ABSTRACTS

A NEURAL NETWORK MODEL FOR PREDICTING CARDIAC RISK INCORPORATING CLINICAL RISK FACTORS AND DOBUTAMINE ECHOCARDIOGRAM RESULTS

PA Nath, SM Day, D Poldermans, D Karavite, KA Eagle, SS Sonnad University of Michigan, Ann Arbor, Michigan

OBJECTIVES Logistic regression and neural network models incorporating cardiac risk factors and dipyridamole-thallium (DTS) results exist to predict the cardiac risk for patients undergoing vascular surgery Dobutamine echocardiogram (DE) is often used rather than DTS to obtain additional information on cardiac risk prior to surgery The purpose of this study was to develop and validate a neural network model using information from clinical history and DE results to estimate perioperative cardiac risk in vascular surgery patients

METHODS We developed a neural network using data from 637 vascular surgery patients from 2 academic medical centers, one in the US and one in Europe Using a two-step Bayesian approach, cardiac risk factors (advanced age, history of myocardial infarction, diabetes mellitus, history of angina, congestive heart failure, and prior coronary artery bypass graft) were considered independently and then in conjunction with DE results (resting wall motion abnormality, ECG change, and new wall motion abnormality) We evaluated model performance using receiver-operating characteristic (ROC) curves and by comparing observed and expected event rates for stratified risk groups

RESULTS Of the 637 patients, 37 (5 81%) experienced post-operative complications (cardiac death or myocardial infarction) The prognostic accuracy (ROC area) of the neural network using only cardiac risk factors was 58% Adding DE results improved model accuracy to 76%, comparable to accuracy of models using DTS results A test of model calibration indicated that risk is overestimated in high-risk patients, as described for existing neural network and logistic regression models Observed versus expected event rates for ordered risk groups were 2 1% vs 2 4%, 17 9% vs 7 4%, 11 4% vs 12 2%, 21 4% vs 17 9%, 20 0% vs 22 9%, 20 0% vs 57 9%

CONCLUSIONS Neural network models provide reproducible and reliable predictions of cardiac risk in vascular surgery patients Moreover, the trends observed using DE results in this study are comparable to those seen in the DTS studies previously described This suggests that the methodology is robust with regard to variation in the type of cardiac imaging data used, a finding that is of particular importance in light of recent trends toward less invasive and less expensive imaging technologies

COST-EFFECTIVENESS OF SCREENING AND TREATING EARLY ALZHEIMER'S DISEASE

SS Araki, KM Kuntz, PJ Neumann, Harvard School of Public Health, Boston, MA

Background: A growing number of individuals are afflicted with Alzheimer's disease (AD), placing tremendous burden on patients, caregivers, and society The U S Preventive Services Task Force has stated that there is insufficient evidence to recommend for or against screening for dementia in asymptomatic elderly persons

Objective: To examine the cost-effectiveness associated with screening for and treating early Alzheimer's disease in a primary care population over 65 years of age

Methods: Previously, we developed a Markov model to simulate AD progression through disease states (mild, moderate, and severe) and settings of care (community and nursing home) to determine the cost-effectiveness of a new drug in treating AD In this study, we estimate the incremental cost per quality-adjusted life-year (QALY) gained with screening versus not screening for AD in the elderly population Our screening strategy consists of a one-time screen using a standard mental status assessment tool in a primary care setting, follow-up testing if the screen is positive, and treatment of afflicted individuals with a drug in the cholinesterase inhibitor class Using available data, we made base case assumptions on disease prevalence (10%), sensitivity/specificity of the screening tool (0 80/0 80), costs of screening ($100) and follow-up tests ($1,000), duration of drug effect (18 months), and the distribution of disease severity when diagnosed (67% mild, 33% moderate), and varied these parameters in sensitivity analyses

Results: Under base case assumptions, the cost-effectiveness (C-E) ratio is $219,000 per QALY The results of the model are highly sensitive to assumptions about disease prevalence, cost of screening, and magnitude and duration of drug effect, and moderately sensitive to assumptions about sensitivity/specificity of screening, cost of follow-up tests, and initial distribution of AD severity Using current estimates, screening for early AD achieves a C-E ratio below $50,000 per QALY only for high-prevalence populations (e g , over 85 years of age) under optimistic assumptions of the duration of drug effect

Conclusion: Our results suggest that screening the elderly primary care population for AD may not be cost-effective Because the findings are sensitive to several key assumptions, the analysis should be revisited as new treatments and screening techniques are developed

USING CLASSIFICATION TECHNIQUES FOR STRATEGY RECOMMENDATION
A COMPARISON OF STANDARD GAMBLE AND VISUAL ANALOG SCALE METHODS

U Chajewska, JW Norman, L Getoor, Stanford University, Stanford, CA

The standard gamble (SG) and visual analog scale (VAS), two frequently used methods for utility assessment, are often found to result in different utilities for the same outcomes We analyzed how SG vs VAS elicitation affects (1) the optimal strategy recommendation, and, (2) the classification of new patients into recommended-strategy groups based on partial utility elicitation

METHODS Our influence diagram models chorionic villus sampling (CVS) and amniocentesis for Down syndrome We used utility data from 51 pregnant women, each of whom was interviewed by a collaborator for an earlier study using SG and VAS to provide utilities for 22 possible birth outcomes For each elicitation technique, maternal age, and utility function, we computed the optimal strategy (4080 cases) For each case, we calculated the loss in expected utility of substituting each of the other strategies for the optimal strategy We then clustered the utility functions using a distance measure based on this utility loss Finally we created a classification tree for each age and elicitation technique, with which we could assign a new patient to a utility cluster (and a recommended strategy) using a small number of preference questions

RESULTS We found that the two elicitation methods generated the same strategy recommendations in 35% of cases, this agreement varied from 25% to 49% when broken down by age, with slightly lower concordance for older patients Using VAS the "no test" strategy was recommended more often than with SG, especially for low maternal age (in some age groups twice as often) The utilities elicited using VAS were more stable, with 61% of the utility functions having the same optimal strategy regardless of maternal age vs 43% for SG Also, if a patient's optimal strategy changed with increasing maternal age, it was likely to change later in life according to the VAS data by more than 5 years For both SG and VAS, amniocentesis was recommended more often with increasing maternal age, and CVS less often VAS elicitation simplified classification, producing more balanced, shallower trees with an average of 5 2 questions vs 7 6 for SG Interestingly, the trees produced using SG relied more heavily on preferences for outcomes related to diagnostic accuracy such as false positive or negative results Cross validation studies revealed no significant difference in classification error between SG and VAS data In both cases, the average utility loss of using a classification tree rather than full utility elicitation was 3 of 100 utility points

DISCUSSION SG and VAS produce different strategy recommendations as well as different utilities

WHICH WOMEN WITH STRESS URINARY INCONTINENCE REQUIRE URODYNAMIC TESTING PRIOR TO SURGICAL MANAGEMENT?

AC Weidner, ER Myers, AG Visco, GW Cundiff, RC Bump Duke University Med Center, Durham, NC

Purpose To determine the predictive value of the symptom of stress urinary incontinence for a urodynamic diagnosis of pure genuine stress incontinence (GSI), and to evaluate the ability of other factors suggested by a published AHCPR guideline to identify patients unlikely to require urodynamic testing prior to surgical management

Methods We evaluated 954 consecutive women (excluding patients with Stage III or IV pelvic organ prolapse) who were referred with symptoms of incontinence Symptoms were recorded using standard forms and definitions as "any stress loss", "only stress loss", "stress and urge loss", "stress and continuous loss", and "no stress loss" Other variables were assessed using a standardized history, physical exam (including urethral axis determination and stress test), one-week validated urinary diary, and post void residual volume (PVR) A multichannel urodynamic diagnosis of pure GSI was used as the gold standard Sensitivity, specificity, and positive and negative predictive values were calculated Logistic regression models incorporating "only stress loss", age, prior prolapse or incontinence surgery, nocturia, voiding frequency, urethral hypermobility, PVR < 100 mL, and race (the factors recommended by the guidelines) as predictors of GSI were constructed to evaluate the predictive ability of the guideline in a subset of 388 patients for whom data on all variables was available

Results	Symptom	Prevalence	Sensitivity	Specificity	+PV	-PV
	Only stress loss	29 7%	38 5%	86 2%	73 7%	58 2%
	Stress & Urge	52 3%	49 2%	56 8%	53 4%	52 6%
	Stress & Continuous	1 8%	2 5%	99 4%	80 0%	50 3%
	Any stress loss	76 0%	91 7%	39 6%	60 4%	82 5%

AHCPR guideline criteria had excellent discrimination (C=0 838) compared to symptoms only (C=0 574), with a +PV of 81 3% However, only 4 1% of subjects met all criteria

Conclusions The predictive value of stress symptoms alone is not high enough to serve as the basis for surgical management AHCPR guidelines improve the predictive value, but may be applicable in only a small subset of patients presenting with urinary incontinence

EMPIRICAL EVIDENCE OF BIAS IN THE EVALUATION OF DIAGNOSTIC TESTS

JG Lijmer, BW Mol, S Heisterkamp, GJ Bonsel, PMM Bossuyt Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands

Background and Purpose The literature contains a large list of potential biases in the estimation of accuracy of diagnostic tests We wanted to determine the quantitative effect of differences in study design on estimates of diagnostic accuracy, to have information on their relative importance when appraising the literature

Methods With a systematic literature search we identified recent meta-analyses summarising the accuracy of diagnostic tests The 11 meta-analyses that were included reviewed 18 diagnostic tests and described the results of 193 original studies All original papers were retrieved The following study characteristics were assessed for each of the 193 studies (1) clinical spectrum (clinical cohort versus case-control), (2) patient sampling (consecutive versus other), (3) verification method (all, partial, different reference standards), (4) interpretation of reference test (blinded or not), (5) data collection (prospective versus retrospective), (6) objective criteria for test reported, (7) objective criteria for reference test reported, (8) description of population

The associations between these characteristics and estimates of diagnostic accuracy were examined with an adjusted summary ROC model The effects were expressed as a ratio of the diagnostic odds ratio's (RDOR) of studies with characteristics of lower methodological quality compared to the reference group

Results Studies evaluating the test in a diseased population and a separate group of 'normals' as controls overestimated the diagnostic performance of a test in comparison with studies that used a clinical cohort (RDOR 3 2, 95% CI 2 1-4 8) Studies in which different reference tests were used for positive and negative test results overestimated the diagnostic performance in comparison with studies using one reference standard for all patients (RDOR 3 0, 95% CI 2 0-4 5) The diagnostic performance was also overestimated when only part of the population was verified (RDOR 1 5, 95% CI 1 1-2 0), the reference standard was interpreted with knowledge of the test result (RDOR 1 6, 95% CI 1 1-2 3), no criteria for the test were described (RDOR 1 7, 95% CI 1 1-2 4) The other characteristics showed no significant effects

Conclusion This study provides empirical evidence that diagnostic studies with designs of lower methodological quality overestimate the power of a diagnostic test, in particularly those including non-representative patients or applying inadequate verification

When Should We Measure Patient's Preference? Utility Assessment in Preventive Surgery of Unruptured Intracranial Aneurysms

Aoki N [1, 2], JR Beck [1], Fukui T [2], Shinbo T [2], Kitahara T [3], Soma [3]
1) Information Technology, Baylor College of Medicine, Houston TX 2) Department of General Medicine and Clinical Epidemiology, Kyoto University Hospital, JAPAN 3) Department of Emergency and Critical Care Medicine, Kitasato University School of Medicine, JAPAN

Purpose: To determine patient characteristics that require the measurement of utility status, in individuals with unruptured intracranial aneurysms

Methods: A decision model of the choice of preventive operation for unruptured intracranial aneurysm (UN-AN) {*Med Decis Making, in press*} was used for this project We defined four disability categories based on a validated neurological trauma index--the Glasgow Outcome Scale (GOS)--and interviewed 140 healthy subjects using the standard gamble method Tornado diagram analysis (TDA) was used to evaluate overall influence of each GOS category on duration of quality adjusted life expectancy (QALE) Threshold analysis (TA) was used to consider the decision context Monte-Carlo simulation analysis (MCA) was conducted to assess the influence of utilities under uncertainty The distribution of probabilities was estimated with meta-analysis and incorporated into the MCA

Results: TDA revealed that the utility value of GOS5 (slight impairment) is the most sensitive to patient preference However, TA showed no utility that influenced the decision in the baseline case of a 40-year old man MCA (10,000 samples) based on age showed preferences for preventive surgery in 100% of cases at age 60, 83% at age 65, 38% at 70, and 4% at 75, using the GOS utility values obtained from the volunteer cohort

Conclusions: Older patients with UN-AN should undergo careful utility analysis with specific elicitation of quality measures for various GOS states Monte Carlo modeling can be used to take the data from a utility study and incorporate it directly into clinical decision making

RELATIONSHIP BETWEEN WILLINGNESS TO PAY AND QUALITY OF LIFE IN RHEUMATOID ARTHRITIS

Bagchi, T Taylor, L Merlino, K Saag University of Iowa, Iowa City, IA USA

Willingness to pay (WTP) is a measure which has not been explored extensively in economic evaluations n health care If WTP is a valid measure of patient preference it should be related to patient quality of life QOL) The objective of this study was to elicit WTP estimates for a cure for rheumatoid arthritis (RA) and to evaluate the relationship between WTP and QOL in an RA population

A WTP questionnaire was developed and pretested in 18 RA patients WTP was elicited using a bidding game on a one time as well as per month basis QOL was assessed using the Health Assessment Questionnaire (HAQ) and the Short Form 36 (SF 36) Regression analysis was used to evaluate the relationship between WTP and QOL

One hundred and seven RA patients were interviewed for the study WTP questions were readily accepted with no refusals The mean WTP for a hypothetical cure for RA was $19693 (SD=34 238 median=8 000) on a one time basis and $346 30 (SD=454 32 median=200) on a per-month basis The monthly WTP was approximately 15% of the subjects' monthly income The mean WTP for a quality adjusted lifeyear was approximately $38,000 Test retest reliability was significantly higher for the one-time format (r=0 98, p < 0 05) than the per-month (r=0 27, NS) format The per-month amounts were subject to starting point bias, but not the one time amounts Regression analyses indicated that WTP was related to QOL dimensional scores with arising, physical functioning, role physical and social functioning having significant coefficients The generic SF 36, which assesses both physical and mental dimensions, was able to increase explanatory power of the regression model over and above the disease-specific HAQ, which has only the physical dimension, indicating that dimensions other than physical were important to the patient

In summary the study was successful in showing a relationship between dimensional QOL scores and WTP The WTP estimates generated in the study are an indication of the burden of RA in this population The one time format might be a better way to assess WTP in RA patients than the per month format The use of generic instruments in addition to disease-specific instruments in QOL assessments or creation of a new disease-specific instrument in RA incorporating dimensions other than the physical is recommended

COUNSELING PERIMENOPAUSAL WOMEN ABOUT POSTMENOPAUSAL HORMONE PROPHYLAXIS ARE MANAGED CARE PLANS BETTER?

