed 0.1, average estimated utility scores changed only 0.076, with 24.4% underestimation. In general, the regression artifacts presented the least threat at the 0.50 utility score, whereas underestimation was most severe at both ends, reinforcing the pronounced role of ceiling/floor effects.

Conclusion: Regression artifacts seemed to be inevitable when mapping HSM to utility based on linear regression algorithm, and may result in underestimation of utility change. Many factors appeared to be responsible for such artifacts including scaling inconsistencies, ceiling/floor effects, and imperfect measurement reliability.

POSTER SESSION - PUBLIC HEALTH; METHODOLOGICAL ADVANCES

CALIBRATION OF A NATURAL HISTORY MODEL OF CERVICAL CANCER USING LONGITUDINAL PRIMARY DATA

Kim J, Kuntz K, Goldhaber-Fiebert J, Weinstein M, Folse H, Sweet S and Goldie S

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 HG lesions was 5.97 and of LG lesions was 5.08 (6.25 and 5.39 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.

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

A TIME-DEPENDENT MARKOV MODEL TO EVALUATE THE COST EFFECTIVENESS OF LOSARTAN IN TREATING PATIENTS WITH HYPERTENSION AND LEFT VENTRICULAR HYPERTROPHY

Anis A[1], Sun H[2], Singh S[2], Woolcott J[3] and Nosyk B[2]

[1]University of British Columbia, Vancouver, BC, Canada; [2]St. Paul's Hospital, Vancouver, BC, Canada; [3]BC Centre for Disease Control, Vancouver, BC, Canada

Purpose: A time-dependant Markov model was developed to perform an incremental cost effectiveness analysis of losartan (Cozaar™) versus atentolol as a first line anti-hypertensive agent in the prevention of cardiovascular morbidity and mortality.

Methods: Using results from the Losartan Intervention for End Point Reduction in Hypertension (LIFE) trial as the source of treatment and effectiveness data, a Markov State Transition model was utilized to extrapolate the outcomes observed during the four-year trial to the patients? lifetime. Four irreversible health states (hypertension, MI, stroke, and death) were defined to calculate costs and quality adjusted life years (QALYs) over patients? projected lifetime. Costs and quality of life (QoL) estimates (utilities) for each of the given states were obtained from published studies The transition probability matrix for each cycle (cycle length 1 year) which allowed the transition probabilities to be time dependent and the state distribution at the end of each cycle were obtained. Moreover, the probability of patients newly transitioning into each state at each cycle was calculated. Similarly, costs were allowed to be time dependent. For example, costs incurred in the first cycle after a patient had a stroke were different from the cost incurred in the subsequent cycles. Costs associated with death due to MI or stroke were also considered. An incremental cost effectiveness ratio (ICER) for losartan versus atenolol, a first-line antihypertensive agent, assuming a discount rate of 3% for both cost and QoL were estimated. Extensive probabilistic sensitivity analyses were performed to examine the impact of a broad range of variation in our model parameters. A societal perspective was adopted. Result: The ICER of losartan versus atenolol in the treatment of patients commencing at age 67 with hypertension and left ventricular hypertrophy was CDN $1,337 (CND$1=US$0.75) per QALY gained. Probabilistic sensitivity analyses with 5000 simulations demonstrated a 50% probability that the ICER would be lower than $1,313 per QALY, a 34% probability that losartan was a dominant strategy (higher QALYs and lower cost) and a 95% probability that the ICER would be less than $18,884 per QALY.

Conclusion: Losartan appears to be an effective and cost-effective alternative to traditional first-line therapies for the treatment of hypertension. The Markov model developed was appropriate because it allowed for time dependent transition probabilities, QoL, and costs.

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

IMPACT OF SECONDARY CARDIOVASCULAR EVENTS ON HEALTH STATUS

Van Stel H[1], Busschbach J[2], Hunink M[2] and Buskens E[3]

[1]UMC Utrecht, Utrecht, Netherlands; [2]Erasmus Medical Center, Rotterdam, Netherlands; [3]University 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 (EQ5D, range 1 to -0.6) and secondary cardiovascular events (death, myocardial infarction (MI), cerebral infarction, 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 = 1405), 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 S-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), cerebral infarction (n = 60, impact score -0.05, p = 0.02) and extracranial bleeding (n = 25, impact score -0.05, p = 0.01), but not for amputation (n = 15). Patients who get an event, except for MI, start substantially (but not significantly, p between 0.15 and 0.22) lower than patients without secondary events. Possible heterogeneity was ignored as the heterogeneity-tests F-max and ICC were negative.

CONCLUSIONS: Subsequent cardiovascular events have significant impact in terms of utility. Impact of secondary events is in the same 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 2002B45.

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

Basu A[1], Meltzer H[2] and Dukic V[1]

[1]University of Chicago, Chicago, IL; [2]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 the negative domains of the disease but also accounts for the severities within each domain. Item-specific scores from the Positive and Negative Syndrome Scale (PANSS) are used to define four severity states 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 of drug treatment. However, haloperidol turns out to be the best in controlling severe positive symptoms for the next two quarters of treatment. Olanzapine appears to be the best drug to control severe negative symptoms across all four quarters of treatment while haloperidol is the worst in this regard. Finally, all three drugs are able to prevent deterioration of negative symptoms as long as they can effectively control the positive symptoms.

Conclusions: Various fine levels of detail on the transition probabilities may serve to estimate quality of life of schizophrenic patients and resource utilizations in this field more accurately. Bayesian synthesis of evidence also reveals considerable uncertainty in the time-profile parameters 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

THE IMPACT OF FAILURE TO CALIBRATE ON RESULTS IN ECONOMIC EVALUATION

Coyle D

Ottawa Health Research Institute, Ottawa, ON, Canada

Purpose: To demonstrate who failure to calibrate economic models can lead to biased estimates of the cost effectiveness of health interventions.

Methods: The Canadian Economic Model of Osteoporosis has been used in several previous evaluations of osteoporosis treatments. It is fully calibrated in that the model replicates population data for the risk of fracture and mortality. Many previous models of osteoporosis are not calibrated: e.g. the population risk of fractures are weighted by the relative risks of fracture with osteoporosis or previous fracture history. This will lead to an overestimation of the risk of fracture for such groups. Analysis assessed the cost effectiveness of alendronate compared to no therapy for a 75-year-old women with previous fracture history with and without calibration for fracture risk and mortality. Analysis identified the optimal age at which treatment with alendronate becomes cost effective.

Results: For a 75-year-old women with previous fracture, the annual probability of fracture without therapy is 1.31% with calibration and 1.58% without calibration. The incremental cost per QALY gained (ICUR) of alendronate is $35 600 when calibrating the model with respect to fracture risks and mortality and $19 400 without. Assuming a QALY was worth $50000, it would be cost effective to treat with alendronate women with previous fracture who were aged 75 and over when the model was calibrated. Without calibration, alendronate would be cost effective for women aged 73 and over.

Conclusions: Although recommended, decision models used for economic analysis are often not calibrated to replicate population data. Failure to calibrate models can lead to substantially different estimates of an ICUR and can lead to differences in policy recommendations. Studies based on decision models need to report what means of calibration were undertaken.

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

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

Delaney B[1], Ford A[2], Qume M[1] and Moayyedi P[3]

[1]The University of Birmingham, Birmingham, United Kingdom; [2]Leeds Royal Infirmary, Leeds, United Kingdom; [3]McMaster University, Hamilton, ON, Canada

PURPOSE: Cost-effectiveness RCTs, where the principal aim is to determine both the difference in effect and the resource costs of alternative treatment strategies, are increasingly 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.

Dyspepsia is a common condition in primary care, for which there are a number of investigative and empirical therapeutic options. A collaborative group has been meeting intermittently since 1996, prospectively registering RCTs with the aim of determining the cost-effectiveness of *Helicobacter pylori* 'test and treat' compared to endoscopy- based management and empirical acid-suppression.

METHODS: Individual patient data meta-analysis pooling effects of endoscopy, acid-suppression and *H. pylori* 'test & treat' on dyspepsia symptoms and costs. Standardized unit costs (UK national standard costs 2001) were applied to resource utilization at patient level and net benefit calculated at patient level using INB = $\lambda \Delta E - \Delta C$. Effect, and net benefit were pooled at study level and a sensitivity analysis on the ceiling ratio (λ).

RESULTS: Five trials that compared *H. pylori* 'test and treat' with endoscopy were pooled. Although a small effect difference was found in favour of endoscopy, Relative Risk of dyspepsia recurrence = 0.95, 95% confidence interval (CI) 0.92 to 0.99), this was not cost-effective. Even at a λ of $1750 per patient 'cured' INB remained negative at -$282 (-$324to -$240). 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 -$156(-249 to -116) at $\lambda = 0$, and was insensitive to increasing λ.

CONCLUSIONS: Endoscopy is not cost-effective compared with *H. pylori* test and treat, or empirical acid suppression. IPD meta-analysis can be used to unpick complex RCTs with both cost and effect data. Using incremental net benefit as the variable to pool maintains the cost-effect correlation at patient level. INB is normally distributed, and can be summarised across studies as a continuous variable.

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

BAYESIAN POSTERIOR DISTRIBUTIONS FOR PROBABILISTIC SENSITIVITY ANALYSIS

Hazen 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 Ades, Lu and Claxton (2004). We outline procedures for using the resulting posterior distributions in Monte Carlo simulation.

Results and Conclusions: We apply these methods to conduct a PSA for a recently published analysis of zidovudine prophylaxis following rapid HIV testing in labor to prevent vertical HIV transmission in pregnant women (Mrus and Tsevat 2004). Zidovudine prophylaxis is cost saving and has net benefit $557 per pregnancy compared to not testing for HIV, assuming a cost of $50,000 per lost QALY (mother and child). We based a PSA on the following data from seven studies of vertical HIV transmission, as well as data for 5 other probability parameters: Given this data, the two parameters (log Risk population mean) and (log Risk Ratio population mean) for vertical HIV transmission have approximate bivariate normal posterior with mean/sd equal to -1.39/0.12 and -1.02/0.23, and correlation -0.108. Using these and other posterior distributions for all 5 remaining probabilities in a PSA yields zidovudine prophylaxis optimal 94.3% (0.13%) of the time, and the expected value of perfect information on all 7 relative effects and probabilities equal to $15.13 (1.35) per pregnancy. These results concur with Mrus and Tsevat's conclusion that the choice of rapid HIV testing followed by zidovudine prophylaxis is not a close call.

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

PROBABILISTIC ANALYSIS AND COMPUTATIONALLY EXPENSIVE MODELS: NECESSARY AND REQUIRED?

Griffin S, Claxton K, Hawkins N and Sculpher M

University of York, York, United Kingdom

Purpose: To demonstrate the necessity of characterising uncertainty in second-order non-linear models with computationally expensive model structures and patient-level simulation.

Methods: The recent methodological guidance on the assessment of health technologies issued by the National Institute of Clinical Excellence (NICE) requires probabilistic sensitivity analysis (PSA) as the appropriate way to quantify decision uncertainty. This requirement may be computationally expensive in more complex models structures, particularly those that employ patient-level simulation. However, where models are second-order non-linear, accounting for second-order uncertainty is required to estimate expected costs and effects as well as decision uncertainty. A review of published NICE technology appraisals identified all cases in which the economic evaluation was undertaken using patient-level simulation. Each case was examined to see if the model structure implied second-order non-linearity, whether PSA was performed, and, where this was not the case, whether alternative modelling approaches were available such as less computationally expensive model structures or the use of emulators.

Results: All of the model structures examined implied second-order non-linearity. The majority of cases did not attempt to account for second-order uncertainty by performing PSA. The reasons cited included the computational expense of using patient-level simulation, and the complexity of the model structure exceeding current availability of data. However, the possibility of implementing less computationally expensive models was demonstrated in two areas: i) two appraisals examined pharmacotherapy for epilepsy, one employed patient-level simulation without PSA due to computational expense, the other used an equivalent semi-Markov model structure where PSA was feasible; ii) one appraisal of treatments for osteoporosis used emulators (Gaussian processes) to address both first- and second-order uncertainty in a complex model structure employing patient-level simulation. In a number of other cases, the use of computational expensive methods was due to modelling treatment switching, rather than evaluating the full range of possible clinical strategies, which may have been more appropriate.

Conclusions: Where model structures exhibit second-order non-linearity, incorporating second-order uncertainty is required for unbiased estimates of expected cost and effect, as well as to characterise decision uncertainty. Given that PSA is necessary and required, the computational expense of alternative model structures should be considered. Where complex and expensive model structures are required further work is needed to identify suitable computational methods and emulators.

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

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

Kongnakorn T[1], Swan J[2] and Sainfort F[1]

[1]Georgia Institute of Technology, Atlanta, GA; [2]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 noxious 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 health 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 model by exploring key variables with a sample of healthy subjects who were given scenarios involving testing for atherosclerotic carotid vascular disease with x-ray angiography (XRA) 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 8.75 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

MORBIDITY COST IN COST-EFFECTIVENESS ANALYSIS

Myers J[1], McCabe S[2] and Gohmann S[3]

[1]Yale University, New Haven, CT; [2]Christine M. Kleinert Institute for Hand and Microsurgery, Louisville, KY; [3]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 argue that morbidity costs are intrinsic 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 difference between illness severity levels in mean QOL ($p < 0.0001$, $F = 29.5$, $df = 2, 178$). No statistically significant interaction between level of illness severity and prior knowledge was observed ($p = 0.5904$, $F = 0.53$, $df = 2, 178$). Therefore, we fit a model removing the interaction term.

Conclusion: The current study demonstrated that those subjects informed of morbidity costs score quality of life lower than subjects uninformed 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

PATIENT ADHERENCE: A BLIND SPOT IN COST-UTILITY ANALYSES?

Rosen A[1], Stone P[2], Greenberg D[1] and Neumann P[1]

[1]Harvard School of Public Health, Boston, MA; [2]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 overstate the cost-effectiveness of interventions, we also examined the relationship between pharamaceutical 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 = 0.65). 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 40% of HIV antiretroviral studies. Among the 70 CUAs in which study sponsorship was disclosed, 40% of pharmaceutical sponsored studies (n = 40) and 50% of non-pharmaceutical sponsored studies (n = 30) included adherence (difference 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.

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

DECISION ANALYSIS OF PROSTATE CANCER WITH VARIABLE LEVELS OF DISEASE AGGRESSIVENESS

Neeley E[1], Cantor S[2] and Cox D[1]

[1]Rice University, Houston, TX; [2]The University of Texas M. D. Anderson Cancer Center, Houston, TX

Purpose: Clinicians recognize that there seem to be variable levels of aggressiveness in many cancers. We performed a decision analysis of the effectiveness of prostate cancer screening with the assumption that there were two levels of aggressiveness: one high and one low.

Methods: We modified the prostate cancer screening model of Cantor SB, et al. (J Fam Pract 1995; 41:33-41). We considered annual screening beginning at age 50, using digital rectal exam (DRE) and Prostate Specific Antigen (PSA) tests.

The sensitivity and specificity of the screening tests were held the same, but we allowed the transition rates from the initial and subsequent stages of cancer to vary in two subgroups: one slowly developing and one rapidly developing. The proportion of rapidly developing cancers was set at 25%, 50%, and 75% and the transition probabilities were recomputed so that (1) the mean time a slow progressing disease spent in a state was twice the mean time of the fast progressing disease, and (2) the average transition probabilities remained the same. We computed effectiveness using the same utilities for health states as in the original paper.

Results: The original analysis, based on a single level of disease progression, indicated that screening resulted in decreased effectiveness by 0.67 quality-adjusted life years (QALYs). In the new analysis with two levels of disease aggressiveness, we found that if 25% of the cancers were fast progressing, then no screening is still preferred, but by only 0.24 QALYs. With 50% and 75% of the cancers fast progressing, screening has slightly higher effectiveness by 0.12 and 0.42 QALYs, respectively.

Conclusions: When more detailed modeling was incorporated in the prostate cancer decision analysis, the net benefits of screening were enhanced. Our analysis showed that relaxing the assumption of a single set of probabilities for transitions between health states can change the optimal recommendations. Often a decision analysis relies on many assumptions about the disease or other conditions which may or may not be true. Our analysis highlights one assumption that can be critical to the outcome of the analysis.

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

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

Bayoumi A[1], Sanders G[2], Munakata J[3] and Owens D[4]

[1]St. Michael's Hospital, University of Toronto, Toronto, ON, Canada; [2]Duke, Durham, NC; [3]VA Palo Alto Health Care System, Menlo Park, CA; [4]VA Palo Alto Health Care System, Palo Alto, CA

The optimal management of multi-drug experienced HIV infected patients is unknown. We evaluated the potential cost effectiveness of aggressive treatment for this group with multiple antiretroviral drugs or novel therapies that are expensive or toxic. This approach may enhance the probability of achieving virologic suppression but could increase costs and decrease quality of life. We developed a Markov analysis model of HIV infection to evaluate standard and aggressive therapy. The model follows individuals from the time antiretroviral therapy is initiated until death. The base case had a viral load of 40,000 copies/mL and a CD4 count of 350 cells/mm3. We assumed that individuals would change regimens after intolerance, virologic rebound, or clinical disease progression and used third line regimen. We modeled standard therapy as a 4-drug fourth line regimen and aggressive therapy as fourth line regimen a combination of at least 4 drugs but with increased costs. Patients intolerant of aggressive therapy would step down to standard therapy. Subsequently, patients started therapy which diminishes, but does not suppress, viral load levels. We modeled the effects of aggressive therapy on the enhanced probability (odds ratio) of virologic suppression and incremental costs, with base case values of 3 and $15,000, respectively. In sensitivity analyses, we increased the risk of drug-limiting intolerance to 50%, and decreased quality of life (QOL) by 10% for aggressive therapy. Aggressive therapy was associated with incremental survival of 6.4 months, discounted quality-adjusted survival of 5.3 months, and a cost effectiveness ratio of $75,500 per quality adjusted life year (QALY), if drug tolerance and QOL were similar to standard therapy. The most important determinants of cost effectiveness were the efficacy and cost of aggressive therapy; incremental costs of aggressive therapy would have to be at most $6000, $8000, or $9000 if the odds ratio for viral suppression was 2, 3, or 4 compared to standard therapy at a cost effectiveness threshold of $50,000/QALY. Incorporating intolerance and quality of life effects, the maximal incremental cost was $4000. Our model suggests that aggressive HIV therapy may be cost effective if adverse effects are minimal. Ongoing trials may characterize the clinical parameters necessary for cost effectiveness, but antiretroviral therapy prices may have to fall considerably for routine use of aggressive therapy to be economically attractive.