Y Wang, MV Pauly Department of Health Care Systems, University of Pennsylvania, Philadelphia, PA

Purpose Even though managed care plans promote the use of preventive services through educational programs and require only a nominal copayment for such services, the gatekeeper system may discourage the use of those services for which referrals are needed We evaluated the difference among health insurance types in counseling perimenopausal women about postmenopausal hormone prophylaxis

Methods In the 1994 National Health Interview Survey, a fraction of perimenopausal women (40 to 60 years old) answered questions concerning menopausal symptoms and ever having discussion with health care professionals about the pros and cons of taking estrogen pills after menopause The study population consisted of women enrolled in fee-for-service (FFS) insurance health maintenance organizations (HMOs), preferred provider organizations (PPOs), individual practice associations (IPAs), and Medicaid We employed the Probit model to assess the effect of health insurance types on the probability of ever having the discussion about postmenopausal hormone therapy

Results Forty three percent of perimenopausal women (N=1873) and sixty three percent of those with menopausal symptoms (N=660) had discussed with health care professionals about postmenopausal hormone therapy Compared with those in FFS plans, perimenopausal women in HMOs were less likely to have the discussion (see details in the table) Other significant predictors included age, menopause, white, and college education (not shown in the table) The HMO effect remained significant for the subsample of women with menopausal symptoms (Estimate=-0 429, Probability>Chi-Square=0 04)

Variable	Estimate (Standard Error)	Probability>Chi-Square
HMO vs FFS	-0 279 (0 129)	0 03
IPA vs FFS	-0 038 (0 078)	0 63
PPO vs FFS	-0 035 (0 137)	0 80
Medicaid vs FFS	-0 075 (0 212)	0 72

Conclusion Both perimenopausal women and women with menopausal symptoms are less likely to receive counselling about postmenopausal hormone prophylaxis in HMOs Managed care plans may provide less preventive care if access to specialists is required

PATIENT AMBIGUITY PREFERENCES

J Hsu, DA Asch, J Baron, PA Ubel University of Pennsylvania, Philadelphia PA, USA

Background Although patients may be influenced by their perceptions of ambiguity when making medical decisions, no standard method for measuring ambiguity tolerance currently exists This study seeks to determine the level of patients' understanding of risk information, and to identify measures of patients' perceived level of ambiguity and attitudes toward ambiguous information

Methods 100 prospective jurors received a hypothetical clinical scenario with three diagnostic tests, the tests varied only in the size of the confidence interval around their average risk estimates The survey tested subjects' ability to identify and to restate point estimates of average risk as well as the boundaries of the confidence intervals The survey also tested for comprehension of the impact of a gene on the chance of developing the disease Another 146 prospective jurors received an identical clinical scenario, examining the relationship between the width of the risk estimate confidence intervals and measures of subjects' ambiguity perception and attitudes toward the information The study looked for a monotonic relationship and for the ability of measures to discriminate between different levels of risk precision

Results Subjects correctly restated factual risk information (80% of the responses among eight items) predicted the impact of a gene on the chance of developing the disease (80%), and identified the diagnostic tests providing the most and least precise risk information (61%) Furthermore, in over 90% of the responses, the measures of ambiguity perception (six items) and attitude (12 items) followed a monotonic relationship with changes in the size of the confidence intervals Specifically, in 93% of the responses subjects' confidence in a test varied monotonically with increases in the confidence interval Similarly, in 95% of the responses, subjects' demand for a test varied monotonically with the confidence interval, and, in 51% of these responses, subjects' demand for a test decreased as the imprecision of the test information increased

Conclusion Subjects demonstrated basic comprehension of risk information Their perceptions of ambiguity level and attitudes towards diagnostic information corresponded with changes in the confidence intervals around the risk estimates These preliminary data support the development of an Ambiguity Preference Instrument and serve as a basis for future reliability and validity studies

A HEALTH INDEX MODEL FOR ULCERATIVE COLITIS DERIVED FROM THE INFLAMMATORY BOWEL DISEASE QUESTIONNAIRE

J Cornwell, L. Lenert, Stanford University, Stanford, CA, USA

Health index models estimate utilities for degrees of impairment of health status Such models can be either general or disease specific We used a database of nearly 500 self identified ulcerative colitis patients to develop a disease specific health index model for ulcerative colitis (UC) based on the Inflammatory Bowel Disease Questionnaire (IBDQ) The database contained information on patients' health status measured using the IBDQ, and their preferences measured using a double anchored visual analog scale (VAS) and the standard gamble (SG) Looking across the four dimensions of health represented in the IBDQ summary measure data (social function, systemic symptoms, bowel symptoms, and emotional function scores), we identified the optimum set of mutually exclusive and collectively exhaustive multi-dimensional health states in UC using k-means cluster analysis To identify the optimum number of states, we used a novel bootstrap approach that added health states to the model one at a time until patterns were no longer mutually exclusive Estimates for the utility of each state were taken to be the mean VAS and SG rating for current health of the patients in each state Results of analysis revealed that the optimal model had four states We created objectively based text descriptions of each state by inverting IBDQ items for later use in studies of the preferences of the general public These can be viewed at <http //preferences stanford edu/ucindex/states html> In split sample validation studies, we have found that the optimum 4-cluster model produced less biased estimates of VAS and SG scores and had at least as great predictive power as models based on simple linear regression (r^2 = 23 vs 0 21, VAS model, r^2 = 0 04 vs 0 04, SG model) These results suggest cluster analysis is a powerful approach for developing disease specific health index models This algorithm allows comparison of patients' preferences for their current health with the general public's preferences for the similar states and compares favorably to linear regression in terms of its statistical properties

EFFECTS OF HIGH PRIORITY PATIENTS ON OTHER PATIENTS WHO HAVE LIFE THREATENING CONDITIONS IN AN EMERGENCY DEPARTMENT

F Boutros, DA Redelmeier Division of General Internal Medicine, Sunnybrook Health Science Centre, University of Toronto, Canada

Trauma victims take priority over other patients because their dramatic injuries require immediate treatment. We hypothesized that such demands disrupt queue discipline and compromise the care of other patients who have potentially lethal conditions, such as those with acute chest pain.

Consecutive trauma victims presenting from April 1, 1995 to April 1, 1997 were identified at one medical center Through administrative database analysis, we selected case patients who had presented with chest pain and arrived immediately following a trauma. Similarly, we selected matching control patients who had presented with chest pain but had arrived on a preceding day when no trauma was in the emergency department. As expected, case and control patients (n=70 and 70, respectively) were similar in mean age (60 vs 60, p=NS), % male (47 vs 53, p=NS), and % ultimately diagnosed as non-cardiac (29 vs 33, p=NS) Case patients spent 81 minutes longer in the emergency department (297 vs 216, p=0 009) Case patients had worse final scores in on the American College of Emergency Physicians Quality Assurance Index (76 vs 84, p=0 027) The most common anomalies were the failure to administer aspirin (62% vs 46%), the failure to adequately treat ongoing pain (58% vs 44%), and the failure to provide instructions regarding follow-up (76% vs 44%).

Patients with special priority can compromise the timeliness and quality of care that other patients receive in an emergency department. Optimizing care requires flexible strategies that respond to special demands without jeopardizing the care of other patients who have potentially lethal conditions.

PREFERENCES AND WILLINGNESS TO PAY FOR PAINLESS DENTAL LASER DRILL

M Goldstein, W Sumner, B Littenberg Washington University School of Medicine, St Louis, MO

We sought to determine the monetary value of painless dentistry Lasers decrease discomfort, do not create a frightening sound, and minimize the need for local anesthetic injections The Er YAG laser has recently been approved by the Food and Drug Administration At present, insurance companies do not cover this service Dentists are reluctant to adopt this technology because of expense and novelty Capital costs are $39,000, and the variable costs are less than $15 per procedure We estimated an insurance premium increase of $2 50 per month would allow insurers to reimburse dentists' costs

Design We prospectively surveyed patients to assess their preferences and willingness to pay for dental laser treatment and insurance coverage, using a U-Titer based computer instrument Patients were asked to fill in the blank with a dollar amount they thought would be expensive but worth it Following this, we did a bisection search anchored at $0, and up to a round number $21-$30 higher than the fill in the blank number

Subjects Patients awaiting conventional drill and fill treatment at two dental offices

Main Outcome Variable Dollar amount patients are willing to pay for laser drill, additional premium they are willing to pay for insurance coverage of laser services

Results Thirty-eight patients were interviewed Two declined the laser drill Of the remaining 36 patients, the median (25th,75th percentile) willingness to pay in extra out-of-pocket costs for the laser drill was $20 (10,30) when asked to fill in the blank, and $40 (19,50) with the bisection technique Three people (8%) were unwilling to pay any extra expenses and eleven (31%)were not willing to pay ≥ $15 for the laser drill when asked to fill in the blank In contrast, only one person (3%) was not willing to pay any additional out-of-pocket costs and seven people (19%) were not willing to pay ≥$15 when the bisection technique was used The median for additional insurance premium patients are willing to pay was $10 (2 4,10) Only two patients (6%) were unwilling to pay any extra premium

Conclusion Demand for the dental laser drill is high Approximately 75% of patients are willing to pay for the laser drill out of pocket, and a similar proportion are willing to pay an additional insurance premium for its coverage The methods used to ascertain the dollar amount may influence the value obtained

DIFFERENT IDEALS OF SHARED DECISION MAKING DEVELOPMENT OF A QUESTIONNAIRE TO ASSESS VARIOUS INTERPRETATIONS OF PATIENT AUTONOMY

AC Molewijk, A Stiggelbout, W Otten, J Kievit, Department of Medical Decision Making, Leiden University Medical Center (LUMC), Leiden, The Netherlands

PURPOSE Although most people seem to support shared decision making (SDM), few realize that SDM can be based on quite different, sometimes even paternalistic, moral values Our goal is to develop an instrument that measures these underlying moral values by assessing interpretations of patient's autonomy (PA) Our PA questionnaire will be used to (a) compare physicians and patients' opinions concerning PA, (b) contrast ideal conceptions of PA versus actual performances, and (c) study differences between ethical theory and ethical practice METHODS A search of the ethics' literature revealed six theoretical views of PA within the process of SDM (such as autonomy paternalism "we want you to be autonomous") A set of 21 items was constructed to reflect these views University students (N = 135) judged how well these items reflected their ideal of PA, assuming they suffered from a serious disease (5-point Likert scale) A general question about the "ideal decision maker" (IDM) served as an internal validity check RESULTS The IDM varied evenly between "physician mainly" (28 9%), "physician and patient equally" (26 7%), and "patient mainly" (38 5%) Factor and reliability analyses showed that one internal consistent scale could be formed, consisting of 7 items (Cronbach's alpha = 79) All these items explicitly referred to the IDM question with a strong impact (p < 0001) Five other items did relate to this scale, dealing with autonomy paternalism, the process of communication and the priority of health over PA Both autonomy paternalism and the priority of health above PA were spread over the domain (respectively 15/14/16/48/42% and 12/19/25/49/30% for the range disagree-agree) The analyses showed that the remaining 9 items could be grouped into "information exchange" or "counsellor role" items, but they did not form reliable scales Interestingly, these latter items did not relate to the decision scale, suggesting that these items cover unique elements of PA CONCLUSIONS Results show that one dimension of PA could be distracted that points clearly at the decision making aspect of PA, whereas ethical theory describes more nuances Surprisingly, information exchange and counsellor's role items did not relate to the decision dimension, implying either their unique contribution to the concept of PA or that these aspects do not affect the perceived degree of PA To replicate and validate these conclusions, the questionnaire will be administered to other groups

COST EFFECTIVENESS OF ANDROGEN SUPPRESSION STRATEGIES IN ADVANCED PROSTATE CANCER

AM Bayoumi, AS Brown, AM Garber University of Toronto, ON, University of Oxford, UK, University of Western Ontario, London, ON, Stanford University, CA, and VA Palo Alto Health Care System, CA

Androgen suppression strategies for patients with advanced prostate cancer differ markedly in cost and toxicity but only marginally in efficacy We evaluated the cost effectiveness of hormonal therapies available in the United States