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

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

Bialy T[1], Washington C[1], Szeto H[2], Whalen J[3] and Chen S[4]

[1]Emory University School of Medicine, Atlanta, GA; [2]Kaiser Permanente, Redwood City, CA; [3]University of Connecticut, Farmington, CT; [4]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 1,000,000 new cases per year. More than 65% of all BCCs and SCCs affect facial sites with both cosmetic and functional sequela. 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 98 consecutive patients with primary facial NMSC was used to obtain baseline cost (Connecticut Medicare 2002 reimbursements) and efficacy (margin 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 using 5-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 ($956.60 vs. $1248.10, and 0.6 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 (granulation, 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

AGGRESSIVE LIPID LOWERING THERAPY IS A COST-EFFECTIVE INTERVENTION FOR SECONDARY PREVENTION IN POST-MYOCARDIAL INFARCTION PATIENTS

Dalal M[1], Karne A[1] and Natarajan S[2]

[1]NYU School of Medicine, New York, NY; [2]NYU School of Medicine and the Manhattan VA Medical Center, New York, NY

BACKGROUND: Recent data indicate that more intense lipid lowering reduces recurrent cardiovascular events. We evaluated the cost-effectiveness of such aggressive lipid lowering with 80 mg atorvastatin versus usual care with 20 mg simvastatin for secondary prevention of recurrent cardiovascular events.

METHODS: The study is a cost-effectiveness analysis by decision analytic modeling using the reference-case approach. Primary outcomes were cardiovascular death, MI, unstable angina requiring hospitalization and/or revascularization measured 2 years from randomization. Effectiveness data were from the Pravastatin or Atorvastatin Evaluation and Infection Therapy - Thrombolysis in Myocardial Infarction 22 study that enrolled patients within 10 days of acute MI in stable condition. The relative risk reduction in the primary outcomes was 16% (95% confidence interval: 5%-26%) with aggressive therapy. We assumed that usual care with 20 mg simvastatin is equivalent to 40 mg pravastatin. Life-years were converted into quality-adjusted life-years (QALYs) using established methods. We used published cardiovascular disease cost data and drug costs from the Veterans Health Administration System. All costs are in 2002 US dollars.

RESULTS: The reference-case analysis yields an incremental cost-effectiveness ratio (ICER) for aggressive therapy with atorvastatin of $64,952 per QALY gained. Raising the price of atorvastatin above $1800 for two years of therapy (115% of the reference-case price) makes the ICER >$100,000/QALY. While aggressive therapy is not cost effective below a relative risk reduction of 12%, it is highly cost-effective at higher estimates. The results are not sensitive to change in compliance rates. When the cost of treating coronary heart disease is varied from 50% to 150% of the reference-case, aggressive therapy is not cost effective at price ranges below 70%. Varying the discount rate from 0% to 7.5% yields results consistently in favor of aggressive therapy.

CONCLUSIONS: A strategy of aggressive lipid lowering provides more QALYs at a modest cost. Except when the costs of treating coronary heart disease are low or if the relative risk reduction by aggressive therapy is small, aggressive lipid lowering therapy is substantially better.

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

BIOPSY OR RESECTION FOR SINGLE SMALL LIVER NODULES IN PATIENTS WITH COMPENSATED CIRRHOSIS- A DECISION ANALYSIS

Bremner K[1], Bayoumi A[2], Sherman M[1] and Krahn M[1]

[1]University Health Network, Toronto, ON, Canada; [2]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 favoured initial hepatic resection, but the gain was small (4.72 quality-adjusted life months, or 5.49 life months). The model was robust to most assumptions; the decision was not sensitive to probabilites 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 contraindications to resection prior to accurate HCC diagnosis by imaging in follow-up screening.

Conclusions: Our model suggests that resection of all suspicious, 1-2 cm liver nodules offers longer 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 EGFR 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 life-years saved (LYs) of three strategies: 1) no test, chemotherapy alone for all patients; 2) test with IHC, cetuximab plus chemotherapy for EGFR-expressers; 3) 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 testing, the addition of cetuximab to standard chemotherapy for all patients increased treatment costs by approximately $66,000, and extended average survival by about 1.2 months, yielding an incremental cost-effectiveness ratio (ICER) of more than $680,000 per LYs. 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 $600,000 per LYs, and the ICER of universal cetuximab therapy was more than $1.5 million per LYs.

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

Janssen M and Buskens E

University Medical Center Utrecht, Utrecht, Netherlands

PURPOSE: Stenting of the arteria carotis interna (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 clearly 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 Ευρο Σανσ∀1.500 higher than those of CEA. Net Health Benefits (at Ευρο Σανσ∀25.000 per QALY) per percent decrease in complications for CAS were estimated at Ευρο Σανσ∀2.400, Ευρο Σανσ∀1.600 and Ευρο Σανσ∀720 for the rates of peri-operative complications "major stroke," "minor stroke" and "death" respectively. For long-term postoperative major and minor stroke rates these figures were Ευρο Σανσ∀9.300 and Ευρο Σανσ∀3.000. The global EVPI was estimated at Ευρο Σανσ∀2400 and the EVPI for peri-operative and long-term complication rates were Ευρο Σανσ∀1600 and Ευρο Σανσ∀1700 respectively. *CONCLUSIONS:* At a global EVPI of Ευρο Σανσ∀2400 an annual intervention rate of approximately 1200 interventions implies a total value of Ευρο Σανσ∀3.000.000 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 Ευρο Σανσ∀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 trail with long term follow-up.

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

COST-EFFECTIVENESS ANALYSIS OF THROMBOLYTIC TREATMENT FOR STROKE

Mar J[1], Begiristain J[2] and Arrazola A[2]

[1]Alto Deba Hospital, Mondragon, Spain; [2]Basque Government, San Sebastian, Spain

Purpose: The treatment of acute stroke has generally been limited to the maintenance of vital functions and the avoidance of complications. Thrombolysis, prompted on the success of in the treatment of the acute myocardial infarction, has also been applied to stroke. The assessment of the NINDS trial a year after the event showed that the intravenous r-TPA decreased the disability by 30% without significant changes in mortality. This article presents a cost-effectiveness analysis, based upon a probabilistic model, of the use of thrombolytic therapy in stroke treatment.

Methods: We carried out a survey with stroke patients during the hospital stay and a year after release from the hospital to obtain data on costs and natural history. We then calculated utility weights, using Euroquol-5D. Efficacy data were obtained from the NINDS trial. Every time the model runs, 4,000 Monte Carlo simulations are undertaken in which each parameter value changes depending on its probability distribution. The incremental cost-effectiveness ratio (ICER) is calculated for each simulation based on the incremental cost and incremental effectiveness. The results are expressed in terms of the cost-effectiveness plane, the CE acceptability curve and the Expected Value of Perfect Information (EVPI).

Results: We studied 435 patients from whom 304 were ischemic strokes. One year later, 216 were alive and 78 were died. The mean utility values were 0.22 in disabled and 0.77 in autonomous patients. The ICER obtained with the means of the parameters is -19.000Ευρο Σανσ∀/AVAC reflecting a saving of 6.000Ευρο Σανσ∀ and a health benefit for patients. The cost-effectiveness plane shows that thrombolysis is dominant in 96.1% of simulations. In the acceptability curves, only a 0.4 of simulations obtain an ICER higher than the societal threshold of 30,000 euros/QALY. The EVPI analysis showed that the total EVPI was low and that the utility values and the costs related with the disability were the parameters which influenced the results uncertainty.

Conclusion: Thrombolytic therapy is judged an intervention to implement because saves cost and wins effectiveness. The key point is the decrease rate in patients disability what produces better quality of life and less costs. The uncertainty of the results was very small.

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

A PROBABILISTIC COST-EFFECTIVENESS ANALYSIS OF ENOXAPARIN VS. UNFRACTIONATED HEPARIN FOR DVT PROPHYLAXIS FOLLOWING MAJOR TRAUMA

Lynd L, Goeree R and O'Brien B

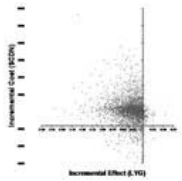
McMaster University, Hamilton, ON, Canada

Objective: 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. Outcomes data were extracted from the only published clinical trail (NEJM 1996; 335 (10):701-7). 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 case-costing 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 DVTs 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 DVTs averted model, 98% of the model iterations fell in the NE and SE 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 λ=$70,000 per life year, there was only a 5% probability that enoxaparin was cost effective.

Conclusions: Only one previous study has evaluated the cost-effectiveness (cost/DVT averted) 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 benefits 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

Resch S[1], Salomon J[2], Murray M[3] and Weinstein M[3]

[1]Harvard University, Boston, MA; [2]Harvard School of Public Health, Cambridge, MA; [3]Harvard 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 is not adequate for treating MDR-TB. Second-line drugs can cure MDR-TB and 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 ineffectively-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 standardized 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 indices 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 $20,500 per death averted, respectively.

Conclusions: Treating MDR-TB after first-line failure improves health outcomes at an affordable cost (less than $100 per QALY gained) even in low-income countries. Ignoring the effect of treatment 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.

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

IMPROVING PATIENT-TO-NURSE RATIOS AS A COST-EFFECTIVE SAFETY INTERVENTION

Rothberg M[1], Abraham I[2], Lindenauer P[1] and Rose D[1]

[1]Baystate Medical Center, Springfield, MA; [2]University of Pennsylvania School of Nursing, Philadelphia, PA

Purpose A growing body of research has shown a link between nurse staffing and patient outcomes. As a result, 14 states have introduced legislation to limit patient-to-nurse ratios, placing a considerable financial burden on hospitals. Although lower ratios could decrease length-of-stay and nursing turnover, offsetting some of the additional labor costs, mandatory ratios could also drive up nursing wages in an already tight market. Our objective was to determine the marginal cost-effectiveness of various nurse staffing ratios.

Methods Adopting the societal perspective, we conducted a cost-effectiveness analysis for general medical and surgical patients comparing patient-to-nurse ratios ranging from 8:1 to 4:1. Cost estimates were drawn from national databases. The effects of nurse staffing on patient mortality and length of stay were based on two large hospital level studies. Outcomes were measured in cost per life saved. The potential contributions of nursing turnover and elasticity of labor supply were addressed in sensitivity analysis.

Results In the base case analysis, 8 patients per nurse was the least expensive ratio, but was associated with the highest patient mortality. Decreasing the number of patients per nurse improved mortality and increased costs, becoming progressively less cost-effective as the ratio declined from 8:1 to 4:1. Nonetheless, the marginal cost-effectiveness did not exceed $85,000 (95% C.I. $42,000 to $260,000) per life saved. Lower ratios generated savings from shorter lengths of stay, but these offset less than half of the increase in labor costs. The model was most sensitive to variations in mortality rates associated with different patient-to-nurse ratios. Several other factors made low ratios more cost-effective: a) low hourly wages, b) high labor elasticity, c) short length of stay, and d) registered nurse-hours decreasing length of stay. However, throughout the ranges of all these variables, the marginal cost-effectiveness of limiting the ratio to 4:1 never exceeded $280,000 per life saved.

Conclusions As a patient safety intervention, patient-to-nurse ratios of 4:1 are reasonably cost-effective and in the range of other commonly accepted interventions. More accurate estimates of the effect of nurse staffing on patient mortality are needed.

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

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

Sanders G[1], Hlatky M[2] and Owens D[2]

[1]Duke, Durham, NC; [2]Stanford 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-model 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 QALYs) 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 31.2%. If the cost of the ICD device were reduced from $25,000 to $10,000, the incremental cost effectiveness of the ICD relative to conventional therapy would improve from $50,900 to $33,500/QALY gained. If the ICD improved quality of life, the cost effectiveness is more favorable than 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 in patients 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 large 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

MEASUREMENT OF THE VALUE OF EXERCISE: A COST-EFFECTIVENESS ANALYSIS OF PROMOTING PHYSICAL ACTIVITY AMONG ADULTS

Roux L[1], Pratt M[2], Yanagawa T[2], Yore M[3] and Tengs T[4]

[1]Centers for Disease Control and Prevention, Vancouver, BC, Canada; [2]Centers for Disease Control and Prevention, Atlanta, GA; [3]Centers for Disease Control and Prevention, Cantonsville, MD; [4]Milliman USA, Denver, CO

Purpose: Our objective was to assess the cost-effectiveness of population-wide strategies to promote physical activity in adults.

Methods: We developed a novel and comprehensive state-transition Markov model to estimate the costs, health gains (QALYs), and cost-effectiveness of 4 alternate public health strategies to promote physical activity. To identify strategies, we selected those that were "strongly recommended" by the US Task Force for Preventive Services. Interventions exemplifying each of 4 strategies were evaluated. A community-wide campaign strategy was represented by a multi-factorial and multimedia-dependent health education intervention. An intervention emphasizing the use of personal trainers and financial incentives exemplified an individually-adapted health behavior change strategy. A social support strategy was represented by an intervention that incorporated organized walking groups, social gatherings, phone calls, and home visits. Finally, a strategy of enhanced access was characterized by an intervention that exposed an entire community to an environment conducive to an active lifestyle (e.g., new bicycle paths, fitness facility hour extension). Each intervention was compared to a no intervention alternative. Efficacy estimates were obtained from randomized controlled trials. A systematic review of disease burden by exercise status was used to assess the relative risk of 5 diseases (coronary heart disease, ischemic stroke, colorectal cancer, breast cancer, and type 2 diabetes) for each of the following physical activity levels: 1) inactive; 2) irregularly active, sufficiently active to minimally meet public health recommendations; and, 4) highly active. Quality of life data by disease state, exercise level, age, and gender were obtained using the Quality of Well Being Scale. Longitudinal medical costs for the disease states were gathered from a 400,000 member claims database and annualized using actuarial methods. Costs and QALYs were assessed from a societal perspective over 10, 20, 30, and 40 year time horizons and discounted back to the present at 3%.

Results: While the most effective strategy focused on enhancing access to physical activity, social support was the most cost-effective strategy at $6400 per QALY, assuming a 40-year time horizon. Enhanced access cost $34,000/QALY, individually adapted $73,000/QALY, and community campaign $110,000/QALY. Results were sensitive to intervention-related costs and efficacy.

Conclusion: For adults, social support offered the best value for money. However, compared with other well-accepted preventive strategies, all physical activity promotion strategies evaluated offered good value for money.

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 U[1], Sroczynski G[1], Aidelsburger P[2], Rossol S[3], Wasem J[2], Manns M[4], McHutchison J[5] and Wong J[6]

[1]Massachusetts General Hospital, Harvard Medical School, Boston, MA; [2]University of Duisburg-Essen, Essen, Germany; [3]University of Heidelberg, Rüsselsheim, Germany; [4]Medical School of Hanover, Hannover, Germany; [5]Duke University Medical Center, Durham, NC; [6]Tufts-New England Medical Center, BOSTON, MA

Purpose: The recently developed German guidelines for antiviral treatment (AVT) in patients with chronic hepatitis C (CHC) recommend basing drug dosage, intended treatment duration, and early stopping rules on the genotype 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 (GEHMO), 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 (IFN), (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 (GUIDE). 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 GUIDE) reduced the 20-year risk for decompensated cirrhosis, hepatocellular carcinoma, liver transplantation, and liver-related death by more than 50%. Compared to NoAVT, PEG increased life expectancy by 5.0 life years and GUIDE increased life expectancy by 4.9 years. GUIDE dominated IFN by strong dominance. Compared to NoAVT, discounted ICERs were 1500 EUR/QALY for GUIDE and 3300 EUR/QALY for PEG. GUIDE saved 3950 EUR per patient. Moving from GUIDE 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 efficiently 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[1], Cohen D[1] and Farley T[2]

[1]RAND, Santa Monica, CA; [2]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 decisions 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 of importance. An optimization program can be built in 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 estimation 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 in STD clinics and raising 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

Mahajan 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 reconsiders 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 410 cases to 1409 controls. The model included patient demographics, disease severity indicators, and clinically-plausible interactions. We used the SAS macro '%match' 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 matches between greedy and optimal matching. Bias ratios >1 indicate 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 increase.

Table Absolute Differences in Propensity Scores

Number of matched pairs	Absolute Difference in Propensity Scores (x 10^{-5}) Greedy Matching (A)	Optimal Matching (B)	Bias Ratio (A/B)
350	24.38	22.72	1.07
300	12.86	12.24	1.05
250	8.36	8.16	1.02
200	5.66	5.56	1.02
150	3.61	3.58	1.01
100	2.01	1.99	1.01
50	0.90	0.90	1.00

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 controls. Hence as the number of matched pairs increase, and competition for controls increases, the greedy match 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 domains) 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 HUI2 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 MCMC and propensity score method with a missing at random assumption. In order to relax the assumption of linear effect SF-36 subscales on utility, we used restricted cubic spline functions (CSF) with four or five knots for each subscale except for RP, 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=9.34, d.f.=3, p<.0001), VT (F=3.50, d.f.=2, p=.0302), MH (F=11.14, d.f.=3, p<.0001), BP (F=42.81, d.f.=2, p<.0001). Total nonlinearity was significant (F=11.76, d.f.=16, p<.0001). 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 0.499 for simple linear model to 0.518 (total R2 from 0.500 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.

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

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

Bendeck S, Hadley J, Bonaccorsi P, Brown K, Lawson D, Murray D, Washington C and Chen S

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 (< one year ago) or old (> one 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	N
I	0.93 (0.099)			14
		1.0 (0)	New Stage II	3
		0.95 (0.074)	Old Stage II	8
		0.90 (0.088)	New Stage III	7
		0.95 (0.088)	Old Stage III	8
		0.94 (0.080)	Combined stages	26
II	0.97 (0.058)			3
		0.97 (0.060)	New Stage I	3
		0.71 (0.24)	Old Stage I	30
		0.73 (0.24)	Combined stages	33
III	0.52 (0.31)			7
		0.54 (0.31)	New Stage I	10
		0.49 (0.30)	Old Stage I	32
		0.50 (0.30)	Combined stages	42

No statistically significant differences were found between real and hypothetical utilities for patients in each individual melanoma health state. When we combined all the patients in different melanoma health states predicting the same state together, the projected utility was not significantly different from those patients who actually had the health state.

Conclusions: Although the number in this preliminary study is small, these results suggest that melanoma patients may be able to realistically imagine alternate melanoma stages.

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

DEFINING EPISODIC COSTS OF INJURIES: THE CASE OF SUICIDE ATTEMPTS

Corso P and Haileyesus T

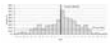
Centers for Disease Control and Prevention, Atlanta, GA

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 $6,023. 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 $p > 0.10$; 95% CI: $148 - $308, mean = $228). Total episodic costs of suicide-related injuries ($8,855), 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

OUTCOMES OF A MULTIDISCIPLINARY PHYSICIAN-LED PRACTICE MANAGEMENT INTERVENTION TO CONTROL COSTS DURING AN ERA OF RAPID TECHNOLOGIC EVOLUTION IN CORONARY INTERVENTIONS

Kamath C, Rihal C, Reller K, Anderson S, McMurtry E, Bresnahan J, Holmes D and Long K

Mayo Clinic College of Medicine, Rochester, MN

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 (66 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 (31% vs. 27%), glycoprotein after PCI (34% vs. 30%), hypertension (64 % vs. 59%), moderate/severe bend in any lesion (48% 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 0.90 days (3.4 vs 2.5; $p<0.001$) and predicted cost savings of $4,667 (95% confidence interval of difference: $4218, $5116).

Conclusion: Physician-led practice management efforts were successful at containing PCI related costs of care 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.

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

PREDICTORS OF USE OF HANDHELD DECISION SUPPORT TOOLS IN THE CLINICAL SETTING

Yu F, Houston T, Ray M and Berner E

University of Alabama at Birmingham, Birmingham, AL

PURPOSE: To identify factors associated with the use of handheld clinical decision support tools by Internal Medicine providers in clinical settings.

METHODS: 82 internal medicine residents of an urban teaching hospital were given personal digital assistants (PDAs), each containing a suite of clinical decision support programs, which include MedMath, MedCalc, and ePocrates, among others. Descriptive data including age, sex, race, years in residency training, previous ownership with PDA and perceived barriers to PDAs were collected on baseline surveys. A tracking program was used to prospectively track the number of times a program was entered. The program was considered to be used by the resident, if the program was accessed 3 or more times during the follow-up period. Our outcome, "breadth of use," was defined as the number of programs used by a physician during follow-up (an ordinal variable ranging from 0 to 6). Multivariate analyses were conducted using linear regression.

RESULTS: 68 physicians were included in the study. Thirty-two percent were postgraduate year one (PGY1), 38% PGY2, and 29% PGY3. Most residents were white (76%), male (75%), and over half had previously owned a PDA (59%). Data were collected at least once (mean data collection interval=5.6 months). Nineteen percent of the residents did not use any of the installed programs. Among those who used the programs, ePocrates was accessed more often than the others. Mean breadth of use was 2.6 programs (SD=2.0, range=0 to 6). Mean use declined by residency year [PGY1=3.4 (SD=1.8); PGY2=2.6 (2.2) and PGY3=1.8 (1.8)]. Univariate analysis showed that higher levels of residency training were associated with a decrease in breadth of use of handheld clinical decision support tools. Independent effects were not observed during multivariable analysis due to a limited sample size.

CONCLUSION: Our data show that increasing resident years is associated with using a smaller range of handheld decision support tools in the clinical setting. Further study is needed to identify barriers to using handheld decision support tools and develop effective strategies to overcome those barriers.

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

RACIAL DISPARITIES IN THE PRESCRIBING OF WARFARIN FOR NONVALVULAR ATRIAL FIBRILLATION

Schauer D[1], Johnston J[2], Wess M[1], Moomaw C[1] and Eckman M[1]

[1]University of Cincinnati, Cincinnati, OH; [2]Eli Lilly, Indianapolis, IN

PURPOSE: Warfarin has been shown to decrease the rate of thromboembolic events in patients with nonvalvular atrial fibrillation (AF) but is frequently under-prescribed. Our goal was to establish whether there are racial disparities in the prescribing of warfarin for patients with newly incident nonvalvular AF in the Ohio Medicaid population.

METHODS: A retrospective cohort of Ohio Medicaid recipients with newly incident nonvalvular AF between January 1, 1998, and May 31, 2002, was identified. Patients were included if they had at least one year of continuous enrollment in Medicaid prior to the diagnosis of atrial fibrillation and two or more International Classification of Diseases, Ninth Revision, Clinical Modification codes for atrial fibrillation. Exclusions included valvular heart disease and warfarin prescriptions prior to the diagnosis of atrial fibrillation. Race was identified from the demographic information in the database, and the analysis was limited to White and African-American patients. Possible confounders included age, sex, history of hypertension, diabetes mellitus, congestive heart failure, renal disease, liver disease, previous stroke, previous bleeding, risk factors for non-adherence and increased risk of falling. Univariate logistic regression was used to evaluate the unadjusted association of potential confounders with warfarin prescribing. To evaluate the independent role of race in warfarin prescribing, we created a multivariable logistic regression model incorporating all predictors significant at $p<.10$ in univariate models.

RESULTS: 6,283 patients were identified as having newly incident nonvalvular AF, 18.5% of which were African-Americans. In general, African-American patients had a higher rate of comorbid illness, with significantly higher rates of risk factors for both stroke (i.e., hypertension, diabetes and congestive heart failure) and bleeding (i.e., renal disease and prior bleeding). 9.4% of White patients and 7.7% of African-American patients were prescribed warfarin. In the univariate analysis African-Americans had an odds ratio of 0.80 (95% CI 0.63, 1.01) for receiving warfarin when compared to White patients. When controlling for significant confounders in the multivariable logistic regression model, African-Americans had an odds ratio for receiving warfarin of 0.74 (95% CI 0.58, 0.94) when compared to White patients.

CONCLUSION: African-American patients in the Ohio Medicaid population between 1998 and 2002 were significantly less likely than White patients to be prescribed warfarin for newly incident nonvalvular AF.

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

ROSUVASTATIN: SUPER-STATIN OR ALSO-RAN? A MULTI-CRITERIA DECISION ANALYSIS OF THE IMPACT OF ROSUVASTATIN ON OPTIMAL CLINICAL USE OF STATINS

Dolan J

Unity Health System, Rochester, NY

Purpose: Prudent decisions about the role of new pharmaceutical agents are important for cost-effective, high quality care. To determine the impact of rosuvastatin, approved for use in 2003, on optimal clinical use of the statins, a multi-criteria decision analysis based on the STEP approach - Safety, Tolerability, Effectiveness, Price - to comparing different pharmaceutical agents. (BMJ. 1996;312:1494) was performed using the Analytic Hierarchy Process.

Methods: Data were obtained from the literature regarding the safety, tolerability, and effectiveness of the 6 statins currently available in the US: atorvastatin (A), fluvastatin (F), lovastatin (L), pravastatin (P), rosuvastatin (R), and simvastatin (S). Average wholesale prices (Medical Letter 2004;46:38) were used to compare out-of-pocket costs. All statins are believed to be equally tolerable, so this criterion was removed from the analysis. Effectiveness was divided into LDL-C lowering and proven patient outcomes; LDL lowering was considered much more important. Safety was divided into 3 equally important sub-criteria: proven long term safety, minimal renal dose adjustments, and low potential for drug-drug interactions. The impact of rosuvastatin was determined by comparing the relative rankings of the statins without and with rosuvastatin.

Results: If cost is disregarded, and effectiveness and safety are considered equally important, atorvastatin (22.6% & 18.7%) and pravastatin (22.0% & 18.2%) are the top two drugs regardless of whether rosuvastatin is available or not. Rosuvastatin becomes the drug of choice when the relative priorities of effectiveness and safety are ≥78% and ≤22% respectively. When safety, effectiveness and price are included and considered equally important, atorvastatin (21.9% & 18.1%), fluvastatin (21.7% & 18.1%) and lovastatin (20.9% & 17.4%) are the three highest ranked statins both without and with rosuvastatin. Rosuvastatin becomes the drug of choice only when the priority of effectiveness is ≥65% and the priorities of safety and cost are ≤17.5% each.

Conclusions: For most patients, previously available statins are better choices than rosuvastatin. Its use should, therefore, be restricted to circumstances where effectiveness is the overriding concern. The proper use of statins depends on the relative priorities of safety, effectiveness, and price; routine assessment and integration of these considerations into prescribing decisions would help promote optimal use of these drugs. Routine, multi-criteria clinical assessment of new pharmaceuticals can help promote both quality of care and effective management of pharmaceutical costs.

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

PROVIDER ADHERENCE WITH QUALITY IMPROVEMENT STRATEGIES FOR MANAGEMENT OF HYPERTENSION: A SYSTEMATIC REVIEW

Walsh J[1], McDonald K[2], Shojania K[1], Sundaram V[3], Lewis R[2], Nayak S[2], Mechanic J[2], Owens D[3] and Goldstein M[2]

[1]University of California, San Francisco, San Francisco, CA; [2]Stanford University, Stanford, CA; [3]VA Palo Alto Health Care System, Palo Alto, CA

Purpose: We systematically assessed the effect of quality improvement (QI) strategies on provider adherence to recommendations for management of hypertension.

Methods: We searched MEDLINE, Cochrane databases, and article bibliographies for experimental evaluations of QI interventions targeting provider adherence with recommendations for evaluation and management of hypertension. Two reviewers independently abstracted data and classified each intervention into one or more of the following: provider education, provider reminders, facilitated relay of clinical information, patient education, audit and feedback or organizational change. Provider adherence outcomes were defined by each study and included adherence to guidelines for the following: evaluation of hypertensive patients, choice of medications, monitoring of blood pressure and counseling of patients. We combined all reported outcomes and calculated a summary provider adherence measure. We compared different QI strategies in terms of the median effects on improvement in proportion of provider adherence (%IMPROVE) to recommended practices for the evaluation or management of hypertension.

Results: 9 articles reporting 11 comparisons met inclusion criteria. Most studies included more than one QI strategy. Across all studies, the median %IMPROVE was 3.3% (interquartile range (IQR): 1.3, 6.4). The largest %IMPROVE was seen in studies that included provider reminders (10.4% (IQR: 7.0, 13.8)). Although the two studies that assessed provider reminders used different approaches, both included a paper reminder placed at the front of the patient's chart at the time of the physician visit. Provider education as a QI strategy was associated with a %IMPROVE of 3.7% (IQR: 1.7, 6.2), facilitated relay of clinical data was 3.3% (IQR: 2.0, 4.0), audit and feedback was 2.7% (IQR: -0.4, 3.7), and organizational change was associated with a %IMPROVE of 2.0% (IQR: -1.3, 3.3). Patient education was associated with a %IMPROVE of 2.0%, but was only included in one study. An increasing number of QI strategies was not associated with increasing provider adherence.

Conclusion: Although several QI strategies have been evaluated, most interventions had small effects on provider adherence. Provider reminders were assessed in two studies and seemed to have the largest effect, improving adherence by 10.4%. Provider behavior is difficult to change. Future studies should investigate innovative approaches to improving provider adherence with hypertension guidelines and should perhaps include simple chart reminders as part of the QI intervention.

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

TARGETING FRAGILITY FRACTURES IN AN ORTHOPAEDIC TREATMENT UNIT: COSTEFFECTIVENESS OF A DEDICATED COORDINATOR

Sander B[1], Maetzel A[1], Elliot-Gibson V[2], Beaton D[2] and Bogoch E[2]

[1]University Health Network, Toronto, ON, Canada; [2]St. Michael's Hospital Hospital, Toronto, ON, Canada

*Purpose:*The orthopaedic unit at a university teaching hospital hired an osteoporosis (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 =1 85, 43%), wrist (n = 124, 29%), humerus (n = 72, 17%) and other (n = 49; 11%); OP 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. Referral uptake, 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 3.2 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% benefits.

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: 1) over reasonable cost ranges, 2) 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.

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

FACTORS AFFECTING RESPONSE TO AND COMPLETION OF PHYSICIAN SURVEYS

Brehaut J, 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 [JB, 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 261/376, or 69.4%. Tests for response bias showed no effect of gender on likelihood to respond $\chi2(1)=2.38$, $p = 0.12$, but a marginal effect of location $\chi2(1) =3.48$, $p = 0.06$; 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 (65.8% for double-sided vs. 73.0% for single sided; $F(1,372) = 2.39$, $p = 0.12$), no significant effect of Sender (72.5% Known vs. 66.3% Unknown; $F(1,372) = 1.76$, $p = 0.19$), and no interaction $F(1,372) = 1.35$, $p = 0.25$. Chi-square tests showed that Print Format did not predict the likelihood of leaving more fields blank $\chi2(1) = 0.70$, $p = 0.40$, but that fewer fields were left blank when the Sender was Known $\chi2(1)= 5.30$, $p = 0.02$.

Conclusions. Print Format and Sender did not affect response rate 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.

POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISON MAKING

QUALITY OF LIFE IN PATIENTS ON RENAL REPLACEMENT THERAPY - A META-ANALYSIS

Liem Y[1], Bosch J[2] and Hunink M[3]

[1]Erasmus MC - University Medical Center Rotterdam, Rotterdam, Netherlands; [2]Erasmus MC, Rotterdam, Netherlands; [3]Erasmus Medical Center, Rotterdam, Netherlands

Purpose: To review and compare published literature on quality of life measured by the MOS 36-item short form health survey (SF-36) in hemodialysis and peritoneal-dialysis patients and renal-transplant recipients.

Methods: An English literature search was performed using Medline and PsycINFO. Studies were included if they reported SF-36-dimension scores of hemodialysis, peritoneal-dialysis and/or transplant recipients and included at least ten patients per treatment group. Demographic, clinical 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 30,405 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. Comparing the eight SF-36 dimensions, all patient groups reported lowest scores on the Role-Physical dimension (mean scores: 34.9 among hemodialysis, 33.0 among peritoneal-dialysis and 52.1 among transplant recipients). Highest scores were reported on the Mental-Health dimension for hemodialysis (mean score 67.3) 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 significantly different from scores of transplant recipients) and the Mental-Health dimension (hemodialysis and peritoneal-dialysis scores significantly higher than scores of transplant recipients). 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, comorbidity, and other potential predictors of quality of life in patients on renal replacement therapies need to be explored in future studies.

POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISON MAKING

VARIANCE REDUCTION WHEN COMPARING POLICIES IN COHORT SIMULATIONS

Shechter S, Braithwaite R, Schaefer A and Roberts M

University of Pittsburgh, Pittsburgh, PA

Purpose: When using Monte Carlo cohort simulation to compare outcomes between different policies, one desires a tight confidence interval (CI) around the difference in sample means for a fixed 100(1-α)% CI. One obvious way to achieve that is to increase the number of replications. The purpose of this study is to demonstrate how the variance reduction technique of common random numbers (CRN) may tighten the CI without increasing the replications.

Methods: To estimate the difference in expected lifetimes for a cohort of patients treated with policy A versus policy B, one would simulate n patients with each policy, form the pair-wise differences, and construct the desired confidence interval on the average. CRN tries to reduce the variance of the differences by making the same patients under each policy use identical random numbers (RNs) for the same reasons. For example, if patient 1 has a probability of .03 of dying from HIV under both policies in month 7, then CRN prevents policy A from generating a RN of .02 (patient dies) and policy B from generating .47 (patient lives). By making them use the same RN, CRN focuses any difference in outcomes on true policy differences rather than differences in luck. We implemented full and partial CRN in a simulation of 10,000 HIV patients undergoing highly active antiretroviral therapy (HAART) and evaluated the mean survival time for patients starting HAART at any CD4 count vs. starting at 200. Full CRN ensured that every RN that may be used by a patient is identical between policies. For partial CRN, we just ensured that the probabilities of dying were identical.

Results: The half-length of the 95% confidence intervals for the difference in mean survival for no CRN, partial CRN, and full CRN were .35, .18, and .12 years, respectively (the average difference in survival was .52 years). This implies, that it would take about 9 times as many replications using independent runs to achieve the same precision obtained using full CRN and about 4 times as many compared to partial CRN.

Conclusion: Though in general the use of CRN is not guaranteed to reduce the estimated variance, we have empirical support for its success when applied to individual cohort simulations such as the one presented here.

POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISON MAKING

INVESTIGATING DIFFERENTIAL ITEM FUNCTIONING IN PHYSICAL FUNCTIONING AND MENTAL HEALTH DOMAINS OF THE SF-36 AMONG POPULATIONS WITH THREE CHRONIC DISEASES (COPD, DEPRESSION, AND DIABETES)

Yu Y, Yu A and Ahn J

University of Southern California, Los Angeles, CA

Purpose: Differential item functioning (DIF) occurs when respondents from different groups have different responses to an item given the same latent ability. The existence of DIF may compromise the validity of group comparison using the instrument. The purpose of this study is to investigate the presence and the extent of DIF among the items of physical functioning (PF) and mental health (MH) domains of the SF-36 regarding to types of chronic illness.

Methods: The study population included three mutually-exclusive cohorts with COPD (n = 779), depression (n = 630), or diabetes (n = 652), who were members of Southern California Kaiser Permanente and completed SF-36 between April 1994 and February 1995. MIMIC (multiple indicators, multiple causes) models were constructed to investigate the effects of covariates including age, gender, education, race, income, marital status, employment status, and presences of various chronic diseases on the underlying factor and the DIF effect of disease type on item responses.

Results: Reliability coefficients were both high in PF (á=0.9) and MH (á=0.8) domains in all three disease groups. After controlling for the level of domain latent trait and other covariates, diabetes and COPD patients were more likely to endorse higher ability of moderate activity (PF2) and bending, kneeling or stooping (PF6) than depression patients, but less likely to rate higher capability of walking one mile (PF7). Diabetes and COPD patients tended to report lower frequency of and being downhearted (MH4), but diabetes patients tended to report higher frequency of being nervous (MH1), feeling down (MH2) and anxious (MH3). No significant DIF effects were found for other items.

Conclusions: In our study population, three out of ten PF items and four out of five MH items of SF-36 showed strong DIF pertaining to disease type. The presence of DIF effect in the items implied the disease-related information beyond the domain they were in. The level of physical function derived from PF domain can be compared among different chronic diseases. However, comparison of mental health status based on MH domain across diseases needs to take DIF into account due to significant DIF effect on major items.

POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISON MAKING

CAN INTERVENTIONS FOR ASTHMA BE MODELED THROUGH AN OBJECTIVE MEASURE OF LUNG FUNCTION?

Araki S[1], Fuhlbrigge A[2], Paltiel A[3], Croom K[2], Weiss S[2] and Kuntz K[4]

[1]Stanford University, Stanford, CA; [2] Brigham and Women's Hospital, Harvard Medical School, Boston, MA; [3]Yale School of Medicine, New Haven, CT; [4]Harvard School of Public Health, Boston, MA

Purpose: To determine the degree to which modeling asthma interventions as a function of forced expiratory volume in 1 second percent predicted (FEV_1) — an objective measure of lung function — predicts long-term outcomes of childhood asthma.

Methods: We developed a first-order Monte Carlo simulation model of childhood asthma using data from the Childhood Asthma Management Program (CAMP), a multicenter clinical trial designed to examine the long-term effects of inhaled corticosteroids (ICS) in asthmatic children. The primary outcomes consisted of acute events: hospitalizations, emergency room (ER) visits, and the need for rescue therapy. Monthly risks of each acute event were based on logistic regression equations estimated from the placebo arm of CAMP with covariates of age, FEV_1, prior hospitalizations, and prior night awakenings. We then modeled the effect of ICS as a percent improvement in FEV_1 (8% for mild, 12% for moderate, 15% for severe impairments in lung capacity) based on prior estimates derived from short-term clinical trials in the literature. Finally, we compared simulated outcomes for ICS with observed outcomes (i.e., from the ICS treatment arm in CAMP).

Results: Simulated versus observed mean number of acute events were similar in the placebo (natural history) group. Predicted [observed] means over 48 months were: hospitalizations 0.25 [0.25]; ER visits 1.05 [1.05]; and need for rescue therapy 4.82 [4.87]. In the treatment group, simulated versus observed number of events closely matched for hospitalizations (0.15 vs. 0.15), but not for ER visits (0.88 vs. 0.57) or need for rescue therapy (3.88 vs. 2.79). While the trial demonstrated treatment effects of 39% reduction in hospitalizations, 46% reduction in ER visits, and 43% reduction in the need for rescue therapy at 48 months; the model simulated reductions of 40% in hospitalizations, 16% in ER visits, and 19% in the need for rescue therapy. By modeling the effect of intervention on FEV_1, we were able to closely predict hospitalizations but underestimated treatment effects on ER visits and the need for rescue therapy.