We assessed the costs and benefits of orchiectomy, diethylstilbestrol (DES), a leutinizing hormone releasing hormone (LHRH) agonist such as goserelin, a non steroidal antiandrogen (NSAA) such as flutamide, and two combined androgen blockade (CAB) strategies an LHRH agonist plus an NSAA, and orchiectomy plus an NSAA The base case was a 65 year old man with local recurrence of moderately differentiated prostate cancer who previously received definitive treatment with radiation or surgery We constructed a Markov model of advanced prostate cancer, explicitly accounting for disease progression, death from other diseases, fatal effects of antiandrogens, minor and dose limiting toxicities, local obstructive symptoms, second line hormonal therapy, efficacy of NSAA withdrawal, and pre terminal chemotherapy Drug efficacy was estimated from a meta analysis of clinical trials Transition probabilities, costs and utilities were estimated from the literature

DES was associated with the shortest survival and least expense The cost effectiveness of orchiectomy relative to DES was $11,100 per life year gained and $17,100 per quality adjusted life year (QALY) Monotherapy with LHRH agonists or NSAA and both CAB regimens are dominated by orchiectomy However, orchiectomy is dominated by other strategies if patients assign a low value to their quality of life after the operation The cost effectiveness of CAB with an LHRH agonist and a NSAA relative to orchiectomy was less than $100,000/QALY only if the risk of disease progression was decreased by 25% or more, whereas the base case estimate was 2% (95% C I 22 to 22%) For CAB with orchiectomy and an NSAA, the required efficacy was 12% with a base case estimate of 6% (95%C I 14 to 22%)

The most cost effective androgen suppression strategy for patients with advanced prostate cancer depends on how individuals value life after orchiectomy Combined androgen blockade yields a small benefit at a high relative cost

KNOWLEDGE-BASED FORMULATION OF DYNAMIC DECISION MODELS IN MEDICINE

CG Wang, TY Leong, APK Leong, FC Seow National University of Singapore, Singapore and Singapore General Hospital, Singapore

We present a knowledge-based model construction approach to support automated and interactive formulation of dynamic decision models, i e , models that explicitly consider the effects of time We incorporate a hybrid knowledge representation scheme that integrates categorical knowledge, probabilistic knowledge, and deterministic knowledge The categorical knowledge captures the definitional and structural relations of the domain concepts, e g , diseases, tests, treatments, and other entities related to the decision problems This type of knowledge provides the power of abstraction and inheritance, it supports modeling at multiple levels of details The probabilistic knowledge captures the interactions among the concepts These relations are needed to support automated derivation of missing information in model construction The deterministic knowledge expresses the deterministic rules and declaratory constraints about the domain The target decision models are expressed in terms of DynaMoL, a high-level dynamic decision modeling language that supports multiple graphical perspectives and allows simple and explicit specification of decision factors and constraints The modeling process begins with the decision context information elicitation, followed by the action space and the state space definition A transition view, which corresponds to the Markov state transition diagram, is then constructed for each action defined in the action space If the action-specific transition functions among the states cannot be directly determined from the knowledge base, the model constructor will try to identify a set of event or chance variables that constitute the effects of the action and a set of probabilistic influences among the event variables These entities constitute an action-specific influence view, which corresponds to the dynamic influence diagram without decision nodes or value nodes, they can be translated into the corresponding action-specific transition functions A set of knowledge-based model modification operations are provided for automatic and interactive generation, abstraction, and refinement of the model components The final model is evaluated or solved by a prototype implementation of DynaMoL to determine an optimal course of action We have conducted a comprehensive case study in colorectal cancer management

The results demonstrated the practical promise of the proposed design through the correctness of and the ease with which some reasonable dynamic decision models in medicine can be constructed

OUTCOMES OF CHILDREN IN THE PEDIATRIC INTENSIVE CARE UNIT

RC Sachdeva, LS Jefferson, JR Beck, BA Brody Baylor College of Medicine, Houston, TX

Outcome measures in the pediatric intensive care unit (PICU) setting have typically included severity adjusted mortality risks using the Pediatric Risk of Mortality (PRISM) scoring system However, these methodologies are limited in that they do not capture health related quality of life (HRQOL), and the perception of the child's outcome from the parent's standpoint Measures of HRQOL include the Health Utilities Index (HUI) and Multi Attribute Health Status system (MAHS), and measures of perception of health outcomes include techniques like linear scaling (LS) and time trade off methods (TTO) The purpose of this study was evaluate these methodologies, and to measure HRQOL and parent's perception of their child's outcome after the PICU stay

This is a prospective cohort study of 260 children performed at the Texas Children's Hospital, Houston, TX, between September 1996 and March 1997 Children admitted to the PICU between the age of 2 and 12 years were eligible to be enrolled after they had stayed for ≥ 24 hours (n=130) These children were individually matched based upon diagnostic category, age, race, and gender with children from general pediatric wards (n=130) Data collected included admission demographics, PRISM score, measurements of "base line" HUI and MAHS scales reflecting the child's status prior to the hospitalization, HUI and MAHS scales at discharge from the hospital, measures of LS and TTO reflecting parent's perception during the initial interview and at discharge

Results showed statistically significant correlations (p<0 05) between HRQOL, PRISM, LS, and TTO in both groups of children in the PICU and general pediatric wards There was an improvement (p<0 0001) in LS and TTO estimates at the time of discharge as compared with estimates at admission HUI domains (physical function, role function, social-emotional function, and health problem) were all statistically significantly (p<0 0001) worse at discharge as compared to "base line" values Similar results, with statistically significantly worse estimates of HRQOL were found on all the six domains (sensory, mobility, emotion, cognitive, self-care, and pain) of MAHS at discharge

HRQOL, LS, and TTO estimates can be used to measure outcomes of children in the PICU and inpatient setting Children who survive the PICU stay still have not reached "base line" HRQOL levels at the time of discharge Therefore, outcome estimation in the PICU needs to include HRQOL estimation in addition to severity adjusted mortality risks *(Supported by the NIH 1996 CAP award)*

COST EFFECTIVENESS OF RADIOFREQUENCY ABLATION FOR TREATMENT OF PAROXYSMAL SUPRAVENTRICULAR TACHYCARDIAS

CHF Cheng, GD Sanders, PA Heidenreich, KM McDonald, BK Lee, MS Larson, MA Hlatky, DK Owens Stanford University, Stanford, CA, and VA Health Care System, Palo Alto, CA

Radiofrequency ablation (RFA) has recently become the preferred treatment for supraventricular tachycardias (SVT), replacing long-term antiarrhythmic drug therapy and management of acute events RFA is costly and may have complications, but may improve the patient's quality of life and reduce long-term costs As part of the Cardiac Arrhythmia Patient Outcome Research Team (PORT) project, we evaluated the cost effectiveness of RFA for initial treatment of SVT

We used a Markov model that compared three treatment strategies. initial RFA, long-term antiarrhythmic drug therapy, and episodic pharmacologic treatment of acute events We evaluated these therapies for patients who have known SVT with etiologies that included atrioventricular nodal reentrant tachycardia, and bypass tracts The model incorporates the effects of therapy on quality of life, assessed by time-tradeoff, based on a survey of 161 RFA patients who were moderately to severely symptomatic The model incorporates complications associated with RFA, including the risk of death and atrioventricular nodal block We estimated efficacy, mortality, and complication rates based on published trials We estimated costs from the literature and the cost accounting system in a high volume referral center We assumed a starting age for patients of 35 years and discounted health and economic outcomes at 3 percent

Over a wide range of severity of symptoms, RFA dominated episodic or long-term drug therapy For our base-case analysis, patients who experienced 5 arrhythmic events per year that resulted in emergency room visits, RFA cost $51000 over the lifetime of the patient, and resulted in 18 85 QALYs Long-term drug therapy cost $66100 and resulted in 17 72 QALYs Episodic drug therapy cost $88100 and resulted in 15 78 QALYs For patients who experienced only one arrhythmic event a year that resulted in ER visit, RFA still dominated other therapies — though the benefit of RFA versus drug therapy was reduced to 0 24 QALYs The effectiveness of drug therapy had to be raised from our base-case estimate of 60% to 99 8%, or the risk of death during a RFA increased 50-fold for long term drug therapy to be as effective as RFA

For moderately to severely symptomatic patients, RFA reduces lifetime expenditures and increases QALYs Because RFA is cost effective if it produces only small improvements in quality of life, the procedure may prove also to be warranted in mildly symptomatic patients

ENHANCEMENT OF DIAGNOSTIC REASONING BY A COMPUTER-BASED DECISION SUPPORT SYSTEM

AS Elstein, CP Friedman, FM Wolf, GC Murphy, TM Franz, PS Heckerling, PL Fine, TM Miller, J Miller Univ of Illinois at Chicago, IL, Univ of North Carolina, Chapel Hill, NC, Univ of Michigan, Ann Arbor, MI

Aims To assess the effects of using ILIAD, a computer-based diagnostic decision support system (DSS), upon the diagnostic opinions of users of varying levels of clinical experience

Design 4 clusters, each with 9 diagnostically challenging cases, were constructed and equated for difficulty Definitive diagnostic findings were deliberately omitted from each case summary Subjects were randomly assigned to a cluster and trained to use ILIAD for diagnostic consultation

Subjects and Procedure 36 general internists, 36 residents in internal medicine, and 36 4th-year medical students (M4), from 3 academic medical centers, each worked up a case cluster, first without and then with ILIAD consultation They were asked to offer a differential diagnosis of up to 6 diagnoses and to list up to 3 next steps in the workup Effects of consultation were measured by number of correct diagnoses over all cases and by changes in mean diagnostic quality scores per physician, averaged across the nine-case cluster The diagnostic quality score for each case was composed of two equally weighted components, ratings of the clinical plausibility of each diagnostic hypothesis weighted by its location on the list, plus an additional location bonus for the correct diagnosis being on the list

Results Over 972 workups, number of accurate diagnoses increased from 389 initially to 429 post-DSS consultation The increase was greatest for M4s Changes were in both directions If the correct diagnosis was not on the initial list, it appeared in about 20% of cases following consultation However, if the correct diagnosis was on the initial list, it was dropped following consultation in about 10% of cases, so the gain was about twice the size of the loss Mean diagnostic quality score increased significantly (p < 001) from pre- to post-DSS consultation, the gain was greatest for M4s Effect size ranged from 0 07 SD for the attending physicians to 0 57 SD for the M4s The vast majority of changes in subjects' differential diagnosis lists did not affect diagnostic accuracy

Conclusion Inexperienced clinicians benefit most from ILIAD consultation The diagnostic reasoning of experienced physicians is affected very little Both gains and losses in diagnostic accuracy were observed The gains were about twice the size of the losses

LOW BACK DISORDERS DECISION MAKING PERFORMANCE OF MACHINE VERSUS CLINICIANS

SA Gracovetsky, A Marriott, NM Newman, MP Richards, S Asselin Concordia University, Spinex Medical Technologies Inc , Centre Hospitalier Universite de Montréal, Montréal, Canada

The evaluation of spinal function requires the distillation of a large number of clinical observations of the patient s stance and motion with regard to deformity, symmetry, speed and coordination Scientific medical literature raises serious questions regarding the efficacy of the clinical examination for assessing spinal function of low back pain patients, with mounting evidence to suggest that this problem is rooted in the clinician's dependency on reported pain To improve the situation, the use of automated machines has been proposed, to process functional information into a diagnosis that is independent of reported pain Such a system comprising 14,000 rules has been developed (the Spinoscope), and its performance assessed against that of clinicians in a strictly controlled, blinded study First, two low back pain expert clinicians established a "gold standard" using the clinical examination for 87 subjects who were classified as normal (46 no low back pain) or abnormal (41 with benign low back pain) The 87 subjects were randomly divided into two groups the first group (23 normal, 21 abnormal) were instructed to report their true condition (objective), the second group (23 normal, 20 abnormal) presented themselves as the opposite of their true condition (non objective) The same subjects were re-examined by two independent clinician evaluators, and by the machine, which collects functional data extracted from the kinematics of skin markers and EMG affixed to the subjects who then carry out pre-determined exercises The diagnostic performance of the machine's 14,000 decision rules was compared to that of the clinician evaluators The clinicians were more accurate with objective subjects (99% vs 78%), the machine more accurate with non-objective subjects (72% vs 30%) and, for the entire population tested, they were equivalent (71% vs 72% concordance) The clinicians were unable to determine the level of objectivity of a subject (26% correct identification) By relying primarily on the subject s self-presentation, often to the exclusion of objective findings, the clinician may err in evaluating low back function when the patient does not report his or her true condition The additional functional analysis provided by the machine offers the clinician objective, pertinent information to complement the clinical examination This experimental study demonstrates that a machine that processes strictly functional data and specifically excludes reported pain and history, which may be unreliable, can reliably determine the diagnosis of benign low back pain

SHORT FORM-12 AS A PREDICTOR OF HEALTHCARE EXPENDITURES IN INDIVIDUALS OVER 65 YEARS OF AGE

PL Cyr, G Lenhart HCIA-Response, Inc Waltham, MA

The Short Form-12 (SF-12) has long been an accepted measure of physical and mental functioning With the elderly population increasing in size and the movement of Medicare to a managed care enviroment, the SF-12 could potentially serve as a predictor of medical expenitures for individuals over 65 years of age

SF-12 surveys were administered to new enrollees in a managed care organization's (MCO) Medicare-risk product The MCO reported over a 50% response rate Results from 3,912 SF-12s were combined with information from paid claims using an outcomes data warehouse A least-sum-of-squares regression model was used to examine the predictive power of the SF-12's physical and mental summary scores for healthcare expenditures The SF-12 physical and mental summary scores were significant predictors The outcome variable was the natural logarithm of dollar amount paid per respondent for a six month period The natural logarithm was used to adjust for the skewness of the claims data The model was adjusted for age, gender, and number of self-reported comorbid conditions (diabetes, hypertension, lung disease, heart disease, irregular heartbeats, skin ulcers, transient ischemic attacks)