Conclusions: Our findings suggest that intervention effects may be modeled through FEV_1 to predict hospitalizations in childhood asthma over time for planning and evaluation purposes. However, intervention effects on FEV_1 alone do not predict ER visits or the need for rescue therapy; additional factors must be considered.

POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISON MAKING

A META-ANALYSIS OF RANDOMIZED CONTROLLED TRIALS WITH CORONARY DRUG-ELUTING STENTS

Sondhi M, Jagannath A and Wong 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% to 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 (Pù.001) the risk of restenosis to 9% (RD -14%, CI -9% 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 DECISON MAKING

RELATIONSHIPS BETWEEN TTO, GENERAL HEALTH STATUS, AND DISEASE-SPECIFIC HEALTH STATUS IN PATIENTS WITH AGE-RELATED MACULAR DEGENERATION

Edwards P, Kongnakorn 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-Tradeoff 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=11): 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 .855 (σ = 0.191) and for group 2 was .608 (σ = .261). 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=0.098, r= .524) and MCS-12 (p=.098, 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 DECISON MAKING

USING CLASSIFICATION AND REGRESSION TREE ANALYSIS TO IDENTIFY CLINICAL PREDICTORS OF UTILITIES IN PARKINSON'S DISEASE

Siebert U[1], Bornschein B[2] and Dodel R[3]

[1]Massachusetts General Hospital, Harvard Medical School, Boston, MA; [2]Ludwig Maximilians University Munich, Germany, Munich, Germany; [3]University of Bonn, Bonn, Germany

Purpose: To identify the most relevant items of a commonly used clinical rating scale in Parkinson's disease for the prediction of utilities.

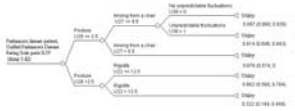
Methods: We used data from a prospective clinical study of the German Competence Network for Parkinson Syndromes (n = 122). We used 38 single items of the Unified Parkinson's Disease Rating Scale (parts II-IV) as potential predictors and utilities derived from EuroQol (EQ-5D) as outcome. We performed a classification and regression tree analysis (CART) with the t-test as test statistic for group selection and adjusted p-values for non-dichotomous variables by the method of Miller & Siegmund and the square-root (n)-method as stopping rule. To determine the explained variance, we entered the identified groups as indicator variables into a linear regression model.

Results: The final CART model had 3 levels with 4 variables partitioning the sample into 5 subgroups. These groups were defined by the degree of posture (item 28), the level of rigidity (item 22), problems with arising from a chair (item 27), and unpredictable fluctuations (item 36) Figure 1 shows the mean utilities in the 5 subgroups with 95% confidence intervals. Explained variance (adjusted R-square) was 0.50.

Conclusions: We successfully applied CART analysis to identify a parsimonious tree that predicts mean utilities based on clinical rating scale values. Our prediction tree is a simple tool that can easily be applied in the routine health care of bedside decision making in Parkinson's disease. However, these results need to be externally validated with independent data.

Figure 1

Utility subgroups in Parkinson's disease represented by a classification and regression tree (CART), based on 122 patients. Results at the "leaves" of the tree denote mean subgroup utilities with 95%-confidence intervals.



POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISON MAKING

OMISSION BIAS INFLUENCES THE MEDICAL DECISIONS OF PULMONARY SPECIALISTS

Aberegg S and Terry P

Johns Hopkins Hospital, Baltimore, MD

Purpose: To determine if status quo and omission bias influence the medical decisions of pulmonary specialists.

Methods: We designed three case vignettes that presented patient information with an associated patient management choice. Two versions of each vignette differed in the status quo state relating to the management choice that must be made. Versions of vignettes 1 and 2 also differed in whether action was required on the part of the decision maker. The case vignettes were administered during the first mailing of an opinion survey sent to 500 practicing pulmonary specialists. Chi-square tests were used for all comparisons.

Results: There were 122 respondents to the first mailing of the survey. Vignette 1 presented a patient with a low probability of pulmonary embolism and respondents were asked to choose between discharge of the patient versus ongoing evaluation. The status quo state varied in the two versions of the vignette by whether ongoing evaluation had already been initiated. Respondents were significantly more likely to choose ongoing evaluation if it had already been ordered and action was required to stop it (71% vs. 53%; RR = 1.6, p = 0.048). Vignette 2 presented a patient in shock and respondents had to decide whether or not to place a central venous catheter in addition to vasopressor therapy. Respondents were significantly more likely to forego placement of the catheter if a vasopressor had already been initiated and omission was an option (71% vs. 50%; RR = 1.70, p = 0.019). Vignette 3 presented a patient scenario and a hypothetical study which suggested that a routinely used but unproven therapy was shown to be harmful. In this vignette, the status quo varied between forms, but both forms required a choice between action and omission, and use of the therapy did not differ significantly based on whether it had already been initiated (50% vs. 55%; RR = 0.91, p = 0.60).

Conclusion: Medical decisions by pulmonary specialists vary according to normatively irrelevant status quo states, but this effect disappears after controlling for the action/omission distinction. Physicians should be aware that status quo and omission bias can influence medical decision making. Further study of the role of these biases in medical contexts is warranted.

POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISON MAKING

DO FALSE POSITIVE MAMMOGRAMS ADVERSELY AFFECT QUALITY OF LIFE? RESULTS FROM THE DMIST ACRIN TRIAL

Tosteson A[1], Fryback D[2], Hammond C[1], Hanna L[3], Brown M[4] and Pisano E[5]

[1]Dartmouth Medical School, Lebanon, NH; [2]University of Wisconsin, Madison, WI; [3]Brown University, Providence, RI; [4]University of North Carolina at Chapel Hill, Liverpool, NY; [5]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 positive 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 mammograms. Telephone interviews were completed shortly after the initial screening mammogram and one-year later. At both time points women completed a short-form of the Spielberger stait-trait anxiety questionnaire (STAI6), 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 negative 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(48%). At baseline, women who required additional work-up had significantly higher anxiety compared to age-matched women who did not (STAI6 35.2 vs. 32.7), but had similar health state values (EQ-5D 0.877 vs. 0.882; RS 84.3 vs. 85.9). At one year, there were no significant differences in STAI6, EQ-5D, or RS between those with false positive and negative mammograms. Women with false positive exams were significantly more likely to report that they would undergo future routine screening than those with a negative 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 DECISON MAKING

A DECISION MAKING TOOL FOR DISTRIBUTION OF HIV PREVENTION RESOURCES

Johnson-Masotti A[1], Brondino M[2], Reed B[3] and Pinkerton S[3]

[1]McMaster University, Hamilton, ON, Canada; [2]University of Wisconsin - Milwaukee, Milwaukee, WI; [3]Medical College of Wisconsin, Milwaukee, WI

Purpose: The goal of the study was to develop and pilot test a 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 though processes by using explicit criteria to evaluate proposed HIV prevention interventions ensure that decision-makers consider all important 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 human 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 attributes were identified as being of greatest importance in the decision to fund HIV prevention interventions. Second, attributes were weighted by decision-makers (HIV Prevention Community Planning 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 respect to HIV prevention interventions to be funded at these 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 HIV 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 proposed 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 similar 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 DECISON MAKING

MORE NUMERATE PATIENTS HAVE MORE ACCURATE IMPUTED EFFICACY JUDGMENTS CONCERNING TREATMENTS OF CERVICAL DYSPLASIA

Hamm R[1], Smith K[1], Hsieh S[1], Turner C[1], Juniper K[1] and McCree D[2]

[1]University of Oklahoma Health Sciences Center, Oklahoma City, OK; [2]Centers for Disease 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 judgment errors is removed. Higher IE accuracy in patients with higher numerical literacy (numeracy) 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 without various treatments. Measures of IE for 5 treatments were calculated. For example, from p(cervical cancer | dysplasia) if screened, and if not screened, we imputed efficacy of Pap screens for preventing 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 numeracy was measured using TOFHLA, REALM, arithmetic, and number comfort. Based on factor analysis, patients were assigned to high and low numeracy groups.

Results. 8 of patients' 10 P judgments were higher than experts', and 4 of their 5 IEs 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 numeracy factor, the more numerate patients had less variable responses than the less numerate for 7 of 10 Ps and for all 5 IEs. Accuracy of different P and IE judgments was related to different numeracy subscales. The more numerate patients were less accurate for 2 probabilities, and more accurate for 3 probabilities (all $p < .05$). However, their IEs were more accurate ($p < .05$) for 3 of 5 IEs.

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 patients' IEs were less variable for all 5 concepts, and significantly less inaccurate for 3 of 5 concepts.

POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISON MAKING

EFFECT OF PROSTATE CANCER SCREENING EDUCATIONAL INTERVENTIONS ON DECISION MAKING PROCESS OUTCOMES

Partin M[1], Nelson D[1], Nugent S[1], Holtzman J[2] and Wilt T[1]

[1]Minneapolis VA Medical Center, Minneapolis, MN; [2]University of Minnesota, Minneapolis, MN

Purpose: This study assessed the effects of mailed video and pamphlet prostate cancer (CaP) screening educational interventions on patient decisional involvement, certainty, and satisfaction.

Methods: The study design was a randomized, controlled trial. A sample of 1152 male veterans age 50 and older with no CaP and scheduled primary care appointments at one of four participating Veterans Affairs medical facilities in 2001, stratified by age (50-69, 70+), PSA in the past year (yes, no), and facility, was randomly assigned to one of three study groups: (1) mailed pamphlet, (2) mailed video, or (3) usual care (control). Intervention materials were mailed to patients two weeks prior to their scheduled primary care appointment, and outcomes were assessed by telephone survey approximately one week after the primary care appointment. A total of 42 participants were excluded from the study (8 deceased, 5 females, 29 CaP diagnoses). A total of 893 of the remaining 1110 participants completed the survey (80% response rate) and were included in analyses. Outcomes examined using adjusted linear regression models included: (1) the 13-item Patient Perceptions of Involvement in Care Scale (PICS), (2) Oconnor's 10-item Factors Contributing to Decisional Uncertainty Scale, and (3) Holmes-Rovner's 6-item Satisfaction with Decision scale.

Results: Roughly 56% of video and 50% of pamphlet subjects reported looking at the mailed materials. Intent to treat analyses revealed that both video and pamphlet subjects had significantly lower Factors Contributing to Uncertainty scale scores than controls (23.8, 23.8 and 24.5, respectively). Pamphlet subjects were significantly more likely than controls to be involved in CaP decision making (PICS scores 1.73 and 1.20, respectively, p=.05) but video subjects were not (PICs score 1.53, p=.21). Neither intervention increased the generally high levels of satisfaction with decision-making (Satisfaction with Decision scale scores 18.5, 18.3 and 18.3 for video, pamphlet, and control subjects, respectively).

Conclusions: The interventions have modest effects on decisional uncertainty and involvement, and no affect on satisfaction. The slightly more pronounced effects of the pamphlet on decision-making involvement may make it an attractive low cost strategy for busy practice settings searching for ways to facilitate patient participation in CaP screening decisions. The impact of the interventions may be enhanced by increasing patient exposure to the intervention material and enhancing provider facilitation of CaP screening decisions.

POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISON MAKING

QUANTIFYING THE EXPECTED BENEFIT OF DROTRECOGIN ALFA (ACTIVATED) IN THE TREATMENT OF SEVERE SEPSIS WITH A PATIENT-SPECIFIC DECISION SUPPORT TOOL

Schauer D[1], Johnston J[2], Hornung R[1], Leonard A[1] and Eckman M[1]

[1]University of Cincinnati, Cincinnati, OH; [2]Eli Lilly, Indianapolis, IN

PURPOSE: The expected benefit of treating severe sepsis with drotrecogin alfa (activated) for an individual patient may depend upon a number of clinical factors including disease severity. Our purpose was to create a decision support tool incorporating patient-specific inputs to estimate the balance of treatment risks and benefits for individual patients.

METHODS: The decision support tool has two components. Logistic regression models were developed to calculate patient-specific mortality risk with and without treatment, which were then used as inputs into a seventy-five state Markov model. Data from the Recombinant Human Activated Protein C Worldwide Evaluation in Severe Sepsis (PROWESS) trial and subsequent open-label studies were used to develop the logistic models and the Markov model. Patient-specific inputs included patient age, gender and twelve readily available clinical characteristics. Expected patient outcomes were expressed in quality-adjusted life years (QALYs).

RESULTS: The expected benefit from drotrecogin alfa (activated) treatment was most dependent upon the underlying disease severity. For example, treatment of a 55 year-old white male with severe sepsis and a 28-day mortality risk of 45% resulted in an expected gain of 3.1 QALYs (16.0 v. 12.9). In a similar patient with less severe disease (28-day mortality risk = 20%), the expected net benefit from drotrecogin alfa (activated) therapy would be 0.8 QALYs (19.6 v. 18.8). However, the treatment decision is sensitive to bleeding risk. If this same patient (28-day mortality risk = 20%) were at an increased risk of serious bleeding prior to treatment (28-day probability of bleeding = 7% vs. baseline of 2%) and an increased relative risk of serious bleeding with treatment (RR of 5.5 vs. baseline of 1.8) then not treating with drotrecogin alfa (activated) would yield the most QALYs (18.7 vs 18.6).

CONCLUSION: A customizable decision model using patient-specific inputs can be used to inform the treatment decision when considering the use of drotrecogin alfa (activated) therapy by weighing the risks versus the benefits of therapy in the treatment of severe sepsis.

POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISON MAKING

DO PHYSICIANS RECALIBRATE PATIENTS' PAIN FOR FUNCTIONAL REASONS?

Marquié L[1], Raufaste E[1], Lauque D[2], Mariné C[1] and Sorum P[3]

[1]Centre National de Recherche Scientifique, Université Toulouse-II, Toulouse, France; [2]Centre Hospitalier Universitaire Purpan, Toulouse, France; [3]Albany Medical College, Latham, NY

Background: Physicians give systematically lower ratings of patients' pain than do the patients themselves. We found (Pain 2003;102:289-96) that emergency department (ED) physicians in Toulouse, France, were influenced by "non-functional" factors like their own gender. We hypothesized that physicians' recalibration of patients' pain is due also to "functional" factors, namely, the physical signs of pain and the apparent pathology responsible for the pain.

Methods: 52 ED physicians in Toulouse (26 males and 26 females) evaluated 45 scenarios of hypothetical patients with abdominal pain. The scenarios consisted of all combinations of 5 levels of the patient's own pain rating (0, 2, 5, 7 and 10 on a 0-10 scale), 3 levels of physical manifestations of pain, and 3 configurations of cues about the severity of the abdominal pathology. The physicians rated patients' pain on a 0-10 visual analog scale; indicated the most likely diagnosis on the basis of the information about history, physical exam, and white blood cell count; and, after completing the scenarios, indicated the expected range of pain ratings for each possible cause of abdominal pain.

Results: Physicians tended to assign middle ratings to patients' pain, thus giving increasingly higher ratings than did the paper patients as the patients' ratings declined below 5 and increasingly lower ratings as the patients' ratings increased above 5. Multiple linear regression found that the extent of this recalibration of the patient's pain was predicted by the patient's pain rating ($\beta = -0.67$; $p < .001$), the manifestations of pain ($\beta = 0.43$; $p < .001$), the cues about severity of pathology ($\beta = 0.13$; $p < .001$), and the minimum expected level of pain for the most likely pathology ($\beta = 0.09$; $p < .001$).

Conclusions: ED physicians in Toulouse used the patient's appearance and signs of abdominal pathology to adjust the patient's own assessment of pain. "Miscalibration" of patients' pain is, therefore, not merely a result of bias and mistrust of patients; it can be functional.

POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISON MAKING

THE PERILS OF INFORMED CONSENT: STANDARD SURVEY INTRODUCTIONS CAN BIAS RESPONSES

Smith D, Schwarz N, Roberts T and Ubel P

University of Michigan, Ann Arbor, MI

Purpose: Research on survey methodology has demonstrated that seemingly innocuous aspects of a survey's design, such as the order of questions, can produce biased results. The current investigation extends this work by testing whether standard survey introductions alter the observed associations between variables. We fielded a brief survey of health and well-being to Parkinson's patients, and varied the survey introduction to test whether informing participants that the study focuses on Parkinson's disease (PD) would make health status more salient, and thus increase the correlation between health and overall life satisfaction.

Method: 153 PD patients agreed to a short phone interview on health and well-being. After initially agreeing to participate, participants were randomly assigned to receive either standard survey instructions, indicating that we were contacting people with PD, or a version that indicated only that we were contacting people living in their area. All participants were fully informed of the purpose of the survey at the conclusion. The survey first assessed life satisfaction, and subsequently health satisfaction.

Results: Health satisfaction was a much larger predictor of life satisfaction for the group who received the "Parkinson's" introduction ($r2 = .48$) than for the group who received the introduction with no mention of the illness ($r2 = .16$; regression interaction $p < .01$). An examination of means indicated that the PD survey introduction significantly decreased reported life satisfaction for participants below the median in health, and slightly increased life satisfaction for participants above the median in health (interaction $p < .05$).

Conclusions: When participants were informed prior to the survey that its purpose was to examine well-being in PD, health satisfaction was a much more important component of life satisfaction, accounting for 3 times as much variation. We hypothesize that the survey introduction primed participants' health status, resulting in an artificially large correlation with life satisfaction (a previously published meta-analysis on the link between subjective health and life satisfaction agrees with the smaller estimate). An implication is that fully informed consent can bias survey results. In innocuous surveys, it may well be preferable to provide full information about the purpose at the conclusion of the survey.

POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISON MAKING

PEDIATRICIANS' DECISION MAKING: A RANDOMIZED CONTROLLED TRIAL OF DECISION SUPPORT WITH FALSE POSITIVE RATE

Sox C[1], Koepsell T[1], Doctor J[1] and Christakis D[2]

[1]University of Washington, Seattle, WA; [2]University of Washington, Seattle, SC

Purpose: To determine whether presenting pediatricians with the false positive rate of screening test affects their 1) estimate of disease probability, and 2) subsequent patient evaluation.

Methods: We sent a mailed survey to 1502 pediatricians practicing in the United States who were randomly selected from the AMA master list. Subjects were sent a four-page questionnaire, with up to four mailings sent to non-responders. The questionnaire presented a clinical case of a healthy 5-year old boy with persistent micro-hematuria detected by screening urinalysis. 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 false positive rate of persistent micro-hematuria in detecting significant renal disease (96%). The non-technical decision support group was presented a non-technical explanation of this false positive rate. Subjects were asked to 1) chose the patient's probability of serious renal disease, 2) refer the patient to nephrology or not, and 3) check his serum BUN and Cr, or not. We conducted chi-square to test the effect of decision support on these three outcomes.