Both physical and mental summary scores were predictive of medical expenditures For a five point increase (better functioning) in the physical summary score there was a 13 5% (p= 0001) decrease in dollar amount paid The relationship between the mental summary score was not as strong as the physical summary score but was still predictive of medical expenditures For example, for a five point increase (better functioning) in the mental summary score there was 7 5% (p= 0001) decrease in dollar amount paid

Our findings suggest SF-12 summary scores may be useful in predicting medical expenditures in an elderly population and may be of help to clinicians and policy makers in assessing and managing elderly individuals' health and healthcare expenditures

WHAT SHOULD SOCIETY BE WILLING TO PAY FOR A QALY? EVIDENCE FROM THE VALUE OF LIFE LITERATURE

RA Hirth, ME Chernew, E Miller, AM Fendrick, WG Weissert University of Michigan, Ann Arbor, MI

Cost-benefit analysis (CBA) provides a clear decision rule regarding whether to use an intervention under investigation undertake the intervention if the monetary value of its benefits exceed its costs Due to a reluctance to characterize health benefits in monetary terms, users of cost-effectiveness analyses (CEA) must rely on arbitrary standards (e g , < $50,000 per QALY) to deem a program "cost-effective" Moreover, there is no clear consensus regarding the appropriate dollar value per incremental QALY gained on which to base resource allocation decisions

To address this issue, we set out to determine the value of a QALY as implied by the value of life literature and compare this QALY value to arbitrary thresholds for cost-effectiveness that have come into common use A literature search from 1966 to 1998 identified 42 estimates of the value of life that were appropriate for inclusion These estimates were classified by method human capital (HK), contingent valuation (CV), revealed preference/job risk (RP-JR) and revealed preference/non-occupational safety (RP-S), and by U S or non-U S origin After these value of life estimates were converted to 1997 US dollars, the life expectancy of the study population and a 3% real discount rate were used to calculate the implied value of a QALY

The median value per QALY was $146,751 Median value per QALY by study type was $22,524 (HK estimates), $112,706 (RP-S estimates), $145,938 (CV estimates) and $302,548 (RP-JR estimates) An OLS regression of the log value of a QALY on study type and national origin explained 52% of the variance across studies and did not reveal any differences between values derived US and non-US studies

Conversions from the value of life literature yielded highly variable estimates of the value of a QALY Interestingly, the median value of a QALY calculated from three different methodologies evaluated (RP-S, CV, RP-JR, HK was an exception) far exceeded the "rules of thumb" that are frequently cited to determine whether an intervention produces an acceptable increase in health benefits in exchange for incremental expenditures

THE INFLUENCE OF PSYCHIATRIC HISTORY ON MEDICAL EVALUATION OF SYMPTOMS

MA Graber, GR Bergus, BT Levy, B Kabele The University of Iowa, Iowa City, IA

Background When evaluating somatic complaints, physicians have been shown to be influenced by a patient's psychiatric history It is not clear whether this history impacts physicians' assessments of the probability of disease or the intervention threshold

Methods Primary care physicians were asked to read 2 scenarios about patients with new, potentially serious complaints (severe headache and severe abdominal pain) Physicians were randomized to 3 experimental groups Group 1 was given no additional past medical history, group 2 was told about a past history of major depression and group 3 was told that the patient had a long history of somatic complaints without obvious organic etiology Physicians were asked to estimate the probability of serious disease and their likelihood of recommending further investigation beyond the history and physical

Results 231 of 300 (77%) physicians returned completed questionnaires Physicians' estimates of the probability of serious disease were influenced by the patient's past history of major depression or past history of multiple somatic complaints without an identifiable organic etiology (p<0 0001 in both cases) Similarly, their likelihoods of recommending immediate further investigation were influenced by these past histories (p<0 0001 in both cases) However, when physicians' estimates of the probability of serious disease were included as a covariate, we found no differences between the groups likelihoods of recommending further investigation with both scenarios (ANCOVA, p=0 36 and 0 18, respectively) This indicates that physicians might respond differently to patients with significant affective disease solely because of different estimates of serious of disease This relationship remained even after physician gender and experience were entered into the analyses

Conclusion Physicians' willingness to recommend further evaluation of severe complaints were significantly affected by the past history of depression or somatization However, the different testing recommendations did not arise because of different intervention thresholds but because of different estimates of the probability of disease

MODELING THE COST-EFFECTIVENESS OF A COMMUNTIY BASED HEALTH PROMOTION PROGRAM

SD Ramsey, DL Patrick, AC Spencer, University of Washington, Seattle, WA

PURPOSE Modeling techniques are uncommonly applied to community-based health promotion programs The model must include the costs and effectiveness of the intervention as well as the promotion program Using data from an RCT, we model the cost-effectiveness of a community promotion of an exercise intervention for adults with osteoarthritis (OA)

METHODS The University of Washington, in collaboration with the Arthritis Foundation, Washington State Chapter, conducted an RCT comparing aquatics exercise versus usual care for adults, aged 55-75, with OA Using the societal perspective, the model estimates the cost-effectiveness of promoting this exercise program for people with OA The time horizon of the trial is 40 weeks The effectiveness of the intervention is modeled over lifetime of the participants Direct costs include cost of recruitment, the aquatics exercise class, and OA-related health expenditures The incremental effectiveness of the aquatics exercise class is measured as quality adjusted life years (QALYs) Future costs and effects are discounted at 3% per annum The model incorporates the probability of individuals responding to the health promotion program, the probability of being eligible, the probability that an eligible candidate chooses to participate, and the probability of adhering to the exercise program for the trial and beyond

RESULTS Of 1,018 individuals who responded during the recruitment phase, 249 decided to participate in the RCT Incremental costs of the promotion effort were $21 per participant Under the assumption that direct medical care costs for the treatment and control groups are similar, equal survival, and maintenance of QOL gains from exercise, the estimated incremental cost-utility of exercise is (1) excluding promotion costs $18,379/QALY gained, and (2) including promotion costs $26,117/QALY gained The result is modestly sensitive to changes in the costs and effectiveness promotion effort

CONCLUSIONS Adding the costs of community promotion modestly reduces the cost-effectiveness of the exercise intervention for OA The ultimate cost-effectiveness of this community promotion program is more dependent on the effectiveness of the intervention than the effectiveness of the promotion effort There are implications for promotion efforts by state and local governments, as well as managed care organizations

DO QUALITY-ADJUSTED LIFE YEARS REFLECT PATIENT PREFERENCES? VALIDATION USING REVEALED PREFERENCE FOR INTENSIVE TREATMENT OF INSULIN-DEPENDENT DIABETES MELLITUS

DO Meltzer, TS Polonsky University of Chicago, Chicago, IL

Purpose Quality-adjusted life years (QALYs) are often used to evaluate therapies in medical cost-effectiveness analysis, but their validity as a measure of patient preferences is unproven The absence of an accepted measure of patient preferences has hindered efforts to validate QALYs We used the clinically accepted standard for identifying the preferences of patients – their informed choices – to assess the "revealed" preferences of 130 patients with IDDM for intensive (INT) vs conventional (CONV) therapy To assess the validity of QALYs as a measure of patient preferences, we then used standard techniques to calculate the change in QALYs with INT and examined whether this predicts patient choice of therapy

Methods Time trade-off questions were used to measure patient beliefs about quality of life (QOL) with each therapy and the major complications of IDDM Frequency assessment questions were used to assess patient beliefs about the efficacy of INT Questions about preference for the present vs future were used to assess discount rates A decision model based on the Diabetes Control and Complications Trial (DCCT) was used to calculate the change in QALYs with INT

Results Patients on CONV reported lower QOL with INT vs CONV (p<0 01), while patients on INT reported no difference in QOL However, patients on INT and CONV did not differ in the QOL they associated with complications of IDDM Patients on INT reported higher estimates of efficacy of INT in preventing complications than did patients on CONV (p<0 01), but overestimated the benefits of INT compared to the results of the DCCT Patients on INT also exhibited lower discount rates than did patients on CONV The change in QALYs predicted reported treatment, with an area under the ROC curve (AUC) of 0 84 Change in QALYs also predicted the behaviors that define INT, but the AUC was only 0 73, perhaps because the correlation of QALYs and reported therapy is enhanced by attempts to minimize cognitive dissonance

Conclusions QALYs correlate with patient preferences as revealed by their choices However, the possibility that answers reflect patient efforts to minimize cognitive dissonance concerning their choice of therapy suggests that this should be considered an upper bound on the extent to which QALYs reflect patient preferences Analyses whose results depend on measurement of QALYs should be interpreted with caution Revealed preference techniques should be used to validate and improve QOL assessment

THE VALUE OF "EXPLICIT" VALUES CLARIFICATION EXERCISES IN PATIENT DECISION AIDS

AM O'Connor, GA Wells, P Tugwell, A Laupacis, T Elmslie, E Drake University of Ottawa, Canada

Guidelines for "value sensitive" decisions recommend patients consider not only the probability of outcomes but also their personal value We demonstrated previously that decision aids improve perceived clarity of values and correlations between values and choices These effects may have been due to 1) describing outcomes in familiar experiential terms to improve implicit judgements of their value, 2) valuing outcomes explicitly using a graphic 'weigh scale' exercise, and 3) describing examples of how other women's decisions were influenced by their values Although all patient decision aids use the first strategy, they vary in their inclusion of the second and third strategies, thereby varying their complexity, duration and cost

Objective: To compare implicit versus explicit methods of valuing outcomes in a decision aid regarding hormone therapy in terms of promoting clarity of values and congruence between values and decisions

Design Randomized controlled trial involving 201 women aged 50-69 who had never used hormone therapy

Interventions Decision aids with detailed descriptions of options, outcomes, and probabilities either followed by **1)** a summary of the main outcomes to implicitly consider benefits versus risks, or **2)** a graphic "balance scale" values clarification exercise to explicitly consider the personal value of each benefit and risk, including examples of how others completed the exercise

Measures A) Perceived clarity of values, a subscale of the decisional conflict scale, B) Accuracy of personal values of benefits and risks (measured from 0-not at all important to me to 10-extremely important) in discriminating between choices (measured as accept, decline, unsure about hormone therapy)

Results There were no statistically significant differences between intervention groups in decisions, perceived clarity of values, and overall discrimination between values and choices However, among those choosing HRT, those exposed to explicit values clarification had significantly better (p=0 05) agreement between values and choices (40%) compared to those exposed to implicit values clarification (0%)

Conclusion The main mechanism explaining previously observed improvements in perceived clarity of values is likely the detailed outcome descriptions The short-term benefit of our explicit values clarification method is limited to those changing the status quo The possible downstream effects on the quality of patient-practitioner communication and long-term persistence with decisions need to be studied

COST-EFFECTIVENESS OF CONSERVATIVE AND INTERVENTIONAL TREATMENT STRATEGIES IN PATIENTS WITH INTERMITTENT CLAUDICATION

S O de Vries, J A de Vries, J B Wong, M C Donaldson, M G M Hunink University of Groningen, the Netherlands, Harvard Medical School, Boston, New England Medical Center, Boston

Objective To compare the costs, benefits, and cost-effectiveness of alternative strategies in the treatment of patients with intermittent claudication

Methods A Markov decision model was developed that incorporated estimates of the risks, benefits, and costs of various therapeutic options in patients with peripheral arterial occlusive disease Therapeutic options included exercise, percutaneous transluminal angioplasty, and bypass surgery Treatment strategies were defined by the initial therapy in combination with secondary treatment options should the initial therapy fail Treatment of critical ischemia was the same for all strategies (angioplasty and/or bypass surgery) The model allowed for both suprainguinal and infrainguinal disease and for bilateral symptoms To capture all relevant clinical details, multiple tracker variables were incorporated in the model Both first and second order Monte Carlo simulation were used to evaluate the model Based on the results from univariate sensitivity analyses, subgroups were defined by age and history of coronary artery disease A bootstrap technique was used to explore the joint distribution of incremental costs and effects

Results Compared to a strictly non-interventional strategy (i e , exercise only), treatment strategies that included angioplasty or both angioplasty and bypass surgery improved the quality-adjusted life expectancy in patients with no history of coronary artery disease The incremental cost-effectiveness ratios varied from $37,000 to $82,000 per quality-adjusted life year gained for strategies that included angioplasty and from $202,000 to $779,000 per quality-adjusted life year gained for strategies that included both angioplasty and bypass surgery In patients with a history of coronary artery disease all interventional strategies were inferior to the strictly non-interventional strategy, which yielded the highest quality-adjusted life expectancy at the lowest lifetime costs Differences in outcomes between strategies that included interventions as a primary options vs those that considered interventions only as secondary options were minimal

Conclusion The results suggest that overall the gain in effectiveness achieved with bypass surgery for intermittent claudication does not justify the additional costs, even if preceded by a trial of exercise Future research efforts are probably best directed at comparing a strictly non-interventional treatment strategy of exercise only to a strategy combining angioplasty and exercise

ANTICOAGULATION AFTER A HISTORY OF FALL IN PATIENTS WITH CHRONIC ATRIAL FIBRILLATION - A DECISION ANALYSIS

Addison R. Tolentino, MD, Thomas McGinn, MD, MPH Montefiore Medical Center, Albert Einstein College of Medicine, Bronx, NY

Background: Atrial fibrillation occurs in about 5% of elderly persons above 65 years old Anticoagulation reduces the rate of cerebral thromboembolic events in these patients from 4 5% to 1 5% per year One very common problem though among elderly persons is the occurrence of fall which may or may not result to serious sequelae among which include intacerebral hemorrhage This decision analysis was performed to evaluate the benefits versus risks of coumadin therapy in a 65-year old man with a history of chronic atrial fibrillation and a recent history of fall

Methods: A one year decision model was constructed to estimate the risks and benefits of anticoagulation in the above mentioned patient A review of literature was done to determine the following key probabilities frequency of falls in the elderly in the community, incidence of and mortality rate from traumatic intracranial bleed on coumadin, annual rate of cerebral thromboembolism in chronic atrial fibrillation on and off coumadin and the mortality rate from them, and the incidence of intracerebral bleed unrelated to traumatic fall in patients on coumadin Utilities ranging from 0 to 1 were assigned depending on the outcome namely, the absence of either thrombotic stroke (resulting from atrial fibrillation) or intracerebral bleed (resulting from a fall/anticoagulation), presence of either one or both of the above, or death Sensitivity analyses were performed over wide ranges of the above key probabilities and utilities

Results: The anticoagulation therapy branch had the higher expected value (0 979 versus 0 956) and remained the preferred strategy after sensitivity analyses were performed on both utilities and probabilities The rate of traumatic intracranial bleed on coumadin had to be more than 10 7% for the decision to be reversed This rate far exceeds the highest probability reported in the literature

Conclusion: A patient who is on anticoagulation for chronic atrial fibrillation and presents with a history of fall should be continued on therapy given benefits derived from anticoagulation and a very small risk of intracerebral bleed from falls or otherwise

COST-EFFECTIVENESS OF TECHNOLOGIES TO IMPROVE SENSITIVITY OF PAP SMEARS IMPACT OF SCREENING STEP (PRIMARY VS RESCREENING) AND INTERVAL

ER Myers, DC McCrory, S Subramanian, S Datta, K Nanda, LA Bastian, DB Matchar Duke University and Durham VA Medical Centers, Durham, NC, and Health Economics Research, Inc , Boston, MA

Purpose To assess the impact of improving the sensitivity of Pap smear screening on the cost-effectiveness of strategies for the prevention of cervical cancer

Methods Cost-effectiveness was measured as dollars/life-year saved ($/LY), discounted at an annual rate of 3% We constructed a 20-state Markov model, including no disease, human papillomavirus (HPV) infection, preinvasive cervical neoplasia, cervical cancer at various stages, hysterectomy, and death from cervical cancer or other causes, and performed a cohort simulation from age 15 to 85 Transition probabilities were obtained from the literature and validated by comparing the age-specific incidence of cervical cancer without screening predicted by the model to reported population-based data Costs in 1997 dollars were estimated using data from MEDSTAT Sensitivity of 78% and specificity of 90% for conventional Pap smear screening were estimated from a meta-analysis We compared conventional Paps to technologies which either increase the sensitivity of the primary screening step to 95% at an incremental cost of $5/slide, or allow rescreening of all smears initially classified as negative at a sensitivity of 95% and an additional cost of $3 75/slide, at varying screening intervals

Results	Strategy	Avg Cost	Life Expectancy	Incremental C/E (compared to prior strategy)
	No Screening	$792 95	27 752498	-
Conventional Pap	q3yrs	$1,234 57	27 764600	$36,490
ImprovedPrimary	q3yrs	$1,285 30	27 765488	$57,117
ImprovedRescreen	q3yrs	$1,517 70	27 765644	$1,490,664
Conventional Pap	q1yr	$2,514 38	27 766520	$1,138,367
ImprovedPrimary	q1yr	$2,649 62	27 766825	$442,463
ImprovedRescreen	q1yr	$3,311 83	27 766825	(Dominated)

Conclusions Technologies which improve the sensitivity of the primary screening step in cervical cytology are more cost-effective than those which improve rescreening sensitivity The cost-effectiveness of either technology relative to conventional Pap is dependent on screening interval

DECISION MAKING IN STAGE IV NON-SMALL CELL LUNG CANCER (NSCLC) PATIENT VALUES FOR SHORT AND LONG TERM GAINS

M D Brundage, R H Cosby, D Feldman-Stewart, R Gregg, W J Mackillop, Radiation Oncology Research Unit, Queen's University, Cancer Care Ontario

The purpose of this study was to determine how patients value survival outcomes when considering palliative chemotherapy [Cisplatin based chemotherapy and best supportive care (C+BSC) over best supportive care (BSC) alone] as treatment for metastatic NSCLC Sixty people previously treated for a GU, breast, or skin cancer participated as surrogate decision makers for a Stage IV NSCLC patient We used a decision-aid interview that included treatment trade-off exercises to clarify participants' values concerning 1 year and median survivals Each exercise involved systematically changing the survival probability expected with C+BSC until the participant reached his/her threshold for choosing C+BSC The difference between the survivals associated with each treatment was the survival advantage threshold (SAT)

Participants' SATs ranged from choosing the more toxic treatment if it offered no survival advantage to declining C+BSC no matter how large its advantage Twenty-four (42%) would take the more toxic treatment for an absolute 1 year survival advantage of 5% (i e 20% for C+BSC vs 15% for BSC) and 33 (58%) would take it for an absolute 1 year survival advantage of 10% (the actual estimated benefit) Eleven (19%) would take C+BSC for a median survival advantage of 2 weeks (4 5 months for C+BSC vs 4 months for BSC) and 21 (36%) would take C+BSC for a median survival advantage of 1 5 months (the actual estimated benefit) Participants' SATs were not significantly related to sex, preferred role in treatment decision making, smoking history, or previous chemotherapy Younger participants (<70 years, U=268 0, p= 0169) and more educated participants (>high school, U=250 0, p= 0185) had significantly lower thresholds but each of these factors accounted for <5% of the SAT variance For 44 (76%) of participants, their stated treatment preference was consistent with their SATs for both median and 1 year survival For two (3%) participants, neither SAT was consistent with the stated treatment preference For the remaining 12 (21%) participants, one SAT was discordant with their treatment preference All participants were asked which of the two time points was more important to their treatment decision and for all 12 discordant cases, the SAT of the more important time point was consistent with their stated treatment preference

There is wide variation in surrogate patients' thresholds for choosing C+BSC with respect to both median and 1 year survivals Participants' SATs cannot be predicted reliably by their demographic characteristics The differences in how participants weigh the survival advantages suggest that patients who wish to participate in decision making should be provided with information about the benefits with respect to both time points

THE INFLUENCE OF INFORMATION ORDER ON PATIENT DECISION MAKING

GR Bergus, IP Levin, C Johnson, The University of Iowa Iowa City IA

Background Human judgment is sensitive to the order in which an individual receives information This phenomenon has been demonstrated with physicians The impact of information order on patient decision making and consent has not been fully investigated

Methods 309 patients at a primary care site were asked to review information about two procedures influenza vaccination and carotid endarterectomy Patients received brochures containing a general introduction followed by specific information about the risks and benefits of the interventions Patients were randomized to review a brochure with the risks detailed either before or after the benefits (Groups RB and BR) Patients were asked to estimate their favorability towards the procedures (on a 100 point scale) after the introduction and again after learning the risks and benefits The final estimates of the two groups and the differences between the initial and final estimates were analyzed using the Wilcoxon Rank Sum Test After reviewing all the information patients were asked whether they would consent to the intervention and completed the Satisfaction with Decision Scale

Results The order of information influenced patients' final favorability towards the influenza vaccine (median 95 vs 80 p=0 02) as well as the change in favorability between the first and final assessments (p=0 01) Additionally 60% of the RB group reported that they would agree to the vaccine compared to 48% of the BR group (p=0 04) Both groups were equally satisfied with their decisions (p=0 43) The order of risks and benefits did not impact patient judgment about undergoing carotid endarterectomy The two groups made similar final estimates of the favorability of surgery (median 80 vs 80 p>0 05) and made similar revisions in course of reviewing all the information (p>0 05) A similar proportion of patients reported that they would agree to the surgery (48% vs 42% p=0 25) and both groups were similarly satisfied with their decisions (p=0 49)

Conclusion The order of information had an inconsistent influence on patient judgment A recency effect was demonstrated with a procedure of little risk but no order effect was found with a procedure of significantly greater risk

BONE MARROW TRANSPLANT VERSUS PERIODIC PROPHYLACTIC BLOOD TRANSFUSIONS IN SICKLE CELL PATIENTS AT HIGH RISK OF ISCHEMIC STROKE

PJ Nietert, M R Abboud, MD Silverstein Center for Health Care Research, Medical University of South Carolina

Objective Measurement of cerebral blood velocity has been used to identify patients with sickle cell disease (SCD) at high risk of ischemic stroke Periodic transfusions have been shown to significantly decrease this risk , although it is uncertain how many years such a patient should undergo transfusions The goal of this study was to use decision analysis to help determine whether bone marrow transplantation (BMT) or periodic prophylactic blood transfusions (PBT) maximizes the number of quality-adjusted life years (QALY's) in such patients

Methods A Markov decision analytical model was constructed to represent the clinical course after bone marrow transplantation as well as the course of disease during periodic blood transfusions Medical literature and expert opinion provided utility values for stroke patients, estimates of transition probabilities from one clinical state to another, and risks of stroke and death for different disease states

Results Assuming 90% compliance with PBT therapy and no discounting of future life years, BMT is the optimal strategy, although the gain in future QALYs is not substantially larger (< 1 QALY) than that offered by PBT If future QALYs are discounted at 3%, PBT is a slightly better choice than BMT (< 1 QALY) As compliance with PBT decreases, BMT is the preferred treatment choice Assuming 90% compliance with PBT and a 3% discount of future QALYs, the choice of whether BMT or PBT is better is dependent on the duration (in years) of PBT therapy

Conclusions Whether these SCD patients would benefit more from BMT or PBT is highly dependent on several variables including the discounting of QALYs, the degree of compliance among those receiving PBT, and the duration necessary for transfusion to sufficiently decrease risk of stroke Although each strategy appears favorable under certain conditions, the gain in QALYs is small

COST-EFFECTIVENESS ANALYSIS OF IRINOTECAN COMPARED WITH BEST ESTIMATED INFUSIONAL 5-FU REGIMEN IN PATIENTS WITH ADVANCED COLORECTAL CANCER

N Neymark, EORTC, I Adriaenssen, EORTC, C Schmitt, ARCOS, Issy les Moulineaux, France

Objective A cost-effectiveness analysis was carried out from the viewpoint of the Belgian health insurance system of Irinotecan (CPT) versus best estimated infusional 5-FU regimens in patients with advanced colorectal cancer

Methods In a phase III RCT comparing CPT with best estimated infusional 5-FU regimens,[1] resource utilization data were collected with the purpose of carrying out an economic evaluation These data are supplemented by a survey of the investigators to determine the resources usually used for the administration of these chemotherapy regimens The de Gramont (DG) regimen is administered either with full hospitalization (DG-F) or with day clinic visits (DG-D), while CPT, the Lokich (LO) and the weekly high dose (AIO) regimens are all administered on a day clinic basis The problem with censored cost data is handled by a recently proposed non-parametric method [2] The prevailing tariffs of the Belgian health insurance system are used for cost valuation Sensitivity analyses are carried out with variability determined by the limits of the confidence intervals estimated for the patient data sampled in the trial

Results The trial documented a statistically significant survival benefit of CPT compared with the 5-FU regimens [1] Expected total costs per patient amount to (rounded figures) 495,000 BF for CPT, 489,000 BF for DG-F, 413,000 BF for DG-D, 332,000 BF for LO and 594,000 BF for AIO CPT is a dominant solution compared to AIO With the survival benefit standardized as life years gained (LY), the incremental costs/LY for CPT becomes 56,000 BF/LY compared to DG-F, 746,000 BF/LY versus DG-D and 1,480,000 BF/LY versus LO The results are quite sensitive to variations in the variables Lacking Belgian reference data, it may be mentioned that the currently often cited limit of acceptability for incremental costs/LY 50 000 $US, corresponds to 1 87 mio BF (using 1997 Purchasing Power Parities)

Conclusions The study shows that the cost-effectiveness of CPT compared to infusional 5-FU regimens varies considerably according to the specific comparator and is decisively determined by institutional policy with regard to the administration of regimens, cf the divergency between the two variants of the DG regimen

[1] van Cutsem et al, abstract *984, p256a in ASCO Proceedings, J Clin Oncol 17, 1998
[2] Cf abstract submitted to this SMDM meeting by C Schmitt et al

HUMAN STICK FIGURES OR BAR GRAPHS? A QUALITATIVE STUDY OF HOW RISK PRESENTATION EFFECTS RISK PERCEPTIONS

MM Schapira*, AB Nattinger*, CA McHorney** * Medical College of Wisconsin, Milwaukee, WI and ** University of Wisconsin, Madison, WI

The purpose of this study was to explore preferences for the format of breast cancer risk information and the effect of format on risk perceptions Four focus groups consisting of 10-11 women each were conducted Focus groups were designed to be homogenous with respect to age (younger 40-50 years, older 51-65 years) and education (college degree vs no college degree) Women of Caucasian, American-Indian, and African-American ethnic groups were represented The focus group guide raised the following topics general interest in quantitative risk information, use of human stick figures vs bar graphs to convey risk estimates, highlighting of human stick figures in random vs consecutive positions, presentation of percentages (i e 10%) vs frequency (1/10), and the presentation of risk information over 1 yr vs 10 yr vs lifetime periods

The focus group analysis revealed a strong interest in personalized and quantitative risk information General population based risk information was not felt to be relevant to personal risk Highlighted human stick figures were preferred to bar graphs as a visual tool to convey risk estimates The human stick figures were felt to make the risk estimates more personal Most subjects felt that consecutively highlighted stick figures were easier to understand than randomly highlighted stick figures when presented with the lifetime risk of breast cancer of 90/1000 However, some subjects felt that the random presentation was a more realistic presentation of risk A larger denominator (9/100 vs 9/10) was felt to be a more reassuring presentation of identical risk information because of the increased number of stick figures represented without breast cancer A larger denominator (9/100 vs 9/10) was also felt to be a more realistic presentation of risk Younger subjects were more likely to want risk information over a 10 year time-span because of its relevance to their immediate future Older subjects were more likely to want lifetime risk Annual risk information was not felt to be helpful because the risk must be faced repeatedly each year

In conclusion, patients may not believe that breast cancer risk information is relevant to them unless it is tailored to their personal risk factors When using visual aids to present risk estimates, highlighted human stick figures have more salience than bar graphs Changes in visual formats such as the number of stick figures used in the denominator and random vs consecutive grouping of highl ghted stick figures can change perceptions of the risk estimate being conveyed

SUBSTITUTING DEATH IN THE STANDARD GAMBLE ARE PATIENT UTILITIES INTERNALLY CONSISTENT WHEN THE HEALTH STATES ARE HIGHLY SALIENT?