Results: 653 subjects returned completed surveys: 208 in the control, 231 in the technical and 214 in the non-technical decision support groups. The estimated response rate was 52%. 81% of participants were board certified in pediatrics, 56% were female, and their mean age was 43 years. The only significant difference between the randomization groups was by gender. Subjects who received non-technical decision support were much more likely than controls to choose the correct disease likelihood (51% vs. 11%, $p < 0.001$), while those who received technical decision support were not (16% vs. 11%, $p = 0.10$). Subjects who received non-technical decision support were more likely to refer the boy to nephrology (30% vs. 19%, $p = 0.01$) and check his BUN and Cr (88% vs. 78%, $p = 0.01$), but those who received technical decision support were not (22% vs. 19%, $p = 0.36$; 75% vs. 78%, $p = 0.48$, respectively). None of these findings were significantly altered after controlling for gender.

Conclusions: Presenting non-technical decision support about the false positive rate of screening urinalysis improved pediatricians' estimations of disease probability and affected patient management, while technical decision support did not.

POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISON MAKING

VALUATION OF HEALTH USING TIME TRADE-OFF AND STANDARD GAMBLE: NON-TRADERS AND NON-GAMBLERS

Suarez-Almazor M[1], Souchek J[1], Byrne M[2], Kelly P[1], Richardson M[1], Pak C[1] and Nelson H[1]

[1]Baylor College of Medicine/Michael E. DeBakey VA Medical Center, Houston, TX; [2]University of Pittsburgh, Pittsburgh, PA

PURPOSE: To examine the characteristics of individuals unwilling to trade-off time or gamble in the valuation of health scenarios using time trade-off (TTO) or standard gamble (SG) methods.

METHODS: As part of a study eliciting preferences for arthritis states, we surveyed 391 individuals using TTO and SG: 193 were members of the public, identified by random digit dialing, and 198 were individuals diagnosed with osteoarthritis (OA). Participants rated two hypothetical scenarios of patients with OA (mild and severe), and own health. Individuals unwilling to trade or gamble to obtain perfect health were compared to the remainder subjects. Differences were analyzed with parametric and non-parametric univariate tests, and logistic regression.

RESULTS: For the mild OA scenario, 9.2 % of the subjects were unwilling to trade-off time for perfect health, and 14.3 % were unwilling to gamble. For the severe scenario, the percentages were 4.8 and 10.5% respectively. Statistically significant differences were observed between the public and patients, with patients being more willing to trade or gamble for perfect health. For own health, 9.1% of the patients and 17.6% of the public were non-traders ($p = 0.01$); 11.6% of the patients and 26.4% of the public were non-gamblers ($p < 0.001$). When compared to the other respondents, non-traders and non-gamblers were significantly more likely to be older, African-American, have lower educational status and difficulties in completing the survey. Religious beliefs, and self-reported health status (after controlling for being a member of the patient or public groups) were not associated with non-trader or non-gambler status.

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.

POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISON MAKING

COST-EFFECTIVENESS OF DALTEPARIN VERSUS WARFARIN FOR ACUTE VENOUS THROMBOEMBOLISM IN PATIENTS WITH CANCER

Aujesky D[1], Smith K[2], Cornuz J[3] and Roberts M[2]

[1]VA Pittsburgh Healthcare System, Pittsburgh, PA; [2]University of Pittsburgh, Pittsburgh, PA; [3]University Hospital of Lausanne, Lausanne, Switzerland

PURPOSE: Although dalteparin has been shown to be more effective than warfarin in preventing recurrent venous thromboembolism (VTE) in cancer patients with acute VTE, its cost-effectiveness is uncertain. We constructed a decision analytic model to calculate quality-adjusted life-years (QALYs) and lifetime costs for treatment with these drugs to compare their cost-effectiveness.

METHODS: Using a societal perspective, our model compared two strategies to treat acute VTE in 65-year-old patients with cancer. In the dalteparin strategy, the daily dalteparin dosage was 200 IU/kg during month 1 and 150 IU/kg during months 2 to 6. In the warfarin strategy, warfarin was given for 6 months at a target international normalized ratio of 2.5 and dalteparin was given for the first 5 days at a dosage of 200 IU/kg. Our model incorporated probability estimates and utilities reported in the literature and published cost data. We conducted multiple one-way and probabilistic (Monte-Carlo) sensitivity analyses to assess the effect of varying baseline estimates on cost-effectiveness.

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 $6,580. 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 <$17 per day (26% 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 22%.

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 unfavorable economically compared with warfarin.

POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISON MAKING

UP-CLOSE-AND-PERSONAL: UNDERSTANDING PHYSICIAN DECISION-MAKING THROUGH VIDEOTAPES OF PATIENT-PHYSICIAN INTERACTIONS

Tai-Seale M

Texas A&M University Health Science Center, Bryan, TX

Background: A decision is defined as a verbal commitment to a definitive course of action. Measuring the process of medical decision-making with micro-level interaction data may enhance our understanding of what happens in medical practices and how to change it to improve quality of care.

Methods: Qualitative critical discourse analysis of videotaped visits complements quantitative investigations which involve designing a coding system to capture the process, content, and rapport of patient-physician information exchange prior to a medical decision and econometric analyses. Duration model explores the role of patient information input in physician's decision-making. The unit of analysis is each topic covered in a visit.

Results: Critical discourse analysis revealed that physicians often did not address patients' key concerns. Affective issues particularly received little attention and were frequently marginalized. Quantitative analyses 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 visit contained an average of 5 topics. Though patients showed verbal or nonverbal cues of mood disorder in over 15% of the topics, only 6% of them involved discussion of affective issues. Patients spoke more than twice as long in emotional topics compared with in other topics. Robust hazard ratios from the duration model were obtained after accounting for the clustering effect of patient-physician pairs. Hazard ratios suggested that patients talked longer when they had initiated the topic, the topic was about emotions, physician showed personal uncertainties, patient showed verbal cues of emotional distress, patient had private insurance, and when patient had better mental health status. The late onset of a topic, being in a managed care organization, longer length of patient-physician relationship, and better scores in physical role functioning, pain, and vitality were associated with shorter patient talk time and short time it took to reach a decision.

Conclusions: Studying the process of care using videotapes of patient visits provides new insights on decision-making. The dynamics of communication play a significant role in patient-physician interaction. Patients should be advised to initiate topics that are important to them early in the visit so that they can give sufficient information on their preferences and receive needed information from their physicians before a decision is made.

POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISON MAKING

VALIDATING AND UPDATING A PREDICTION RULE FOR NEUROLOGICAL SEQUELAE AFTER CHILDHOOD BACTERIAL MENININGITIS

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. In order to validate this rule, it was applied to 628 patients, collected from almost all hospitals in The Netherlands (validation set), and the probability of neurological sequelae was estimated for each patient. We assessed calibration with a calibration line that described the relation between the predicted probabilities and the observed frequencies, and with the Hosmer-Lemeshow goodness-of-fit test. The discriminative ability was studied with the ROC area. Finally, we updated the original rule by adding extra predictors and re-estimating the regression coefficients using the merged derivation and validation sets. Results The calibration plot of the original rule in the validation set showed poor agreement between predicted probabilities and observed frequencies for the patients of the validation set. This was confirmed by a significant Hosmer-Lemeshow test (p-value < 0.01). The ROC area was 0.65 (95%CI: 0.57-0.72), which was significantly lower than the area found in the derivation set (0.87 (95%CI: 0.78-0.96)). In the merged data sets, gender was no longer an important predictor. In addition to atypical convulsions, body temperature at physical examination, and type of pathogen, the use of anti-epileptica > 2 days during hospital admission, and presence of petechiae and/or ecchymoses appeared to be important predictors of neurological sequelae. The ROC area of this updated rule was 0.75 (95%CI: 0.67-0.83) after correction for overoptimism and its calibration was acceptable. Conclusion The former developed prediction rule for neurological sequelae after childhood bacterial meningitis showed poor performance when applied to another much larger population. The updated prediction rule showed adequate calibration and discrimination in the merged data sets. Further study of model validity may stimulate application in clinical practice.

POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISON MAKING

HOW WIDELY USED IS THE CANADIAN C-SPINE RULE BY EMERGENCY PHYSICIANS?

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

Ottawa Health Research Institute, Ottawa, ON, Canada

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 of 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-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 correct addresses. Overall response rate was 261/376, or 69.4%. 216 (82.8%) reported already being familiar with the rule, while 163 (62.5%) reported already using it. Of the 98 (37.5%) 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%, Easy to learn 74.6%, Easy to remember 60.4%. Factors associated with never having seen the rule include older age $F(2,218) = 4.84$, $p = .009$, non-specialist status $\chi2(2)= 5.78$, $p = 0.055$, part time status $\chi2(4)= 16.48$, $p = 0.002$, and community hospital setting $\chi2(4)= 8.89$, $p = 0.06$.

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.

POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISON MAKING

THE CHEAPER THE BETTER? SHOULD LOVASTATIN BE THE STATIN OF CHOICE FOR ROUTINE USE IN PRIMARY CARE?

Dolan J

Unity Health System, Rochester, NY

Purpose: The statins are among the most frequently prescribed prescription drugs. Their appropriate use is essential for both quality of care and control of prescription drug costs. Managed care companies have encouraged clinicians to prescribe lovastatin (available generically) as the statin of first choice. The goal of this study was to determine if this policy is consistent with optimal use of statins in clinical practice.

Methods: A multi-criteria decision analysis was performed using the Analytic Hierarchy Process based on the STEP approach -Safety, Tolerability, Effectiveness, Price - to comparing different pharmaceutical agents. (BMJ. 1996;312:1494) Data were obtained from the literature regarding the safety, tolerability, and effectiveness of the initial doses of the 6 statins currently available in the US: atorvastatin (A), fluvastatin (F), lovastatin (L), pravastatin (P), rosuvastatin (R), and simvastatin (S). Prices were obtained from a 3-tier formulary used by a managed care organization that recommends first line use of lovastatin. This formulary divides drugs into 3 categories, with patient co-payments of $5 for each prescription of lovastatin, $20 for atorvastatin or pravastatin, and $35 for fluvastatin, rosuvastatin, or simvastatin. All statins are believed to be equally tolerable, so this criterion was removed from the analysis. Effectiveness was divided into LDL-C lowering and proven patient outcomes; LDL lowering was considered much more important. Safety was subdivided into 3 equally important sub-criteria: proven long term safety, minimal renal dose adjustments, and low potential for drug-drug interactions.

Results: If Safety, Effectiveness, and Price are considered equally important (priorities all equal to 0.33), lovastatin is the drug of choice (L 21.4%; A 18.5%; P 18.2%; R 14.3%; F 13.8%, S 13.8%) Sensitivity analysis shows that lovastatin is the drug of choice if its priority is: ≥0.21 if Effectiveness and Safety are considered equally important, ≥0.195 if Effectiveness is considered twice as important as Safety, and ≥0.23 if Safety is considered twice as important as Effectiveness.

Conclusions: Advice to prescribe lovastatin as the statin of first choice is reasonable as long as patient out-of-pocket cost is given a priority of at least 20% relative to Effectiveness and Safety. Routine assessment and incorporation of patient priorities regarding these criteria would help both maintain quality of care and effectively manage pharmaceutical costs.

POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISON MAKING

NATURAL HISTORY OF CERVICAL INTRAEPITHELIAL NEOPLASIA: A META-ANALYSIS

Cantor S[1], Atkinson E[1], Cardenas-Turanzas M[1], Benedet J[2], Follen M[3] and MacAulay C[3]

[1]The University of Texas M. D. Anderson Cancer Center, Houston, TX; [2]University of British Columbia and BC Cancer Agency, Vancouver, BC, Canada; [3]British Columbia Cancer Research Center, Vancouver, BC, Canada

Purpose: The objective of this study was to determine the probabilities of transition of stages in the natural history of cervical cancer by conducting a formal meta-analysis of published studies of the topic.

Methods: We identified health states of interest in the natural history of cervical pre-cancer, identified all possible papers from MEDLINE (years 1966-2002) that met selection criteria, developed relevance and acceptability criteria for inclusion, then thoroughly reviewed the selected studies. Four transitions were investigated in detail: (1) high-grade squamous intraepithelial lesions (HGSIL) to cancer, (2) low-grade squamous intraepithelial lesions (LGSIL) to HGSIL, (3) HGSIL to LGSIL, and (4) LGSIL to normal. We converted data to determine 6-month transition probabilities, as this is the time frame for follow-up after an abnormal Papanicolaou smear. To determine the transition probability data we used a random effects model that assumed the parameters were drawn from a gamma distribution.

Results: 28 studies were found that met acceptability and relevance criteria; not all studies provided data for all 4 health-state transitions. Our final analysis included 9 studies for HGSIL to cancer (follow-up 0-336 months), 12 studies for LGSIL to HGSIL (follow-up 0-76 months), 3 studies for HGSIL to LGSIL (follow-up 0-132 months), and 12 studies for LGSIL to normal (follow-up 0-77 months). The homogeneity test failed for each transition under investigation, i.e. the studies were not homogeneous. The 6-month mean predictive transition probability (95% confidence intervals with "prediction interval" in parentheses) for HGSIL to cancer was 0.0037 (0.00004, 0.03386); for LGSIL to HGSIL was 0.0362 (0.00055, 0.23220); for HGSIL to LGSIL was 0.0282 (0.00027, 0.35782); and for LGSIL to normal was 0.0740 (0.00119, 0.42672).

Conclusion: The transition probabilities between cervical cancer health states for 6-month intervals are small. However, there was significant variation in the probabilities for the various health-state transitions. The results of this study, when used in a decision-analytic model for cervical cancer screening, will have to undergo extensive sensitivity analysis.

POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISON MAKING

THE COST-EFFECTIVENESS OF SURGERY PLUS RADIATION FOR REGIONAL ORAL CAVITY CANCER IN THE ELDERLY

Hollenbeak C[1], Stack B[1] and Piccirillo J[2]

[1]Penn State College of Medicine, Hershey, PA; [2]Washington University School of Medicine, St. Louis, MO

Purpose. Treatment for regional cancer of the lip and oral cavity may consist of surgery or surgery followed by radiation. The purpose of this research was to determine whether the addition of a course of radiation therapy following surgery for regional cancer of the lip and oral cavity was cost-effective compared to single modality surgery.

Methods. We studied 660 Medicare patients treated for known regional cancer of the lip and oral cavity with surgery plus radiation (N = 337) or surgery alone (N = 323) between 1984 and 1994 using the SEER-Medicare linked database. We first estimated survival functions for these patients stratified by treatment modality using the product limit estimator of Kaplan and Meier. We next computed average accumulated costs up to five years, stratified by treatment. Finally, we estimated the cost-effectiveness of surgery plus radiation compared to surgery alone by computing an incremental cost-effectiveness ratio. The incremental cost-effectiveness ratio represents the additional cost per life saved in five years if a treatment strategy of surgery plus radiation were used instead of surgery alone.

Results. As seen in the figure, five years after diagnosis, 59% of patients with treated with surgery plus radiation had survived compared to 40% of patients receiving surgery alone ($p = 0.0007$). Furthermore, in spite of the added cost of radiation, patients receiving surgery plus radiation also had lower average five-year costs than patients receiving surgery alone (313,302 vs. 149,884). Combining the five-year survival and accumulated costs implies an incremental cost-effectiveness ratio of -$909,091 per life saved, which implies that surgery plus radiation dominates surgery alone. Conclusions. Radiation in addition to surgery is highly cost-effective in the treatment of regional cancer of the lip and oral cavity in elderly Medicare patients. In this study of 660 Medicare patients, radiation plus surgery had both lower average costs and lower mortality rates at five years. Further research is needed to determine whether these findings are obtained for other sites and stages of disease and for younger head and neck cancer patients.

POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISON 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

Labarere J[1], Bosson J[2], Belle L[3] and Robert C[4]

[1]CHU BP 217, GRENOBLE Cedex 9, France; [2]TIMC-IMAG, Grenoble Cedex 9, France; [3]CH de la region annecienne, Annecy, France; [4]Institut National de Recherche en Informatique et en Automatique, Montbonnot, 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: We 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 patients), "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 each 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 over the 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 DECISON MAKING

HOW OFTEN SHOULD HEALTHY ADULTS RECEIVE TETANUS BOOSTER VACCINES?

Mitchell C[1] and McCabe S[2]

[1]University of Louisville, Louisville, KY; [2]Christine M. Kleinert Institute for Hand and Microsurgery, Louisville, KY

Purpose: Vaccines are given to healthy individuals with the intention of preventing morbidity and mortality. When the disease the vaccine is targeting is exceedingly rare, like tetanus, the risks associated with immunization must be considered in the overall vaccine policy. Major authorities continue to recommend healthy adults receive tetanus booster vaccinations every 10 years. An alternative strategy of a single booster at 50 years of age may be as effective and safer. 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 20 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 76.77 QALY and provided 2,000 (undiscounted) QALYs per 100,000 cohort members. It was cost saving at $188/QALY compared to $407/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.

POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISON MAKING

CAN DECISION SUPPORT BE SUCCESSFULLY INTEGRATED INTO CLINICAL CARE?

Moore C[1], Collins E[1], Clay K[1], Kearing S[2], Gallagher S[3], Amon D[3], Llewellyn-Thomas H[2] and O'Connor A[4]

[1]Dartmouth Hitchcock Medical Center, Lebanon, NH; [2]Dartmouth Medical School, Hanover, NH; [3]Dartmouth Medical School, Lebanon, NH; [4]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 documenting: physical and emotional health; distress levels; and decisional needs (e.g., understanding of options, social pressures, and values clarity). Responses were instantly summarized in a report and social workers were automatically alerted to intervene if a patient reported high levels of distress. Patients then viewed the video and proceeded to surgical consultations. To assess feasibility, we surveyed patients about questionnaire content, response burden, and relevance to care; we surveyed staff and surgeons about impact on scheduling and clinic flow. To assess benefits, we utilized survey data of the patients' perceptions of the video, as well as measuring the number of interventions for distress, and physician and staff response to the new process.

Results: Feasibility: All eligible patients have completed the new protocol. Most (28/35) were satisfied with the questionnaire's content and length. No physicians experienced clinic disruption. Staff members reported some problems scheduling surgery appts, but no problems scheduling the intake (video and questionnaire appointment). Benefits: The video helped most (97/102) patients to: organize thoughts, identify questions, talk with their doctor, and make better decisions. Many questionnaire subjects(19/35) reported distress or mental health symptoms that triggered intervention. All staff (n = 6) and physicians (n = 4) agreed that adding decision support to routine care improved services, and that benefits outweighed the burden. All physicians agreed that the report was helpful. (Additional data will be available at presentation.)

Conclusions: Integrating decision support with clinical care is operationally feasible and yields benefits for patients and physicians. Overall, the clinic's transition to the new protocol has been nearly seamless. Results demonstrate a need for distress intervention, which can be folded into decision support. The video and report help prepare both patients and physicians for the initial consultation. With further testing this process can be adapted to a variety of health conditions.

POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISON MAKING

VALIDATION AND FEASIBILITY OF THE "VAN DEN BERGHE" INSULIN PROTOCOL TO MAINTAIN A TIGHT BLOOD GLUCOSE CONTROL IN THE CRITICALLY ILL

Oeyen S, Poelaert J, Colpaert K and Decruyenaere J

University Hospital Ghent, Ghent, Belgium

Purpose: The Van den Berghe group developed an insulin protocol (IP) to target glycemia (G) between 81-110 mg/dL which resulted in a mortality reduction of 50% (N Engl J Med 2001 345:1359-67). We evaluated the compliance with the IP in the ICU of an university hospital and thereby the efficacy and feasibility of this protocol.

Methods: Evaluation of the IP was started after lessons and a 1 month initiation period. All patients with an arterial line and an expected ICU-stay (ILOS) > 48 h were included. Nurses measured G on undiluted arterial blood samples using a bedside glucometer (GlucoTouch, LifeScan, Benelux). The according rate of the continuous insulin infusion (Humulin Regular, Lilly) and the frequency of G measurements were exclusively managed by ICU-nurses. A correct compliance with the IP was seen as a right insulin dose adjustment according to the protocol and the condition of the patient. Efficacy was established by comparing the number of G values within the correct range to the total number of G measurements. We defined 2 groups adjusted to the median percentage of correct G range: (1) G values ≤ 39% and (2) > 39% in the normal range. Data are presented as mean ± SD or as percentage. Statistical analysis was performed using an independent samples T-test. Significance was accepted when $p<.05$.

Results: Out of 6016 G measurements, (30 patients - APACHE II 26±8) a good compliance with the IP in 70.9% (N = 4267) was found. Only 41.6% (N = 2504) of all G values were regulated in the desired range despite hourly or two-hourly G measurements (17 ± 7 per day). Group 1 showed higher G values G>250 mg/dL 1.5±1.5 vs 0.7±0.9; $p < .006$. In group 2 more G measurements were < 40 mg/dL 0.1±0.4 vs 0.5±0.9; $p < .01$. The 2 groups were comparable concerning APACHE II score, number of diabetic patients, ILOS, compliance with the IP and number of daily G measurements.

Conclusion: The compliance with the paper-based "Van den Berghe" IP was acceptable but can be improved. A good compliance does not imply a high percentage of G values within the target range. Further study is needed to see if computer-based and more complex guidelines, implemented bedside in an Intensive Care Information System, could lead to a better blood glucose control.

POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISON MAKING

FACILITATED PROCESS IMPROVEMENT: A PRAGMATIC APPROACH TO GUIDELINE IMPLEMENTATION

Patwardhan M, Samsa G and Matchar D

Duke University Medical Center, Durham, NC

Purpose: In order to promote decision-making based on available evidence, we developed Facilitated Process Improvement (FPI) as a pragmatic approach to implement new guidelines for advanced chronic kidney disease (CKD). The approach is based on the Theory of Constraints (TOC), and is intended to be a general technique for developing effective practice improvement strategies for busy clinicians.

Methods: FPI is distinct from other forms of process improvement because it moves the formative work of guideline implementation from the practice to a national team consisting of a professional organization and practice improvement methodologists. FPI is based on the three TOC questions about change: *what* to change (understanding current processes, undesirable effects and their root causes), what to change *to* (establishing functional specifications), and *how* to cause the change (developing tools to address the root causes). The table illustrates the steps of FPI and the participants in this CKD practice improvement effort.

FPI Steps	Participants/Data sources
1. *Establish functional specifications*: what the care processes intend to accomplish	CKD guideline
2. *Investigate processes of care* in variety of site-types to understand the process, reasons *why processes fail to meet* functional specifications	Focus groups: nephrologists, non-nephrologists, patients
3. *Develop a practical tool* to attack root causes of process failures	Literature search Nephrologist interviews Focus groups: nephrologists, non-nephrologists, patients National expert working group (WG)
4. *Develop a meta-tool* to guide local tool selection	WG

Results: The result is the "Advanced CKD Management Toolkit" consisting of implementation tools (e.g. a flow-sheet, a referral form to ease communication between providers) and the meta-tool, which is in two forms. The "quick start" form is based on typical practice vignettes (e.g. "a busy primary provider with no interest in managing advanced CKD patients"). The "insight" form guides providers through an exercise to understand local processes and the potential for specific tools. The *application* of the meta-tool will be the only function of the site involved in guideline implementation.

Conclusions: FPI provides an explicit link between evidence and practice that takes advantage of the conceptual framework of TOC, and builds on principles of total quality management (TQM), but reduces TQM's substantial demands. In effect, FPI transmits what is known about process improvement to local personnel committed to guideline implementation but limited by resources and experience.

POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISON MAKING

DIAGNOSTIC PROCEDURES FOR DETECTING PANCREATIC CANCER AND ASSESSING RESECTABILITY: A CLINICAL DECISION ANALYSIS

Schink T[1], Böhmig M[1], Hur C[2], Koch I[1], Rosewicz S[1], Wernecke K[1] and Siebert U[2]

[1]Charité, Humboldt University Berlin, Berlin, Germany; [2]Massachusetts General Hospital, Harvard Medical School, Boston, MA

Purpose: To assess different strategies for diagnosis and determining resectability in patients with suspected pancreatic cancer (PC).

Methods: We used the data from a prospective study of 193 patients with suspected PC who were enrolled at the Charité University Hospital from 08/1999-11/2000. These patients underwent each of the following six different diagnostic procedures to determine diagnosis of pancreatic cancer and assessment of resectability: ultrasound (US), magnetic resonance imaging (MR), computed tomography (CT), endoscopic ultrasound (EUS), fluorodeoxyglucose positron emissiontomography (PET), and endoscopic retrograde cholangio-pancreaticography (ERCP). We developed a decision tree to predict diagnostic accuracy and resectability using the conditional probabilities derived from our study. We evaluated 41 strategies comprising strategies based on single tests and combinations of two tests. As both correct diagnosis and resectability assessment are crucial for an adequate treatment plan, 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 conditional probabilities to perform a cohort simulation. We evaluated the impact of a reduced initial prevalence in a one-way sensitivity analysis. As numbers in some sub-branches were small, we also explored the degree of uncertainity of our result by means of probabilistic sensitivity analysis using Monte-Carlo-simulation.

Results: In the base-case analysis, the best test performance was achieved with MR alone, classifying 79% of patients correctly in benign, malignant/resectable, and malignant/unresectable. The combination of PET and EUS had the worst performance with only 61% of patients correctly classified. The choice of the best strategy depends on prevalence. If the prevalence is less than 23%, US followed by CT has the highest expected value of correctly classified patients. For a prevalence between 23% and 53%, MR followed by ERCP is optimal. Probabilistic sensitivity analysis showed robust results. In a MC-simulation with 10,000 samples, the probability of beeing the optimal strategy was 57% for MR alone (man accuracy 78%, standard deviation 3.1%) and 24% for MR followed y ERCP (mean accuracy 76%, standard deviation 3.3%)

Conclusion: Among the evaluated strategies, we suggest performing MR alone as first choice for patients with suspected pancreatic cancer in the setting of a specialized medical center with high prevalence of PC. Our model will be expanded to perform a cost-effectiveness analysis.

POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISON MAKING

QUALITY IMPROVEMENT STRATEGIES FOR HYPERTENSION: A SYSTEMATIC REVIEW

Walsh J[1], McDonald K[2], Shojania K[1], Sundaram V[2], Lewis R[2], Nayak S[2], Mechanic J[2], Owens D[3] and Goldstein M[2]

[1]University of California, San Francisco, San Francisco, CA; [2]Stanford University, Stanford, CA; [3]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 median effects on changes in blood pressure or changes in the percentage of individuals 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 mmHg (interquartile range (IQR): 1.5, 11.0) and 2.1 mmHg (IQR: -0.2, 5.0), 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 mmHg (IQR 4.2, 14.0) and 4.2 mm Hg (IQR 0.2, 6.8), and median increases in %GOAL for SBP and DBP of 21.8% (IQR: 9.0, 33.8) 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 median reductions in SBP and DBP of 8.0 mm Hg (IQR: 2.5, 12.3) and 1.8 mm Hg (IQR -0.1, 4.5), and median increases in %GOAL for SBP and DBP of 25.1% (IQR: 17.0, 34.2) and 2.0% (IQR: 1.6, 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 DECISON MAKING

QUALITY OF LIFE IMPACT BY MELANOMA AS MEASURED BY UTILITIES

Bendeck S, Hadley J, Bonaccorsi P, Brown K, Lawson D, Murray D, Washington C and Chen S

Emory University, Atlanta, GA

Purpose: The goal of our study was to evaluate the impact of melanoma on patients' quality of life by calculating utility scores.

Methods: Utilities 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 (< one year ago) or old (> one year ago). The patients' utility score for their melanoma health state was elicited after they were shown a brief presentation of its 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 109 patients (mean age 49.9 years, 56.9% female, and 100% Caucasian) were recruited from our melanoma clinics and database. The following table shows mean utilities, standard deviations, and number of participants for each melanoma health state:

	N	Mean Utility (SD)
Stage I		
New (diagnosis one year ago)	14	0.93 (0.099)
Old (diagnosis ? one year ago)	69	0.93 (0.11)
Stage II		
New	3	0.97 (0.058)
Old	8	0.87 (0.16)
Stage III		
New	7	0.52 (0.31)
Old	8	0.89 (0.13)

Mean utilities for new vs. old diagnoses within the same stage were only significantly different for Stage III ($p = 0.010$). There was also a significant difference in the utilities across new melanoma health states ($p < 0.001$); post hoc analysis demonstrated that the significant differences lie between Stages 1 and 3 ($p < 0.001$) and Stages 2 and 3 ($p = 0.002$). There was no significant difference in the utilities across old melanoma health states.

Conclusions: Although the number of subjects in this preliminary study is small, these results suggest that Stage I melanoma has a relatively small QOL impact. QOL impact increases significantly with stage for new (II vs III and I vs III), but not old diagnoses. Increasing time from initial diagnosis did not significantly affect QOL impact for Stages I and II; however, it did lessen the impact for Stage III. Further work needs to be done to obtain more data, especially for new Stage II diagnoses.

POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISON MAKING

COMPARING SIDE EFFECTS & PERCEIVED MEDICATION EFFECTIVENESS USING NON-STEROIDAL ANTI-INFLAMMATORY DRUGS AND CYCLO-OXYGENASE TYPE II INHIBITORS IN THE TREATMENT OF OSTEOARTHRITIS

Mahajan S[1], Anstrom K[1], Kauf T[1], Chang J[2], Omar M[2], Kahler K[2] and Schulman K[1]

[1]Center for Clinical and Genetic Economics, Durham, NC; [2]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—diarrhea, nausea/vomiting, heartburn, stomach pain, headache, and dizziness.

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 1 decimal places. Using this new group of propensity-matched COX-2 inhibitor and NSAID users, we compared side effect burden and perceived medication effectiveness.

Results: Prior to matching COX-2 inhibitor users (n = 410) were older, were more likely to be unemployed, experienced more pain, and missed more hours of work/activity due to their osteoarthritis compared to NSAID users (n = 1409; all $p < 0.02$). 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 396 matched pairs of COX-2 inhibitor and NSAID users which revealed no significant differences in demographics or disease severity variables. Propensity matched COX-2 inhibitor users experienced fewer side effects than NSAID users ($p = 0.04$), particularly headaches ($p < 0.0001$). This is significant because headaches were a strong predictor of lowered productivity in a separate regression ($p < 0.0001$). 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 DECISON MAKING

PROSPEQT: A NEW PROGRAM FOR COMPUTER-ASSISTED UTILITY ELICITATIONS

Bayoumi A[1] and Dale W[2]

[1]St. Michael's Hospital, University of Toronto, Toronto, ON, Canada; [2]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; random ordering 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 an 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 FT elicitations, participants took a median of 30 (interquartile range 12 to 88), 15 (7 to 54), and 28 (12 to 87) seconds, respectively. The initial SG utility elicitation took 136 (40 to 233) seconds compared to the final SG elicitation, which took 17 (9 to 48) seconds. In 2.8% of responses, participants retracted a response. ProSPEQT is a web-based program for utility elicitations 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 elicitations.

POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISON MAKING

ACCURACY OF DIAGNOSTIC CODES: ONLY AS GOOD AS THE DIAGNOSIS

Singer M[1], Barzilai K[1], Bonomo R[2] and Jakubowycz M[2]

[1]Case Western Reserve University School of Medicine, Cleveland, OH; [2]Louis Stokes Cleveland VA Medical Center, Cleveland, OH

PURPOSE: Retrospective analyses rely on the accuracy of diagnostic codes, yet it is unclear in some cases, such as outpatient pneumonia, that the diagnosis itself is reliable. We examined the accuracy of outpatient diagnosis of pneumonia in a Veterans Affairs facility.

METHODS: We used an encounter database to identify patients seen at a Cleveland area VA facility during 2001-2002 with a primary diagnosis of pneumonia, who were treated as outpatients. Patients were included if they filled a prescription for an oral antibiotic within 3 days of the index visit. Patients were excluded if they were hospitalized or filled an antibiotic prescription in the last 30 days, or had already been evaluated for the current episode prior to the index visit. Patient records were abstracted by a nurse, recording results of chest x-ray (CXR) and the following symptoms: cough, shortness of breath, pleuritic chest pain, temperature>38, crackles or rales on auscultation. Pneumonia was considered to be "probable" if CXR positive, or unclear with >=2 symptoms; "possible" if unclear CXR and <2 symptoms, or no CXR and >=2 symptoms; "unlikely" if negative CXR, or no CXR and <2 symptoms. Chi-square tests, t-tests were used to compare patients with negative CXR to others.

RESULTS: 106 cases were eligible for the study. The mean age was 65 years. Interrater reliability showed agreement in at least 20/21 cases for result of CXR and each of the symptoms of pneumonia. 23/106 patients had a major comorbidity. 88/106 patients had a CXR, with 28 positive, 10 unclear and 50 negative. Likelihood of pneumonia: 31% probable, 17% possible, 52% unlikely. There were no statistically or clinically significant differences between patients with a negative CXR and everyone else. Mean number of pneumonia symptoms was 1.9 in the CXR negative group and 2.2 for everyone else (p=.08). Treatment for patients with negative CXR was 45% quinolones, 39% macrolides vs. 51% quinolones, 37% macrolides for others. A comorbidity was noted in 22% of CXR negatives and 21% of others.

CONCLUSIONS: In this study of VA outpatients diagnosed and treated for pneumonia, approximately half were unlikely to have pneumonia. Retrospective analyses of pneumonia should be cautious in their conclusions as the diagnostic codes may accurately reflect the physician's diagnosis, but may not reflect the true diagnosis.

POSTER SESSION - CLINICAL STRATEGIES; JUDGMENT AND DECISON MAKING

MARKOV STATE TRANSITION MODELS BASED ON AUTOREGRESSIVE MULTINOMIAL LOGISTIC REGRESSION FOR THE PREDICTION OF CHANGES IN SLEEP STRUCTURE INDUCED BY AIRCRAFT NOISE - THE GERMAN AEROSPACE CENTER STUDY

Basner M[1] and Siebert U[2]

[1]German Aerospace Center (DLR), Cologne, Germany; [2]Massachusetts General Hospital, Harvard Medical School, Boston, MA

Purpose: By dividing polysomnographic recordings into intervals of 30 sec, human sleep can be classified in six distinct states: Awake, stages 1&2 (light sleep), stages 3&4 (slow wave sleep, SWS), and rapid eye movement (REM) sleep. The sleep states differ in their contribution to the restorative power of sleep. Environmental noise is a potential disruptor of the sleep process and may cause changes in the structure of sleep. The goal of this study was to predict changes in total sleep structure depending on sound pressure levels and time patterns of aircraft noise events (ANE).

Methods: In four laboratory studies with 128 subjects lasting from 1999 to 2003, the Institute of Aerospace Medicine of the German Aerospace Center (DLR) investigated the influence of aircraft noise on human sleep. Quiet baseline nights of 125 subjects were used to build and validate a model for the simulation of noise-free nights based on autoregressive multinomial logistic regression. Data of 33,000 ANE and related events were used to incorporate the effects of ANE on transition probabilities. Three noise scenarios (see results) with constant maximum sound pressure levels of 65 dB(A) were compared regarding their impact on total sleep structure.

Results: A second order autoregressive model fit the validation criteria best. Comparison of mean sleep stage fractions of baseline nights and 10,000 first-order Monte Carlo trials showed good agreement (model vs. raw data: Awake -0.7%, S1 +27.5%, S2 +0.5%, S3 +2.5%, S4 -8.8%, REM -1.5%). Noise restriction between 11 pm and 5 am (scenario 2: 59.3 min SWS, 47.2 min awake) revealed clear benefits compared to unrestricted traffic (scenario 1: 43 min SWS, 63.5 min awake), although these benefits were reduced if the traffic that formerly took place between 11 pm and 5 am was rescheduled to the time before and after the silent period (scenario 3: 58 min SWS, 54.2 min awake).

Conclusions: It was possible to validly reproduce key features of noise-free baseline nights with a Markov state transition model based on multinomial autoregressive logistic regression. The extension of the model based on extensive data on the reactions to ANEs allows for the comparison of sleep structures induced by different noise patterns and may serve as a valuable tool for structuring nocturnal air traffic and for political decision making.

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

COST-EFFECTIVENESS OF CHILDHOOD ADENOTONSILLECTOMY; A RANDOMISED COMPARISON WITH WATCHFUL WAITING

Buskens E[1], Van Staaij B[1], Van den Akker J[2], Hoes A[1] and Schilder A[2]

[1]University Medical Center Utrecht, Utrecht, Netherlands; [2]University Medical Center Utrecht, Utrecht,, Netherlands

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 milder symptoms of throat infections or adenotonsillar hypertrophy. Methods Economic evaluation alongside an open randomised controlled trial. Setting: Multi-center: 21 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 measures: 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 Ευρο Σανσ∀1,196 as opposed to Ευρο Σανσ∀804 in the watchful waiting group (49% increase; 100% certain). During a median follow-up period of 22 months children in the adenotosillectomy group experienced fewer episodes of fever 0.21 (95% CI 0.54 to -0.12), throat infections 0.21 (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 Ευρο Σανσ∀2,333, Ευρο Σανσ∀1,444 and Ευρο Σανσ∀788 respectively. Conclusion For the majority of Dutch children currently undergoing adenotonsillectomy, i.e., with relatively mild symptoms or adenotonsillar hypertrophy, operation results in a significant increase in costs without relevant clinical benefit. Additional research is required to identify subgroups in which operation may be worthwhile.

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

EFFECTIVENESS OF ADDING ANTICOAGULATION TO MECHANICAL PROPLYLAXIS FOR PREVENTING DVT IN CRANIOTOMY

Sonnad S, Danish S, Burnett M and Stein S

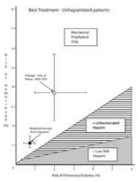
University of Pennsylvania, Philadelphia, PA

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 3%. 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.