ED Kennedy BI O'Connor M Varkul Z Cohen AH Steinhart HA Llewellyn-Thomas AS Detsky RS McLeod University of Toronto Toronto Ontario CANADA

Previous work implies that standard gamble (SG) utilities for hypothetical health states are internally inconsistent when the gamble's loss outcome is systematically varied, but it is unknown if this inconsistency would remain when non-hypothetical health states are considered The study purpose was to determine if patients' SG utilities for salient health states remain internally consistent when the gamble's loss outcome is modified

One hundred patients with Crohn's disease (CD) performed 3 SG tasks SG1 elicited the utility for Certain Recurrence of CD, using Perfect Health and Death as the gamble outcomes Next, utilities were obtained for taking 5-ASA with 25% chance of recurrence (SG2), and for not taking 5-ASA with 40% chance of recurrence (SG3) in each of these tasks Perfect Health and Certain Recurrence of CD were used as the gamble's gain/loss outcomes Rationally the individual retaining his/her utility for Certain Recurrence of CD will be indifferent when the probabilities are 75% 25% in SG2 and 60% 40% in SG3

In SG1, 18 patients refused to accept any chance of death in the gamble The results of this subgroup of highly risk-averse individuals are not included in the analysis reported here For the remaining participants, the rationally-predicted (75%) and observed (76%, +/-14%) indifference points in SG2 were not significantly different For SG3, the rationally-predicted (60%) and observed (70%, +/-15%) indifference points were significantly different ($t = 6\ 67$ $p < 0\ 05$), however, this did not appear to translate into utility instability When we used the observed utility for Certain Recurrence of CD (0 90 +/- 0 11) and the axiom of substitutability, the predicted utilities for SG2 and SG3 were 0 97 +/-0 02 and 0 96 +/-0 04 respectively, the relevant observed utilities (0 97 +/- 0 04 and 0 97 +/- 0 05) were not significantly different from the predicted utilities

When death was replaced by a salient clinical outcome in the SG the shift in the reference point predicted by prospect theory was not observed The substitute gamble outcome itself was highly-valued (utility > 0 90) and may have generated this stability thus further methodologic work is indicated to identify the threshold utility for the loss outcome at which shifts in the reference point are stimulated

DO STUDENTS REMEMBER WHAT THEY WERE TAUGHT ABOUT DECISION ANALYSIS?

HP Lehmann, CM Wiener Johns Hopkins University, Baltimore, MD

Objective· The SMDM and the AAMC are both publishing learning objectives in decision analysis (DA) There are few data on teaching DA to medical students.

Subjects. 24 second-year medical students self selected themselves to take a 5-week (10-hour) selective in decision analysis (DA) in the first two months of their second year and 96 who took another selective (nonDA)

Outcome measure As part of the final pathophysiology exam 6 months after the selectives, all students wrote a justification for obtaining a computed tomogram in the setting of a young woman's "worst headache" she'd ever had Responses were scored in a blinded fashion for the presence of language consistent with DA principles taught in the selective, following the acronym You SHOULD You (decision maker), Structure (DA argument), History, Outcomes, Uncertainty, aLternatives and Desires/traDeoffs

Results Data were collected on 114 students (95%) No single or aggregate score had a significant association with DA status, including grades No student addressed issues of who the decision maker is and none drew a decision tree Hierarchical clustering found no meaningful pattern

	Y	S	H	O	U	L	D	Total (SE)
DA	0%	18%	76%	62%	85%	72%	47%	3 7 (0 3)
nonDA	0%	14%	81%	76%	86%	62%	52%	3 6 (0 2)

Conclusions Students evaluated in a non-DA setting have an implicit understanding of DA principles unaffected by having taken a DA selective To assess the impact of learning DA explicitly, we need different measures (or a better selective!), but our design can be used by others to assess the effectiveness of their teaching DA

COST-EFFECTIVENESS OF TROPONINS IN PATIENTS WITH CHEST PAIN WHO ARE AT LOW RISK FOR MYOCARDIAL INFARCTION

Sundar Natarajan, Bruce Kinosian, Division of General Internal Medicine, University of Pennsylvania, Philadelphia, PA 19104

Background: Using clinical and electrocardiographic criteria it is possible to identify subjects at low risk for myocardial infarction (MI) Hospitalization of such low-risk patients to rule out myocardial infarction (MI) imposes a heavy burden on the health care system Thus different strategies have been developed to identify patients likely to have MI

Objective: The objective of this study was to evaluate alternate strategies of managing low-risk patients with chest pain

Design: Cost-effectiveness analysis using decision analytic modeling

Methods: The strategies we compared included A) Hospitalization, B) Discharge, and C) Testing all low-risk patients and admitting test positives The testing strategies included i)Troponin I, ii) Troponin T, iii) Echocardiography, and iv) Observation unit Data inputs for probabilities and test characteristics were from the published literature Costs were estimated from Medicare relative value scales and from published data The reference case is a 50 year-old man The time frame was 30 days while the analytic horizon was the lifetime of the individual Life expectancy was derived from the Framingham heart study and actuarial data The results are expressed as incremental cost-effectiveness ratios (CER)

Results: The incremental CER of Troponin T compared to discharging all patients was $64, 024 per life year saved (LYS) The incremental CER of Troponin I was $59,031 per LYS and the incremental CER of hospitalization was $215,482 per LYS The echocardiography and observation strategies were dominated The relative advantage of the Troponin tests was preserved in sensitivity analyses

Conclusion: The results suggest that testing for troponins is a relatively safe and cost-effective strategy for managing low-risk patients in the emergency department Among the troponins, testing using Troponin I seems to be the superior test

Decision Analysis of Endoscopic Mucosal Resection Therapy for Mucosal Gastric Cancer.

Aoki N [1, 2], Kajiyama T [3], JR Beck [1], Fukui T [2]
1) Information Technology, Baylor College of Medicine, Houston Tx 2) Department of General Medicine and Clinical Epidemiology, Kyoto University Hospital, Kyoto, JAPAN 3) First Department of Internal Medicine, Kansai Denryoku Hospital, Osaka, JAPAN

Objectives: To evaluate the effectiveness of Endoscopic Mucosal Resection (EMR), a new minimally invasive therapy for mucosal gastric cancer (mGC) patients and the factors that affect the decision

Methods: Markov decision analysis was used to compare EMR and elective surgery for patients with mGC over a five-year period Outcomes were defined in terms of cumulative quality-adjusted life months (QALMs) Transition probabilities were obtained from published clinical studies and integrated with meta-analysis Utilities were elicited from expert panels Tornado diagram analysis was performed regarding all probability and utility values to clarify the influence of each factor Monte-Carlo probabilistic sensitivity analysis was undertaken to simulate the uncertain clinical situation

Results: Expected utility analysis for 50-year-old men showed that 47 780 QALMs were achieved by the EMR cohort and 47 549 with elective surgery A tornado diagram analysis revealed that there were three variables having a large potential effect on QALMs, incidental lymph node metastatic (LN (+)) rate, and mortality due to elective operation and EMR procedure Monte-Carlo sensitivity analysis of 10,000 samples showed that the EMR strategy had greater QALMs in 96 38 percent of cases Stratified Monte-Carlo simulation showed that the percentage of greater QALMs in EMR strategy was 100%, 86%, 1%, and 0% if all subjects had protruded mGC, all depressed mGC, all poorly differentiated mGC and submucosal gastric cancer, respectively, (regardless of any other factors)

Conclusions Decision favoring EMR was shown highly dependent on incidental LN (+) rate and mortality of therapeutic procedure Many factors affect LN (+) rate, such as size, existence of ulceration, macroscopic type, and histological findings, which directly influence the decision Further clinical trials may identify a reliable method to predict the incidental metastatic rate of mGC

COMPUTER-ASSISTED POPULATION-DERIVED UTILITY ELICITATION FOR HEPATITIS C-RELATED CIRRHOSIS AND HEPATOCELLULAR CARCINOMA

JC Dube, AMK Rostom G Nichol, A Laupacis Division of Gastroenterology University of Ottawa Ottawa ON and Clinical Epidemiology Unit, Ottawa Civic Hospital Ottawa, ON Canada

Population-based measures of preference for cirrhosis are not currently available and are needed for the conduct of cost-effectiveness analyses based on the recommended Reference Case The use of computerized interviews for utility elicitation offers the advantages of reproducibility of the conduct of the interview with minimal assistance and bias from the interviewer It also theoretically provides respondents with more control over the pace of the interview than conventional methods of utility elicitation

The aim of this project was to obtain population-based measures of preference for hepatitis C-related compensated cirrhosis, decompensated cirrhosis and hepatocellular carcinoma (HCC)

Methods Scenarios for each of these three health states were constructed after review of the literature on quality of life in hepatitis C and end-stage liver disease as well as review with an expert panel The scenarios were constructed using the descriptors of the Health Utility Index, a generic preference measure We used a Windows based software authoring package to create a computerized utility program which integrates text, audio, and graphical visual aids into a 30 minute interview yielding standard gamble derived utilities on hypothetical scenarios The program allows respondents to review previously presented material, and prints a summary of the utility scores A convenience sample of 40 respondents were randomly selected among individuals who were accompanying outpatients undergoing an endoscopic procedure at our institution Results Mean utility scores for compensated cirrhosis, decompensated cirrhosis and hepatocellular carcinoma were 0 93, 0 51, and 0 28 respectively The interview was completed by all included respondents Results obtained for compensated cirrhosis were consistently higher than those of decompensated cirrhosis which were, in turn, higher than the HCC utility scores

In conclusion, this project has enabled us to generate population-based utilities for health states related to chronic liver disease It also shows the feasibility of programming customized multimedia computer aids for utility elicitation

DEVELOPMENT AND PILOT EVALUATION OF AUTOMATED COMPUTER-BASED CREATION OF SITE-SPECIFIC CLINICAL-PRACTICE GUIDELINES FROM DECISION MODELS

GD Sanders, RF Nease Jr , DK Owens Stanford University, Stanford, CA, Washington University, St Louis, MO, and VA Health Care System, Palo Alto, CA

Previous research indicates that variation in patient populations and practice settings often warrants tailoring clinical-practice guidelines (CPGs) to specific practice settings Despite such needs, development of CPGs requires experts in medicine and evidence synthesis who may be unavailable in many practice settings We developed a computer-based system that enables developers and users to create, disseminate, and tailor CPGs, using normative decision models (DMs)

We developed ALCHEMIST, a web-based system that analyzes a DM, creates a CPG in the form of an annotated algorithm, and displays the optimal strategy as determined by the DM ALCHEMIST's interface enables remote users to tailor the guideline by changing input variables for the guideline and observing the new annotated algorithm that is developed automatically We used a DM that evaluates strategies for staging non-small-cell lung cancer to perform a pilot evaluation of the system Subjects (n=15) compared the automatically created guideline to published guidelines for staging non-small cell cancer and critiqued both the computer-based and published guidelines with a previously developed instrument to assess the quality of the guidelines Each subject rated the CPGs for usability, accountability, and accuracy on a scale of 0 to 2

The mean overall score for the ALCHEMIST CPG was 1 502, compared to the published-CPG score of 0 987 (p = 0 002) The ALCHEMIST CPG scores for usability, accountability, and accuracy were 1 683, 1 393, and 1 430 respectively, the published CPG scores were 1 192, 0 941, and 0 830 (each comparison p < 0 05) On a scale of 1 (worst) to 5 (best) users mean rating of ALCHEMIST s ease of use, usefulness of content, and presentation format was 4 76, 3 98, and 4 64, respectively

Our study demonstrates the feasibility of a web-based system that automatically analyzes a DM and creates a CPG as an annotated algorithm The system enables remote users to develop a site-specific CPG tailored to the user's practice setting Our pilot evaluation indicates that the ALCHEMIST guideline meets published criteria for quality and compares favorably to published guidelines by national organizations The high ratings for usability and usefulness suggest that such computer-based systems can provide an important tool for guideline development

COST-EFFECTIVENESS OF CHOLESTEROL-LOWERING THERAPIES ACCORDING TO INDIVIDUAL PATIENT CHARACTERISTICS

LA Prosser, AA Stunnett, PA Goldman, MGM Hunink, L Goldman, MC Weinstein Harvard School of Public Health, Boston, MA, University of Alabama at Birmingham, AL, Erasmus University Medical School, Rotterdam, The Netherlands, and University of California, San Francisco, School of Medicine, San Francisco, CA

The purpose of this study is to evaluate how the cost-effectiveness (CE) ratios of cholesterol-lowering therapies for primary and secondary prevention of coronary heart disease vary with different risk factors age, gender, smoking status, blood pressure, LDL-cholesterol level, and HDL-cholesterol level