CONCLUSION: The clear advantage of heparin prophylaxis in general surgery is offset by increased risk of intercranial 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

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

Sanders G[1], Lin D[1], Chan J[2] and Gould M[3]

[1]Duke, Durham, NC; [2]Stanford University, Stanford, CA; [3]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 recommended downstream decisions and to perform sensitivity analyses. The diagnosis and management of solitary pulmonary nodules (SPN) is one such clinical area in which multiple combinations 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 clinically plausible sequences of 5 diagnostic interventions 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

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=7538) 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, 10-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 (¥ÄR2 <0.035), evidencing satisfactory DIF validity.

Conclusions: In this study, PF and MH items showed no 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.

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

THE IMPACT OF AIDS-RELATED EVENTS AND NON-AIDS SERIOUS ADVERSE EVENTS ON HEALTH-RELATED QUALITY OF LIFE IN A MULTINATIONAL TRIAL OF ANTIRETROVIRAL THERAPY

Araki S[1], Guh D[2], Singer J[3], Joyce V[4], Sundaram V[5], Munakata J[4], Hill A[4], Sun H[2], Griffin S[6], Sanders G[7], Bayoumi A[8], Holodniy M[5], Brown S[9], Sculpher M[6], Yu W[4], Owens D[5], Anis A[3] and Investigators t[10]

[1]Stanford University, Stanford, CA; [2]St. Paul's Hospital, Vancouver, BC, Canada; [3]University of British Columbia, Vancouver, BC, Canada; [4]VA Palo Alto Health Care System, Menlo Park, CA; [5]VA Palo Alto Health Care System, Palo Alto, CA; [6]University of York, York, United Kingdom; [7]Duke University, Durham, NC; [8]St. Michael's Hospital, University of Toronto, Toronto, ON, Canada; [9]Bronx VA Medical Center, Bronx, NY; [10]VA Connecticut Health Care System, West Haven, CT

Purpose: In a three-nation randomized trial of management strategies (drug-free period and/or treatment intensification) for advanced HIV disease, we evaluated the impact of AIDS-related events (ARE) and non-AIDS related serious adverse events (SAE) on health-related quality of life.

Methods: From the 282 patients enrolled as of April 2004 in OPTIMA (Options in Management with Antiretrovirals - an ongoing study for patients with virologic failure and multi-drug resistant virus), we identified: 1) patients with ARE; 2) patients with SAE; and 3) a random control group of patients with no clinical events. We assessed utilities using the Health Utilities Index Mark 3 (HUI3) and EuroQol (EQ-5D) at baseline and at regularly scheduled visits thereafter. Concurrent events and those with missing utility data (at visit just before or after event) were excluded from the analysis. We performed t-tests to evaluate differences between pre- and post-event scores for each group.

Results: Excluding 38 deaths, there were 91 ARE in 53 patients (e.g., esophageal candidiasis, pneumocystis carinii pneumonia) and 205 SAE in 88 patients (e.g., chest infection, anemia); utility data were available for a total of 103 non-concurrent events. Mean utility scores did not change significantly from baseline (HUI3 0.58; EQ-5D 0.68) 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 (-0.09, p = 0.07) and SAE (-0.09, p = 0.02), while mean EQ-5D scores fell slightly (ARE: -0.04, p = 0.43; 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.

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

ASSESSING RESPONSIVENESS OF THE SELF-ESTEEM AND RELATIONSHIP QUESTIONNAIRE IN MEN WITH ERECTILE DYSFUNCTION TREATED WITH VIAGRA(r) (SILDENAFIL CITRATE) IN AN INTERNATIONAL, DOUBLE-BLIND, PLACEBO-CONTROLLED TRIAL

Cappelleri JM[1], Glina S[2], Duttagupta S[3], Sherman N[3], Stecher V[3], Siegel R[3] and Tseng L[3]

[1]Pfizer Global Research & Development, Groton, CT; [2]Instituto H. Ellis, Sao Paulo, Brazil; [3]Pfizer Inc, New York, NY

Purpose: Using a stringent definition of a responder, we assessed responsiveness of the Self-Esteem And Relationship (SEAR) questionnaire to changes in erectile function after treatment with Viagra.

Methods: A 12-week, double-blind, placebo-controlled, flexible-dose (25, 50, or 100 mg), multicenter, international trial was conducted in men aged ?18y with clinically documented ED and baseline score ?75 on the SEAR questionnaire Self-Esteem subscale. Patients who had taken another phosphodiesterase type 5 inhibitor, >6 doses of Viagra, or were receiving nitrates or a nitric oxide donor were excluded. A treatment (Viagra or placebo) responder was defined as someone who responded "yes" to 2 end-of-treatment (EOT) questions indicating improved erections and sexual intercourse and had an EOT score ?22 on the Erectile Function domain of the International Index of Erectile Function. Change scores from baseline on the 2 domains (Sexual Relationship, Confidence), Confidence domain subscales (Self-Esteem, Overall Relationship), and Overall score of the SEAR questionnaire were analyzed between responders and nonresponders using an ANCOVA model (adjusting for center and baseline score).

Results: A total of 149 (mean ± SD: age, 54±12y; ED duration, 4.7±4.7y) and 151 patients (mean ± SD: age, 56±11y; ED duration, 4.3±4.5y) were randomized to placebo and Viagra, respectively. Overall, 169 of the intent-to-treat population (55 placebo, 114 Viagra) were treatment responders and 114 (84 placebo, 30 Viagra) were nonresponders. Responders had significantly greater improvements from baseline on all SEAR components (P <0.0001); nonresponders had no change or significant decrements. Adjusted mean change from baseline scores for Sexual Relationship (46.3), Confidence (48.9), Self-Esteem (50.5), Overall Relationship (45.5), and overall score (47.3) were significantly higher (P <0.0001) than corresponding nonresponder change scores (-5.21, -1.8, -1.0, -3.6, and -3.7, respectively).

Conclusions: Regardless of treatment, responders who reported improved erectile functioning also showed improvement in self-esteem, confidence, and relationships, whereas nonresponders who lacked improvement in erectile functioning showed no such improvements in these psychosocial factors. The results confirm the accuracy of the SEAR questionnaire for measuring changes in psychosocial factors as they relate to changes in erectile functioning.

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

A LITERATURE REVIEW OF THE RELIABILITY, CONSTRUCT VALIDITY, AND RESPONSIVENESS OF HEALTH UTILITY MEASURES IN HIV POPULATIONS

Baker F

VA Palo Alto Health Care System, Menlo Park, CA

Purpose: In order provide guidance on which of six different health utility measures to use when assessing the desirability of a health state, research that used these measures within HIV populations was reviewed.

Methods: Ovid and PubMed were used to find all peer reviewed research articles (from 1966 to Jan 2004) in which people with HIV were asked to assess the desirability of their own health state using at least one of six health utility measures: Rating Scale (RS), Standard Gamble (SG), Time Tradeoff (TTO), Health Utility Index (HUI), EuroQol (EQ-5D), or Quality of Well-Being Scale (QWB). Information concerning the reliability, validity, and responsiveness for these six health utility measures was extracted.

Results: The review yielded 36 articles. No articles were found in which the retest reliability or the internal consistency of these measures were specifically tested. The difference in health utility between people classified as having HIV versus AIDS depends on the measure (see table). The RS and EQ-5D seem to be responsive to changes in health. There is a decent amount of information available on the construct validity of the RS, EQ-5D, and QWB, but there is little information on the SG, TTO, and HUI. The RS, EQ-5D, and QWB all correlate reasonably well with health status measures and markers of disease progression, while the SG and TTO do not.

Conclusions: Evidence for the reliability of these measures within HIV populations is desperately needed. Whether the measures differ in validity ultimately rests on both empirical evidence, as presented in this review, and rational argument for whether they are tapping into the construct of interest. Based on the available empirical evidence within HIV populations, the RS, EQ-5D, and QWB seem to have the most desirable measurement qualities.

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

HORMONE THERAPY IN 2003: WOMEN'S DECISIONS AND RATIONALE

French L, Holmes-Rovner M, Smith M and Holtrop J

Michigan State University, East Lansing, MI

Purpose: To survey postmenopausal women in primary care practice regarding decision-making about hormone therapy (HT) after publication of the Women's Health Initiative results.

Methods: Questionnaires (n=133) including open-ended questions regarding women's HT decisions were collected August through October 2003 in one family practice waiting room. Women born 1933-1952 attending for office visits were invited to participate. Questions addressed current views about HT, change in views in the prior year, personal rationale for taking or not taking HT, self-rated health status, symptom bother, and demographic variables. A coding scheme for qualitative analysis was developed. Two investigators coded the responses to each question. Chart review provided HT rates for the clinic population in the study window and the previous 2 years.

Results: Rate of HT use in the practice decreased to 26% at the time of survey from 49% during the previous two-year period. Survey respondents were predominantly white, healthy, well educated, highly symptomatic, long-term (>5years) HT users, and currently taking HT at a higher rate (36%). The predominant belief pattern was that use of HT is riskier than previously believed. Less common were confirmation of previous personal skepticism, rejection of new recommendations, and belief that HT is beneficial with apparent lack of awareness of new research findings.Of 50 women who had stopped HT in the previous year 12 had restarted. Restarters were struggling with their decision to continue or not. Some women who had quit were considering restart. A pattern of grief was articulated by some quitters. Women who had continued HT without stopping were a relatively young and healthy group, but the majority had used HT> 5 years. Expressions of concern/fear regarding their HT use were present in the responses of former users, quitters, restarters, and continuers.

Conclusions: This sample of care-seeking women remains concerned about pros and cons of HT. Future studies should account for both symptoms and chronic disease risk vs. benefit as factors in decision-making.

HT Use Category	Number	Moderate/Severe Hot Flashes (%)	Good/Excellent Health Status (%)
Never users	18	40	93
Former users	18	26	64
Recent quitters	38	55	89
Restarters	12	40	58
Continuous users	28	4	92

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

IDENTIFICATION OF MEN WHO WANT HELP WITH PROSTATE CANCER TREATMENT DECISIONS

Knight S[1], Chren M[1], Sands L[2], Grollman J[3], Nadler R[4] and Bennett C[4]

[1]San Francisco VA Medical Center, San Francisco, CA; [2]Purdue University, West Lafayette, IN; [3]Rosalind Franklin University of Medicine and Science, North Chicago, IL; [4]Feinberg 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 their 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. *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 make decisions without physician input. White men who were in better health (i.e., better physical function, fewer symptoms) and were more optimistic preferred greater autonomy in decision making(ps=0.001-0.05). When men saw 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-0853 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 = 2,743) were asked to be interviewed in-person, and 1,367 (50%) accepted. The respondents were representative of the target population in terms of sex and age. The respondents were asked to imagine that 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 while 12% were uncertain. The proportions consenting were 39%, 52%, 56%, 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 81% 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.

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

WHAT WOULD MAKE THE SOCIETY AND THE ANNUAL MEETING BETTER? A SURVEY OF THE SOCIETY FOR MEDICAL DECISION MAKING MEMBERSHIP

Krahn M[1], Witkos M[1], Sanders G[2], Sugano D[3] and Cantor S[4]

[1]University Health Network, Toronto, ON, Canada; [2]Duke, Durham, NC; [3]Schering Plough Inc., Kenilworth, NJ; [4]The University of Texas M. D. Anderson Cancer Center, Houston, TX

Purpose: The Board of Trustees of the Society for Medical Decision Making (SMDM) commissioned a survey to assess member attitudes in order to guide planning for future annual meetings, broaden the membership and enhance the Society's relevance.

Methods: We generated survey items by soliciting selected board members' opinions regarding attributes that might improve the attractiveness of the meeting and the Society. The survey involved 54 questions about meeting attendance, content and quality of previous meetings, attitudes toward the Society, and factors which would improve meeting attendance. E-mails soliciting participation in the web-based survey were sent to the Society membership on three occasions over 4 weeks. We employed descriptive statistics, and stratified analyses by demographic characteristics to analyze the results, as well as compiling free text responses by theme.

Results: 297 of 918 members (32%) responded. In comparison with Society members, more survey respondents were women (40% vs. 31%, $p = 0.0023$), and fewer physicians (44% vs. 51%, $p = 0.05$). Most respondents were between 30 - 45 years old (57%), male (60%), from the US (79%), and have an academic primary affiliation (80%). Responders 'strongly agreed' that the meeting is scientifically useful (75%), fun (47%), and aligns with their research interests (49%). 58% of responders 'strongly agreed' that they strongly support the Society. 58% and 47% 'strongly/somewhat' agreed that their support is increasing with time, and the Society is their primary affiliation, respectively. Physicians stated they were more likely to attend if the meeting was more clinically relevant ($p = 0.0005$). Members from outside North America would be more likely to come the meeting if they had to travel less ($p = 0.0009$). The most frequent free text comments about the meeting were that it should be more practical, policy, and clinically relevant (25), it is great as is (16), should be less expensive (15) and better structured for networking (14). The most frequent comments about the Society were that it should be more practical, policy, and clinically relevant (20), and more inclusive/open (15).

Conclusions: Members in general, have a positive attitude towards the SMDM and the annual meeting. Issues of relevance, high cost, and inclusiveness were the predominant issues identified by the survey responders and will be targets of the current board and meeting chairs.

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

Norris 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 73.8 (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 >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.

[1] Ware JE Jr. SF-36 health survey: manual and interpretation guide. Boston: The Health Institute, New England Medical Center; 1993.

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

BREAST CANCER PREVENTION: A FRAMEWORK FOR DECISION MAKING

Ozanne E[1], Adduci K[2], Annis C[2] and Esserman L[2]

[1]Massachusetts General Hospital, Boston, MA; [2]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%. At 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.

Conclusions: Initial 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 preventive 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 - UTILITY THEORY; HEALTH ECONOMICS; PATIENT & PHYSICIAN PREFERENCES; SIMULATION; TECHNOLOGY ASSESSMENT

ACCEPTABLE FAILURE RATES FOR ANTIBIOTIC THERAPY OF CENTRAL VENOUS CATHETER (CVC) ASSOCIATED BACTEREMIA

Perencevich E[1], Kaye K[2], Strausbaugh L[3], Bradham D[1], Fisman D[4] and Harris A[1]

[1]VA Maryland Health Care System and University of Maryland School of Medicine, Baltimore, Baltimore, MD; [2]Duke University Medical Center, Durham, NC; [3]Portland VA Medical Center, Portland, OR; [4]Drexel University, School of Public Health, Philadelphia, PA

Purpose: Antimicrobial resistance is an increasing public health problem. Decreasing the duration of antibiotic therapy is an attractive strategy for delaying emergence of antimicrobial resistance. The paucity of data about optimal treatment durations for most clinical infections hinders the adoption of this approach and encourages overtreatment. Using a case scenario, this study used the methodology developed for contingent valuation analysis to identify acceptable failure rates of infectious diseases consultants (IDCs) for treatment of CVC-associated bacteremia. Quantifying an acceptable failure rate may help reduce excess treatment duration and improve numerical statements in physician risk communication.

Methods: The authors developed a case scenario for a representative patient who developed uncomplicated CVC-associated coagulase-negative staphylococci bacteremia. The patient received standard-of-care therapy including catheter replacement and intravenous vancomycin therapy. In August 2003, the Infectious Diseases Society of America's Emerging Infections Network (EIN) distributed the case description to its members. The EIN gave individual members one of 10 failure rates and asked if they would accept or reject the given value. The authors used logistic regression to evaluate the relationship between specific failure rates offered to EIN members and their willingness to accept them.

Results: Overall, 374 (54%) of 687 EIN members responded to the questionnaire. Mean years in practice for responders was 15.9 years. Logistic regression analysis determined that the median acceptable failure rate was 6.8%. Thus, half the IDCs would have found a failure rate of 6.8% to be acceptable. 75% of IDCs would have found a failure rate of 1.6% acceptable and 25% of IDCs would have found a failure rate as high as 11.9% acceptable. Adjusting for IDCs years in practice did not alter these results. As expected, the acceptable failure rate and ranges in this study were all lower than those reported in a previous study using the same methodology in diabetic foot osteomyelitis (Perencevich, et al., Clin Infect Dis Feb 15, 2004)

Conclusions: The quantified physician acceptable failure rates when combined with patient preferences to avoid treatment failure and duration-specific therapy failure rates might assist in selecting an optimal treatment duration. Thus, the methodology used may prove useful in delineating acceptable treatment failure thresholds, a first step in reducing durations of antimicrobial therapy. In addition, the numerical failure rates may help improve physician risk communication.

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

THE HAPPINESS GAP: A MATTER OF CONTEXT?

Pond H[1], Fagerlin A[2], Goldstein C[2], Smith D[2], Loewenstein G[3] and Ubel P[2]

[1]VA Ann Arbor Health Care System, Ann Arbor, MI; [2]University of Michigan, Ann Arbor, MI; [3]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 experience with 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 240 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 (context condition), or only 1 of these 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 = 45.15) 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 emphysema and thus cannot readily account for the happiness gap.

POSTER SESSION - UTILITY THEORY; 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

Schackman B and Teixeira P

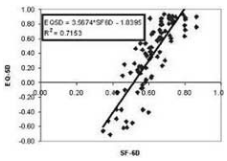
Weill Medical College of Cornell University, New York, NY

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 were -0.33 (0.33) for the SF-6D, and -0.14 (0.43) and -0.11 (0.26) 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[1], Waickus C[2], Baker E[2], Cortez E[2] and McNutt R[2]

[1]University 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 (SDM) 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 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: 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 not enrolled in the course completed the instruments. Analyses reported here compare students' (a) cognitive skills, (b) attitudes toward shared decision making, and (c) opinions of the course's usefulness as a clinician and as a patient, compared with anatomy and behavioral science.

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 felt both patient and physician decision participation were 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 SDM; M1 students further rated behavioral science as more useful than SDM. 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. Course evaluations and focus groups revealed dissatisfaction with course process and conflicting messages about decision making from other M1 courses.

Conclusions: Students were dissatisfied with the first offering of the M1 course, and did not perceive it as useful. Nevertheless, students improved MDM-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

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

Weinfurt K[1], Depuy V[1], Castel L[1], Seils D[1], Sulmasy D[2], Schulman K[1] and Meropol N[3]

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

Purpose: Patients' expectations of benefit from phase I trials often far 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, patients' profiles were 28.46 points higher in level and displayed greater scatter (difference in SD, 15.15) than their physicians' profiles. The average correlation between the shapes of patients' and physicians' profiles was 0.03, 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 likelihood 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

IMPACT OF DISCUSSION ON UTILITY VALUES ELICITED IN A GROUP SETTING

Stein K[1], Crabb T[1], Brazier J[2], Ratcliffe J[2], Milne R[3] and Round A[1]

[1]Peninsula Medical School, Exeter, United Kingdom; [2]School of Health and Related Research, Sheffield, United Kingdom; [3]University 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 group members were then discussed for five to ten minutes and participants given the opportunity to revise their initial preference following discussion. The number of changes made and their impact on the range and summary utility values for the group were calculated. The importance of group discussion was explored with participants in qualitative interviews.