Treatment strategies evaluated were Step 1 Diet, pravastatin (primary prevention), simvastatin (secondary prevention), and no preventive treatment The analysis used a state-transition computer simulation model, the Coronary Heart Disease Policy Model, to estimate the effects and costs of each strategy in different risk groups for a period of thirty years Incremental CE analysis of primary prevention strategies assumed secondary prevention with simvastatin Secondary prevention strategies were compared to no preventive treatment Modeling assumptions followed the US Public Health Service Panel recommendations for conducting a "reference case" cost-effectiveness analysis

Optimal strategies for primary prevention ($/QALY) based on a threshold of $100,000/QALY:

Treatment	All Persons with LDL>159 mg/dl		Very High Risk[1] Persons with LDL>189 mg/dl		Very Low Risk[2] Persons with LDL>189 mg/dl	
By Age	Men	Women	Men	Women	Men	Women
35-45	Diet	No Therapy	Pravastatin	Diet	Diet	No Therapy
45-55	Diet	No Therapy	Pravastatin	Diet	Diet	No Therapy
55-65	Diet	Diet	Pravastatin	Diet	Diet	Diet
65-75	Diet	Diet	Pravastatin	Pravastatin	Diet	Diet
75-85	Diet	Diet	Pravastatin	Pravastatin	Diet	Diet

[1] 3 risk factors according to NCEP guidelines (DBP>94, Smoker, HDL<35)
[2] -1 risk factors according to NCEP guidelines (DBP<94, Non-smoker, HDL>49)

Using a threshold of $100,000/QALY, pravastatin is cost-effective only in high risk subgroups Pravastatin becomes cost-effective for some very high risk individuals at a threshold of $60,000/QALY and for some very low risk individuals at $150,000/QALY Secondary prevention appears to be cost-effective across patient subgroups (CE ratios < $50,000/QALY)

RELATIONSHIP BETWEEN PATIENT PREFERENCE VALUES AND SATISFACTION WITH DIALYSIS CARE

S Wills, E Bass, N Fink, H Rubin, M Jenckes, A Levey, K Meyer, J Sadler, N Powe Johns Hopkins University University of MD Baltimore MD New England Medical Center Boston MA

Numerous studies have examined how patient preference values (PVs) relate to other measures of quality of life but little is known about how PVs relate to satisfaction with medical care The aim of this study was to assess how PVs assigned to current health (CH) by patients (pts) with end-stage renal disease (ESRD) relate to their satisfaction with dialysis care We interviewed ESRD pts using a time trade-off technique scaled between 0 (death) and 1 (perfect health) to measure pts PVs for their health on their current dialysis treatment We also asked pts about their satisfaction with ease in reaching the doctor ease in reaching staff accuracy of information provided by the doctor attention to cleanliness sensitivity to pts needs staff availability in an emergency information provided at treatment onset and overall quality of care using a 5-point ordinal scale

A total of 188 pts were interviewed 109 received hemodialysis (HD) 57 received continuous ambulatory peritoneal dialysis (CAPD) and 22 received continuous cycling peritoneal dialysis (CCPD) The mean age was 53 3 years 42% were female 73% were African American and 33% did not have a high school degree In the total group PVs for CH were positively correlated with satisfaction with 3 aspects of their dialysis care ease in reaching the doctor (Spearman rho=0 15, p< 05) accuracy of information provided by the doctor (Spearman rho=0 24 p< 001) and staff availability in an emergency (Spearman rho=0 15 p< 05) In multivariate analyses after adjusting for differences in pts age gender race type of dialysis (HD vs PD) education and score on the Beck Depression Inventory PVs for CH were independently related to pts ratings of the accuracy of information provided by the doctor but not to other measures of patient satisfaction with care

These results highlight the importance of the information that physicians give to pts with serious chronic diseases like ESRD and suggest that the accuracy of information may help patients adjust to the limitations imposed on them by ESRD and dialysis care

PREDICTORS OF INFLUENZA VACCINE ACCEPTANCE AMONG HEALTHY ADULTS

EJ Coups GB Chapman Rutgers University, New Brunswick, NJ

Previous studies of predictors of influenza vaccine acceptance have typically focused on populations at high risk for complications from influenza, or healthcare workers Many predictors identified in these studies relate to risks and benefits of the vaccine, although little attention has been paid to the timing of such risks and benefits The present studies differ from these studies in two ways Firstly, predictors of vaccine acceptance were investigated among healthy adults, a population for whom the vaccination has significant health-related and economic benefits Secondly, the role that personal time preferences may play in the decision to obtain vaccination was examined Time preferences indicate the extent to which people value future events relative to present events The decision to accept influenza vaccination implicitly involves an intertemporal choice the inconvenience and possible side effects occur immediately whereas the benefits of influenza reduction occur months later Thus influenza vaccine acceptance should be related to future oriented time preferences

In the first study 79 University employees (39 vaccine acceptors and 40 decliners) were interviewed and asked questions concerning their medical history attitudes and beliefs regarding the influenza vaccine and health and monetary time preferences Three significant predictors of vaccine acceptance were identified perceived effectiveness of the vaccine (OR=2 27) perceived likelihood of a reaction to the vaccine (OR=0 41) and receiving the vaccine last year (OR=9 22) Time preferences did not provide any additional predictive power

A second study was carried out with a sample of 435 corporate employees (64% response rate) The following were significant predictors of vaccine acceptance perceived effectiveness of the vaccine (OR=3 07) perceived likelihood of a reaction to the vaccine (OR=0 61) receiving the vaccine last year (OR=27 6) age (OR=1 37) and percentage of co-workers accepting the vaccine (OR=1 59) Monetary (but not health) time preference was a significant predictor (OR=3 12), even after accounting for these other factors

The predictors of vaccine acceptance that we identified in our samples of healthy adults are largely congruent with previous findings on high-risk populations and healthcare workers We also found some evidence of an association between time preferences and influenza vaccination

MANAGEMENT OF MYCOBACTERIUM AVIUM COMPLEX IN AIDS USING MONTE CARLO METHODS TO ASSESS STRATEGIES FOR DISEASE PREVENTION AND TREATMENT

DS Zingmond, SA Bozzette VA Health Care System, San Diego, CA and University of California, San Diego

Disseminated Mycobacterium Avium Complex (MAC) is a systemic opportunistic infection which is a cause of increased morbidity and mortality in AIDS Macrolide antibiotics are the basis for the most effective prophylaxis and treatment regimens However, macrolide resistance is frequent among patients failing prophylaxis, forcing the use of less effective regimens for treatment Using Monte Carlo Methods, we performed simulations to evaluate trade-offs in total survival and disease-free survival between different combinations of prophylaxis and treatment regimens by CD4 strata and rates of macrolide resistance

A state transition model was created to model the combination of prevention and treatment strategies and address trade-offs in predicted total survival and disease-free survival Time-dependent state transition probabilities and risk adjustments were calculated directly from primary clinical trial data Disease-free survival was defined as time without disease plus time with symptom-free disease MAC bacteremia was used as a proxy for symptomatic disease Sensitivity analysis on spontaneous macrolide resistance rates after prophylaxis failure was performed to estimate the breakpoint favoring prophylaxis versus no prophylaxis

For CD4 positive cell count strata less than 100 cells/mm3, median total survival varied from 18 months for no prophylaxis and no treatment to 43 months for prophylaxis of azithromycin plus rifabutin followed by treatment with clarithromycin, rifabutin, and ethambutol Disease-free survival was 12 and 40 months respectively No prophylaxis followed by treatment had intermediate results Across a wide range of spontaneous macrolide resistance rates, estimated total survival remained superior among those receiving prophylaxis For symptom-free survival, estimated breakpoints for rates of resistance were uniformly greater than published spontaneous resistance rates

Across CD4 positive cell count strata of less than 100 cells/ mm3, prophylaxis with a macrolide-based regimen has superior total survival and disease-free survival versus a 'watch-and-wait" approach of treatment alone in patients becoming infected with MAC

PATIENT PREFERENCES AND TEMPORAL DISCOUNTING FOR HEALTH STATES ASSOCIATED WITH HORMONE REPLACEMENT THERAPY A PILOT STUDY

AL Drapkin and ER Myers Duke University Medical Center, Durham, NC

Purpose To compare utilities for health states associated with hormone replacement therapy (HRT) elicited by visual analog scale (VAS) and standard gamble (SG), and to determine if preferences for delayed outcomes as predicted by the exponential discount model differ between health states

Methods We developed standardized descriptions for five health states associated with hormone replacement therapy menopausal symptoms, side effects of HRT, breast cancer, myocardial infarction, and osteoporosis Using the software program U-Titer II , we assessed utilities using VAS and SG for 4 time periods the present and 10, 20, and 30 years in the future Perfect health and death were used as anchors Discount rates for each health state were calculated using standard methods Patient opinions about the instrument were obtained at the end of each session We recruited women between the ages of 45 and 55 without a history of breast or endometrial cancer who considered themselves eligible for HRT

Results Sixteen women who met our criteria completed the computerized interview within 30-60 minutes Both utilities for health states and discount rates varied widely between subjects

	UTILITY				DISCOUNT RATE			
	VAS		SG		VAS		SG	
	MEAN	RANGE	MEAN	RANGE	MEAN	RANGE	MEAN	RANGE
Menopause	0 582	0-0 995	0 864	0 9-1 0	2 45%	-2 6-15 5%	0 43%	-0 35-7 3%
HRT	0 616	0 03-1 0	0 895	0 5-0 99	1 28%	-1 4-9 8%	-0 04%	-4 8-1 5%
Breast CA	0 421	0 03-0 88	0 700	0 12-0 99	1 04%	-2 0-6 8%	0 11	-7 1-7 2%
MI	0 389	0 48-0 83	0 767	0 1-0 99	2 30%	-1 2-16 2%	0 31	-6 7-6 0%
Osteoporosis	0 483	0 33-0 9	0 756	0 45-0 99	1 70%	-4 5-6 5%	0 89	-1 5-24%

Patient responses to the instrument were uniformly positive

Conclusions Although there was wide variation, both the utility and discount rate for breast cancer were lower than those for other health states associated with HRT for many of our subjects This may reflect a magnitude effect for breast cancer Eliciting individual utilities and discount rates may be important in the design of clinical decision analytic models for HRT counseling

USING ESTROGEN TO REDUCE SERUM CHOLESTEROL IN POST-MENOPAUSAL WOMEN A COST-EFFECTIVENESS ANALYSIS

RM Werner, AI Mushlin University of Pennsylvania, Philadelphia, PA, and University of Rochester, Rochester, NY

Objective To determine the incremental cost-effectiveness of using estrogen replacement therapy and HMG-CoA reductase inhibitors compared with standard therapy of Step I diet in post-menopausal women with hypercholesterolemia

Methods We developed a decision analytic model from a societal perspective with a one year time horizon The model specified that patients were either treated with step I diet alone, HMG-CoA reductase inhibitors alone, or estrogen with or without an HMG-CoA reductase inhibitor, depending on response to estrogen Outcomes were measured as the liklihood of acheiving the LDL-C goal set by the National Cholesterol Education Program for three CHD risk groups <2 risk factors, ≥2 risk factors, or CHD Costs (1998 US dollars) were derived from the Medicare reimbursement schedule and average retail prices for pharmaceuticals Sensitivity analyses were conducted on the costs of pharmaceuticals and the probabilities of acheiving LDL-C goal for each risk group and therapy

Results In the baseline analysis, the incremental cost-effectiveness ratio for starting therapy with estrogen, with the option of adding an HMG-CoA reductase inhibitor if needed, ranges from $1,196 to $3,764 for each one percent of women who achieve LDL-C goal, depending on their risk group The strategy of starting therapy with an HMG-CoA reductase inhibitor, instead of estrogen, is dominated by estrogen for all risk groups Estrogen is cost-effective across a wide range of costs and estimates of effectiveness

Conclusions For post-menopausal women with hypercholesterolemia, initial treatment with estrogen appears to be a cost-effective stategy to reduce serum cholesterol HMG-CoA reductase inhibitors alone achieve more effect than estrogen alone, but by using estrogen as a first line therapy, the need to use relatively costly HMG-CoA redectase inhibitors is reduced

A COMPARISON OF METHODS FOR ACCOUNTING FOR TRUNCATED TRIALS OF LIFETIME THERAPIES

H. Glick, J Cook, J Heyse, B Kinosian University of Pennsylvania and Merck Research Labs, Philadelphia, PA

Objective To compare two methods for evaluating economic impacts from intermediate trials with observed mortality differences 1) projection of the costs and effects for the patients' lifetime (Lifetime projection) vs 2) evaluation of the costs observed within the trial and assignment of average life expectancy to those who survived the trial (Add on)

Methods We used data from 3 heart failure trials (follow-up from 3-5 years) and one cholesterol-lowering trial (5+ years of follow-up) to compare the two methods For method 1, we estimated lifetime survival using parametric failure time models, we estimated lifetime costs of hospitalization, ambulatory care, and the intervention based on rates observed within the trial For method 2, costs were taken from within the trial, those surviving the trial were assumed to have a life expectancy derived from the results of the failure time models for the placebo group Comparison of the two methods was based on the point estimates of the resulting ratios of cost per QALY saved (CER) as well as the 95% CI for the ratios (estimated using a jackknife procedure)

Results The point estimates and CI for the 2 methods in each of the 4 trials are reported in table 1 While the point estimates in general were similar, the CI appeared to differ For example, for trial 3, the lifetime method suggested the intervention was highly cost effective (ranging from dominance to an upper limit for the CI of $44,000, the add on method, on the other hand, suggested therapy might dominate (-$1000, quadrant IV) or be dominated (-$1783 million, quadrant II)