Results: 40 health states relating to seven specific conditions were valued, giving 445 individual results. Mean utility values ranged -0.27 to 0.98. 14 (3.1%), changes were made in values for 12 scenarios (30%) by seven individuals. One individual changed five values. Changes ranged from -0.075 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 value discussion highly.

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

Elkin E, Saranchuk J, Touijer A, Scardino P, Eastham J and Kattan M

Memorial Sloan-Kettering Cancer Center, New York, NY

PURPOSE: The most favorable outcome that can be achieved after radical prostatectomy (RP) is complete tumor resection with full recovery of continence and potency. Risks of erectile dysfunction, urinary incontinence and disease recurrence have been described, but separately these estimates do not adequately inform patients of the likelihood of returning to their preoperative functional status while remaining cancer-free. The purpose of this study was to estimate the probability of achieving optimal clinical outcome—absence of disease with full recovery of potency and continence—after RP.

METHODS: We used a Markov state-transition model to simulate clinical outcomes in the first 4 years following RP. The primary endpoint was time until full recovery of continence and potency with no evidence of disease. The model consisted of three mutually exclusive health states: (1) incomplete recovery (impotent or incontinent with no evidence of disease); (2) complete recovery (potent and continent with no evidence of disease); and (3) disease recurrence, as demonstrated by elevated serum prostate specific antigen (PSA). All patients started in the incomplete recovery state. Disease recurrence was possible from both the incomplete and complete recovery states. Transition probabilities were estimated from a clinical database of 647 men who underwent RP for clinical stage T1 to T3 prostate cancer. The model used a one-month Markov cycle length.

RESULTS: Complete recovery was achieved in 30% of men at 12 months, 42% at 24 months, and 53% at 48 months postoperatively. In the first 4 years following surgery, the cohort spent an average of 17.6 months in the optimal outcome state. Fewer than 10% of patients experienced disease recurrence by the end of 4 years.

CONCLUSIONS: Our results suggest that a majority of prostate cancer patients can achieve full potency, continence, and cancer control by 48 months after RP. Information about the individual risks of impotence, incontinence and disease recurrence may be less useful to patients than estimates of the likelihood of achieving the optimal post-operative outcome and the expected duration of that outcome. Markov modeling may provide valuable information in clinical situations where multiple dynamic outcomes are relevant to the decision-maker, and where estimating the joint probability and duration of these outcomes is not straightforward using standard statistical methods.

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 and 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 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

THE CLINICAL BENEFITS AND COST-EFFECTIVENESS OF A HYPOTHETICAL CATHETER-BASED STRATEGY FOR THE DETECTION AND TREATMENT OF VULNERABLE CORONARY PLAQUES, A DECISION ANALYTIC APPROACH

Bosch J[1], Beinfeld M[2], Muller J[2], Brady T[2] and Gazelle G[2]

[1]Erasmus MC, Rotterdam, Netherlands; [2]Massachusetts General Hospital, Boston, MA

Purpose: Currently, stenotic coronary plaques are detected and treated without considering their "vulnerability" to rupture. In this study, we evaluated the potential clinical benefits and cost-effectiveness of a hypothetical catheter-based strategy for the detection and treatment of vulnerable/high-risk plaques in patients with coronary artery disease using decision analysis.

Methods: In a new hypothetical strategy, vulnerable coronary plaques are detected with a catheter-based test and treated with a drug-eluting stent, regardless of degree of stenosis. We developed a Markov-decision model to compare the new strategy with current practice (angioplasty followed by stent placement in stenotic arteries). Monte Carlo simulations were performed from a societal perspective, costs were converted to year 2003 U.S. dollars, and future costs and outcomes were discounted at 3%. Sensitivity analyses were performed to evaluate the effect of assumptions such as the prevalence of plaques and treatment effect.

Results: In 60-year old male patients with coronary stenoses the new strategy would be less expensive and more effective than current practice ($43,103 vs. $44,003 and 10.17 vs. 9.86 quality-adjusted life years, respectively). The benefits of the new strategy were robust in sensitivity analyses (e.g., if the prevalence of vulnerable plaques in this patient group was 50% or more and the sensitivity and specificity of the new test were at least 0.80).

Conclusion: The detection of non-stenotic vulnerable plaques with a catheter-based test followed by their treatment with a drug-eluting stent could be a less expensive and more effective strategy than current practice in patients with coronary artery disease. If applied to 1 million such patients in the US undergoing catheterization, the new strategy could add 310,000 quality-adjusted life years and save $826 million dollars per year.

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

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

Biesheuvel C[1], Siccama I[2], Grobbee D[1] and Moons K[1]

[1]University Medical Center Utrecht, Utrecht, Netherlands; [2]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; 95%CI: 0.64-0.82) than that of the logistic regression model (0.68; 0.59-0.77). The calibration of both models was similar, indicating similar amount of overoptimism.

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

Nieuwkerk P, 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- with follow-up measurements. 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 baseline- 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

Fagerlin A, Jankovic A, Langa K, Weir D, Willis R, Alexander G and Ubel P

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. 1010 participants answered four numeracy questions. We randomized subjects to one of 3 contexts: 1) numeric context: e.g. "What is 15% of 1000?," 2) health context: e.g. "If the chance of getting a disease is 15%, how many people out of 1000 would be expected to get the disease?," or 3) 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%-82% 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

AN INTERNET SURVEY OF PUBLIC PREFERENCES FOR CHARACTERISTICS OF HEALTH INTERVENTIONS

Prosser L1, Concannon T[1], Pearson S[1] and Hammitt J[2]

[1]Harvard Medical School and Harvard Pilgrim Health Care, Boston, MA; [2]Harvard School of Public Health, Boston, MA

Purpose: To measure the relative effect of specific intervention characteristics on willingness-to-pay (WTP) for coverage of health interventions and to assess the feasibility of an insurance coverage question for capturing such preferences.

Methods: An internet-based survey was conducted on a random sample of respondents whose characteristics were matched to those of the U.S. general population (n=1999). Respondents were asked their willingness-to-pay for an increase in their annual insurance premium to cover a hypothetical health intervention using dichotomous-choice double-bounded questions. Each respondent was randomized to one of four scenarios which simultaneously varied age (children or adult) and prevention (vaccination) or treatment and one of two scenarios comparing an intervention for a condition perceived to be involuntary (heart disease) with a similar intervention for a condition perceived to be voluntary (obesity). Scenarios also included information on efficacy, side effects, cost, size of the affected population, and cost-effectiveness that did not vary between scenarios. Survival analysis was used to estimate the median WTP for each additional coverage option both adjusted and unadjusted for respondent characteristics. Covariates included sociodemographic variables, type of health insurance, and familiarity with and self-assessed risk for conditions described in the hypothetical scenarios.

Results: Respondents were not willing to pay a significantly different amount for an intervention in children compared with adults or for a prevention compared with a treatment. They were, however, willing to pay more (p<.01) for an intervention for a condition perceived to be involuntary (heart disease) as compared with a similar intervention for a condition associated with voluntariness (obesity). Results were consistent using both adjusted and unadjusted models. Important covariates were voluntariness of condition, plan type, and total out-of-pocket expenses.

Conclusions: Perceived personal accountability for health conditions appears to influence values the public places on health interventions. In contrast to other studies, we did not find differences in willingness-to-pay associated with the age of a patient or when comparing prevention with treatment. Possible explanations are that our insurance coverage question was insufficiently sensitive to detect preferences among intervention characteristics or that vaccines may be associated with greater value than other preventive interventions. Further research on measuring public preferences for specific intervention characteristics could help inform public and private coverage decisions.

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

Farahati 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.

Methods: Two-part model were performed to predict the weekly working hours (after adjusting for other chronic disorders and socio-demographic) of diabetics among respondent to the Canadian Community Health Survey 2001. Logistics regression and multiple OLS analyses were used to predict the probability of working and weekly work 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 10% 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 reated 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 reated comorbidities/complications (all at p-values < 0.05). The predicted weekly work 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 and related comorbidities/complications were 17 and 30, respectively.

Conclusions: This is the first study to estimate the impact of diabetes on Labour market outcomes 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 are 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 Y[1], Han S[2] and Cantor S[1]

[1]The University of Texas M. D. Anderson Cancer Center, Houston, TX; [2]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 0.28 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, 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 pharmacoeconomic models would yield more accurate projections of the ICER and enhance decision making.

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

A SYSTEMATIC COMPARISON OF THE PERFORMANCE OF GOMPERTZ FUNCTION AND DECLINING EXPONENTIAL APPROXIMATION OF LIFE EXPECTANCY IN DECISION TREES

Siebert U 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 outcomes of decision-analytic Markov models, but their evaluation in decision trees is less 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 calculate undiscounted and discounted RLE and QALE. Results were compared to actuarial life table analysis representing the gold standard for age-dependent RLE and QALE. For the discounted basecase, we used a 3% annual discount rate. Bias was defined as percent deviation from the sum of RLE for ages 30-89. All analyses were performed separately for men and women. DSM and discount rates were varied in one-way sensitivity analysis.

Results: Both approximation methods underestimated the actual undiscounted RLE for both the additive and multiplicative model. Basecase results for men: for the multiplicative model, GPA (bias -4%) performed better than DEALE (-49%), whereas for the additive model, DEALE (-6%) was superior to GPA (-25%). Results for women showed similar patterns regarding magnitude and direction of bias. The use of time-independent disease-specific utility decrements yielded similar patterns for QALE. When varying 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.

Conclusions: 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

PREDICTING DERMATOLOGICAL UTILITIES: AN EXPLORATION OF CENSORED UTILITY DATA PREDICTION MODELS

Chen S[1], Sternberg M[2], Bayoumi A[3], Olkin I[4], Soon S[1], Cruz P[5], Sexton S[1], McCall C[1] and Goldstein M[6]

[1]Emory University, Atlanta, GA; [2]Emory University, School of Public Health, Atlanta, GA; [3]St. Michael's Hospital, University of Toronto, Toronto, ON, Canada; [4]Stanford University, Palo Alto, CA; [5]University of Texas at Southwestern, Dallas, TX; [6]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 states. Moreover, regression models based on linear regression (LR) are only valid only when certain assumptions are met. Violations of these assumptions can lead to biased prediction models. If the population under study rarely records utilities of 0 and 1 it may reasonable to disregard the ceiling or floor effect and use OLS. However, we have demonstrated that for most dermatological health states, utility scores will cluster close to the ceiling of 1.

PURPOSE: We explored several different regression methods to identify the best method to predict health utilities from a validated skin-specific health status measure, Skindex.

METHODS: We interviewed 250 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 trade-off 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 of 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: 1.0 - .9487). The MAPE and IQR were as follows: LR 0.042 (0.075); 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

A SUMMARY MEASURE OF POPULATION HEALTH

Stewart S[1], Woodward R[1] and Cutler D[2]

[1]Harvard Interfaculty Program for Health Systems Improvement, Cambridge, MA; [2]Harvard University and the National Bureau of Economic Research, Cambridge, MA

Purpose: We propose a method for quantifying non-fatal health that addresses concerns raised about existing measures by: 1) incorporating a range of symptoms and impairments from mild to severe, 2) basing disutility weights on current health ratings rather than counterfactual scenarios, and 3) accounting for multiplicative relationships between health problems. We then quantify the effects of diseases such as diabetes and heart diseases on health-related quality of life, based on how the diseases affect specific symptoms and problems.

Methods: Alternative disutility weights for health problems from the Quality of Well-Being Scale (QWB) are derived by examining their effects on global self-rated health status (SRHS) and time-tradeoff (TTO) ratings of current health. Data are from 1420 respondents age 45 to 89 in the Beaver Dam Health Outcomes Study. Ordered probit and OLS regression are used, with interaction terms testing for non-additive relationships. Disease weights are based on our weights derived for symptoms/problems, and on probit equations that examine the effect of each disease on each symptom/problem.

Results: Health problems with the greatest impact on quality of life were: limited ability to work, physical activity limitations, pain, and taking medications or following a prescribed diet. Other problems, such as sensory impairments, speech problems, sexual problems, and problems with weight or appearance, had little independent effect. Disutilities across eight domains were similar to comparable QWB scores, and mean overall scores were similar, with worse scores on our measure among those with problems in some domains. The TTO distribution was truncated (with 59% accepting no trade-offs), and SRHS was more closely related to objective measures of health and yielded higher disutility weights. Diseases with the worst disutility included heart and respiratory conditions, mood disorders, pain, sleep problems, ulcers, and type I diabetes.

Conclusions: Our method yields weights and scores similar to those from the QWB scale, with some important differences. Deriving disutility weights for diseases via their impact on specific symptoms/impairments allows examination of the mechanisms through which changes in mean quality of life with a specific disease occur over time. Replication of these methods in a larger, more representative sample is recommended as the basis for a summary measure of population health.

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

A RECIPE FOR INCOHERENCE: AVERAGING TIME-TRADEOFF OR STANDARD-GAMBLE UTILITIES ACROSS HEALTH ATTRIBUTES

Hazen G

Northwestern University, Evanston, IL

Purpose: Health outcomes are often specified using multiple health attributes. Procedures for assigning QALY coefficients to multiattribute health states include, for example, the Health Utilities Index and the EuroQol. For some cost-effectiveness analyses, the HUI or EuroQol attributes are not specific enough to address important issues. In such cases, modelers may be tempted to assess time-tradeoff or standard gamble utilities one attribute at a time, and then combine the assessed utilities by averaging over attributes. We point out why this procedure is mathematically incoherent, and show what errors in the inferred QALY coefficients may occur as a result.

Methods: We consider the case in which health status $q = (q_1,q_2)$ is described by two health attributes, and modelers wish to use TTO or SG techniques to assess utilities u_1, u_2, and then form a weighted average to obtain overall QALY coefficient $U(q) = k_1u_1+k_2u_2$, where k_1 and $k_2 = 1-k_1$ are the weights. We assume that when a subject specifies a TTO or SG response r_1 for a level q_1 of one attribute, s/he by default assumes the other attribute q_2 is at its best level, and vice-versa.

Results: Under the averaging model $U(q) = k_1u_1+k_2u_2$, the standard of taking $u_1 = r_1$ and $u_2 = r_2$ is no longer valid. If the modeler does so and then averages as just described to obtain the QALY coefficient $U(q)$, the resulting theoretical error in $U(q)$ is $DU(q) = (1-r_1)k_2 + (1-r_2)k_1$. The error $DU(q)$ is largest for attribute levels q_1,q_2 farthest below their best possible levels, and can be as large as 0.5 on a scale from 0 to 1 when attributes are equally weighted. The only way to avoid such errors is to replace the averaging rule by the multiplicative combination rule $U(q) = u_1u_2$.

Conclusions: Assessing time-tradeoff or standard gamble utilities one attribute at a time, and then averaging the assessed utilities to obtain an overall QALY coefficient is mathematically incoherent and can lead to large errors in the resulting QALY coefficients.

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?

Mrus J[1], Sherman K[2], Leonard A[2], Sherman S[2], Mandell K[2] and Tsevat J[1]

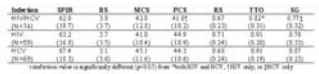
[1]Cincinnati VA Medical Center and University of Cincinnati, Cincinnati, OH; [2]University of Cincinnati, Cincinnati, OH

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 tradeoff (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 FACIT-SpEx 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 42% 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 (R^2=0.62); TTO with BS and SPIR (R^2= 0.24); and SG with infection type (HCV moninfection better), PCS, and BS (R^2=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 both for 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

VALUATION OF ARTHRITIS HEALTH STATES ACROSS ETHNIC GROUPS AND BETWEEN PATIENTS AND COMMUNITY MEMBERS

Souchek J[1], Byrne M[2], Kelly P[1], Richardson M[1], Pak C[1], Nelson H[1] and Suarez-Almazor M[1]

[1]Baylor College of Medicine/Michael E. DeBakey VA Medical Center, Houston, TX; [2]University of Pittsburgh, Pittsburgh, PA

PURPOSE: To examine differences in the valuation of health states by patients and community members from different ethnic backgrounds.

METHODS: We surveyed 193 community members identified by random digit dialing: 64 white (W), 65 African-American (AA) and 64 Hispanic (H). The patient sample included 198 individuals diagnosed with osteoarthritis (OA) and drawn sequentially from a health-provider institution clinic lists, 66 per ethnic group. Participants were interviewed face to face and asked to rate two different scenarios describing patients with arthritis (mild and severe) using visual analog scale (VAS), standard gamble (SG) and time trade-off (TTO). Differences were adjusted for cohort, age, age-squared, gender, and education.

RESULTS: The difference between the utility scores for mild OA and severe OA was significantly smaller for AA than W by the VAS, TTO, and SG methods. The difference between mild and severe states was smaller for H than W by 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 for mild OA. The public gave the severe OA state a higher preference score than patients did using the VAS 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 by VAS for older individuals. Education ameliorated the effects of other variables on TTO and SG scores.

CONCLUSIONS: Our findings show significant differences between ethnic groups in the valuation of health, with AA reporting less difference between the mild OA state and the severe OA state than W by the VAS, TTO and SG methods of valuating health states. H were less willing than W to risk death to move from a severe OA state to a mild OA state. Members of the public were less willing than patients to risk death to achieve perfect health. These differences suggest that in health decision-making, valuation of health states cannot be used interchangeably across ethnic groups.

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

THE GAP EFFECT: DISCONTINUITIES OF PREFERENCES AROUND DEAD

Stalmeier P[1], Busschbach J[2], Lamers L[3] and Krabbe P[1]

[1]University MC Nijmegen, Nijmegen, Netherlands; [2]Erasmus Medical Center, Rotterdam, Netherlands; [3] Erasmus MC Rotterdam, Rotterdam, Netherlands

Background: The assessment of negative values for health states considered to be worse than dead is a controversial issue. Objective: To investigate how health states are valued when they are close to dead.

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 health states were given by 3395 respondents. For each respondent, states were rank ordered by their VAS and TTO scores. For these rank ordered states, better than dead, equal to dead, and worse than dead preferences were defined. Differences between the valuations of adjacent rank ordered states for the VAS and the positive and negative TTO scores were calculated.

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

Wittenberg E[1], Divi N[1], Halpern E[1], Araki S[2], Prosser L[3] and Weeks J[4]

[1]Massachusetts General Hospital, Boston, MA; [2]Stanford University, Stanford, CA; [3]Harvard Medical School and Harvard Pilgrim Health Care, Boston, MA; [4]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 in combination with either age, race or gender information on respondents. Gender and mean/median age were both non-significant in this model, and race was reported too infrequently to include (adjusted R2=0.60). In the pooled analysis (n = 974), white respondents provided relatively higher ranked utilities for the same health state compared with non-whites, (p= .002); age and gender were non-significant (adjusted R2=0.40).

Conclusion: 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.

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

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 (p04) to assign a bipolar diagnosis than a psychotic diagnosis. In addition, Black patients were significantly less likely ($p<.001$) to receive a bipolar diagnosis than a psychotic diagnosis, when compared to White patients. Furthermore, male patients were significantly less likely ($p<.026$) 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.