Table 1 Cost per QALY Saved ($000) and 95% CI

	Trial 1	Trial 2	Trial 3	Trial 4
Lifetime	12 (7 to 27)	14 (6 to 54)	5 (-2 to 44)	-332 (-20 to 15)
Add on	14 (7 to 49)	4 (-21 to 27)	5 (-1 to -1,783)	-2 (152 to 11)

Conclusion The point estimates for the ratios were similar for the two methods (although in theory they could differ) The CI varied substantially (likely due to differences in the correlations between costs and effects), which may affect policy recommendations Given that the two methods can differ and that the theoretical justification for the add on method is not clear, researchers should adopt it with care

POTENTIAL BIASES IN EVALUATING CANCER SCREENING A COMPARISON OF TWO APPROACHES FOR MODELING UNDERLYING DISEASE PROGRESSION

SJ Goldie, KM Kuntz Harvard School of Public Health, Boston, MA

Evaluating cancer screening often requires modeling the *underlying* disease process instead of *observed* disease, particularly in the absence of direct evidence linking screening to a survival benefit Natural history can be modeled as progression from no disease (ND) to precancerous lesions (PCLs) to invasive cancer (CA) The mechanism of screening is such that an individual with a PCL may be detected and if treated successfully, returned to a ND state There are data to support that these individuals are at a higher risk of progression to CA than those without prior PCLs

To examine the effect of two alternative approaches to modeling disease progression, we developed a simple cancer screening model to compare triennial screening with no screening Both approaches assume that the monthly transition from ND to PCL for persons with a PCL history (α_1) is greater than that of persons without a PCL history (α_0) The FIRST approach assumes that the transition from PCL to CA for persons with a PCL history (β_1) is greater than for persons without a PCL history (β_0), while the SECOND approach assumes that they are equal The screening test was assumed to be perfect and riskless The model was calibrated such that the FIRST and SECOND approaches predict the same 10-year cancer incidence

In the FIRST approach (α_0=0 002 α_1=0 10 β_0=0 015, β_1=0 03) the screening benefit was negative in that the life expectancy (LE) of no screening was 1 8 months greater than screening Using the SECOND approach (α_0=0 001, α_1=0 26, β_0=0 03, β_1=0 03), the screening benefit was positive (LE of screening was 0 36 months greater than no screening) To explicitly demonstrate the bias we used extreme values for the monthly transitions However, when applying these two modeling approaches to a more realistic analysis of cervical cancer screening in HIV(+) women, the FIRST ($\alpha_0<\alpha_1$, $\beta_0<\beta_1$) approach consistently underestimates screening benefit compared to the SECOND ($\alpha_0<\alpha_1$, $\beta_0=\beta_1$) approach, provided that both models are calibrated using the same observed long-term outcome data

When modeling the underlying disease process in cost-effectiveness analyses for cancer screening one should avoid the potential bias associated with an approach which assumes both the transition from ND to PCL and from PCL to CA is greater with a PCL history than for persons without a PCL history

DIAGNOSING STREP USING THE HISTORY AND PHYSICAL A META ANALYSIS

MH Ebell, HC Barry, MA Smith, M Carey Michigan State University, East Lansing, MI

The purpose of this study was to perform a meta-analysis of studies of the history and physical to diagnose group A beta-hemolytic (GABHS) pharyngitis We searched MEDLINE and personal files to identify candidate articles Each candidate article was independently reviewed by two investigators for inclusion criteria, study quality, and test characteristics Discrepancies between reviewers were resolved by consensus Only articles with the following characteristics were included predominantly enrolled patients with sore throat, used an independent, blind throat culture as the reference standard, collected data prospectively, and each patient received both clinical examination and throat culture Summary values for sensitivity, specificity, positive likelihood ratio (LR+), negative likelihood ratio (LR-), and area under the receiver operating characteristic curve (AUROCC) were calculated using standard methods Data for 64 clinical variables in 30 studies were included in the analysis

Symptoms and signs with the greatest AUROCC were tender anterior cervical nodes (0 72), exudate (0 71), history of fever (0 70), tonsillar enlargement (0 69), measured temperature > 100 F (0 67), enlarged anterior cervical nodes (0 66), and absence of cough (0 64) Those with the greatest LR+ (best for ruling in disease) were autumn season (3 7), chills (3 3), severely red throat (3 2), anorexia (2 9), hypertrophic papillae of tongue (2 5), positive throat culture in last year (2 3), and physician estimate (2 3) Signs and symptoms with the lowest LR- (best at ruling out disease) were cough and no recent exposure (0 23), no pain with swallowing (0 37), duration <= 3 days (0 38), hoarseness (0 39), no exudate or enlarged nodes (0 40), no history of fever (0 42), no marked redness of tonsils (0 46), and physician estimate (0 49) While the AUROCCs were consistent with previous work to develop clinical prediction rules, other variables such as autumn season, chills, anorexia, short duration of symptoms, no pain on swallowing, and hoarseness were useful in either ruling-in or ruling-out GABHS, but not both These variables should be measured in future studies that attempt to develop a clinical prediction rule for diagnosis of GABHS

CAN RISK OF MORTALITY WITHIN 12 MONTHS OF REVASCULARIZATION PROCEDURES BE PREDICTED BY VARIOUS QUALITY OF LIFE DOMAINS SCORES?

Robert Tilden, Carl Heltne, John Streitz, Frank Romano, Diane Kolquist, Colleen Reiner, SMDC Health System, Duluth, Minnesota

Objective: To review *Quality of Life* measures associated with 12 month risk for mortality among patients undergoing cardiac revascularization procedure

Methods: We collected *Quality of Life* information from patients with coronary artery disease (CAD), receiving Percutaneous Transluminal Coronary Angioplasty (PTCA) and those receiving Coronary Artery Bypass Graft (CABG), prior to the procedure, and at 6 and 12 months after the procedure We used the Health Status Questionnaire (HSQ) 2 0, a modified version of the SF-36 Four hundred seventy seven patients agreed to participate, of which 20 died within 12 months We reviewed baseline scores for those that died in eight domains of *Quality of Life* using z-scores (the number of standard deviations from age/gender specific standards) to control for age/gender differences

Results: Those experiencing mortality within 12 months reported lower scores at baseline for three domains of *Quality of life Role Emotional, Social Functioning*, and *General Health* The mean *Role Emotional* score for those dying was -0 9 standard deviations (sd) lower than the survivors, the *Social Functioning* score was 1 1 sd lower, and the *General Health* was 0 6 sd lower The risk of mortality was 2 9 times higher (95% ci 1 24 < rr < 6 83) for those cardiac revascularization patients with *General Health* scores < –1 sd from the norm, 3 13 times higher (95% ci 1 31 < rr < 7 49) for those with *Role Emotional* scores < -2 sd from the norm, and 3 9 times higher (95% ci 1 48 < rr < 10 31) for those with *Social Functioning* scores < –1 sd from the norm Those patients that were below the thresholds for all three scores had an 5 26 increase risk of mortality (95% ci 2 15 < rr < 12 86)

Conclusion: Patients reporting low scores for their age and gender for *Role Emotional, Social Functioning*, and *General Health* prior to revascularization had an increased risk of 12 months mortality following the procedure If extra care is taken in the rehabilitation of patients with low *Quality of Life* scores in *Social Functioning, General Health*, and *Role Emotional*, risk of mortality for cardiac revascularization patients could be reduced

WHAT INFORMATION IS NECESSARY FOR PATIENTS WITH CURABLE PROSTATE CANCER TO DECIDE ON TREATMENT?

D Feldman-Stewart[1], MD Brundage[1], BA McConnell[1], RH Cosby[1], CRR Hayter[1], JC Nickel[2], WJ Mackillop[1] Radiation Oncology Research Unit[1], Dept of Urology[2] Queen's University, and Cancer Care Ontario

The study was designed to determine what information men with early-stage prostate cancer think is necessary to know in order to choose their treatment from three options (surgery, radiotherapy, or "watchful waiting") Our previous research identified 59 items of information that might be considered necessary A survey was distributed to all men diagnosed with early-stage prostate cancer in a regional urology practice between July 1995 and Dec 1996 who had not participated in our previous survey Respondents were asked to consider the treatment decision for a specific hypothetical patient and to judge whether or not each of 59 items was "necessary" for them to know in order to make the treatment decision Items were presented as factual statements Respondents provided basic demographic information (age, education and marital status)

Of 89 eligible patients who received surveys, 69 (78%) responded The number of items that individual respondents identified as necessary ranged from 9-58 (mean 32, sd 12) Each item was considered necessary by between 20-93% of respondents (mean 54, sd 18) Of the 59 items, only 5 were necessary to at least 80% of respondents, 17 were necessary to at least 67% of respondents and 7 were not necessary to at least 67% of respondents Responses were more evenly split (maximum disagreement) on the remaining 28 items Although, for 16 of the items, there were statistically significant associations between responses and at least one demographic characteristic, only 5 of the associations accounted for more than 10% of the variance

These results suggest that there is large variation among patients on the number of information items deemed necessary for them to know for the treatment decision and little agreement on the necessity of most individual items Demographic characteristics are weak predictors of what information is necessary to the individual patient Therefore, the process of informing patients with curable prostate cancer needs to be flexible and able to accommodate a comprehensive range of information

UNCERTAINTY IN COST-EFFECTIVENESS ANALYSES A COMPARISON OF FOUR CONFIDENCE INTERVAL ESTIMATION METHODS

AP Johnson-Masotti, SD Pinkerton, Center for AIDS Intervention Research, Medical College of Wisconsin, Milwaukee, WI,
D Laurson, LA Aday, L Franzini, School of Public Health, University of Texas, Houston, TX

This research compared four statistical approaches (Taylor, Fieller, percentile bootstrap and percentile bias-corrected bootstrap) to estimating confidence intervals (CIs) around the cost-effectiveness (CE) ratio, and applied these techniques to data from the National AIDS Demonstration Research (NADR) project, which include indices for both sexual and drug-injection associated HIV risks

A Monte Carlo simulation was conducted for hypothetical CE data sets with diverse characteristics The results of this simulation study were then matched to the particular characteristics of NADR data, and the four CI estimation techniques were applied The main performance criterion was the percentage of times the estimated CIs contained the "true" CE ratio (specified a priori) A secondary criterion was the average width of the CIs that each method generated

Overall, the Taylor method outperformed the other methods when the coefficient of variation (CV) of effectiveness and of costs were small (10%) It might therefore be preferable to apply the simpler Taylor series method in this case Nevertheless, the percentile bootstrap performed better than the bias corrected bootstrap and the Fieller methods when the CV of effects was large (30%), and vice versa when the CV of effects was small (10%)

In conclusion, how well a particular CI estimation technique performs depends upon the characteristics of the particular cost and effects of the data set This suggests a need for more comprehensive assessment and reporting of important distributional parameters for clinical trials of health care and health promotion interventions

LINKING EVIDENCE-BASED MEDICINE (EBM) TO CLINICAL DECISION ANALYSIS (CDA)

B Djulbegovic and I Hozo University of Louisville, KY,Indiana University Northwest,IN

Evidence-based medicine (EBM) has emerged as a powerful problem-oriented practice of medicine that seeks to improve clinical practice by understanding the quality of clinical evidence and ensuring that only "best evidence from clinical research" is used in the management of individual patients EBM has particularly contributed to our understanding of the meaning of effects of treatments as reported in the literature, and often expressed this in terms of useful therapeutic summary measures However, from the point of view of CDA, EBM therapeutic summary measures are utilities and alone cannot be used in medical decision making The tenets of CDA hold that the preferred management strategy is the one with the highest expected value and not the one the highest utility value Determination of optimal treatment strategy within framework of two or more competing treatment strategies is traditional realm of CDA That is, decision analysis based medical decision making should be pre-requisite for any use of EBM and that effective application of EBM process requires its integration into analytic model of CDA (J Health Serv Res Policy 1996,1 104) However, no method has been developed to enable direct integration of EBM measures of therapeutic efficacy within CDA We showed that it is possible to link all EBM measures of therapeutic efficacy to CDA This direct link is especially clearly expressed in the following equation, particularly useful when competing treatment alternatives are treatment vs no-treatment

$$pt = R*NNT$$

where R is the risk (harms) of the treatment, NNT (number of patients who need to be treated to achieve one additional favorable outcome), and pt is the treatment threshold probability at which physician is indifferent between treating and not treating his/her patient If the probability that patient has the disease is lower than pt treatment is not indicated, if the probability of the disease is greater than pt treatment should be given This probability can sometimes be difficult to estimate in clinical practice Therefore, of equal importance is to know how low NNT should be before treatment is worth administering As our equation shows the treatment should only be considered if NNT<1/treatment risk We also show that other therapeutic summary measures used within EBM movement (EBM 1998,3 33) can be integrated within CDA framework and in other clinical situations such as comparing two different treatments or use of diagnostic tests

BIAS IN PREFERENCES WITH VIOLATION OF PROCEDURE INVARIANCE

JR Treadwell and LA Lenert, Stanford University Medical Center, Stanford, CA

In studies of health preferences, utilities for hypothetical health states cannot be always be successfully measured One marker for unsuccessful measurement is violation of "procedure invariance", when the preference ranking of two health states varies across assessment procedures For example, a person might assign a higher rating to health state A than to health state B when using the visual analog scale, but then have a lower indifference probability for A than B using the standard gamble No monotonic mapping function between methods can explain this pattern Using preference values based on unsuccessful measurement may result in misinterpretation of patients' attitudes about health We sought to determine whether people who violated procedure invariance had significantly different preferences than people who satisfied it

We performed secondary analyses of three completed studies that used the same two assessment procedures visual analog scale and standard gamble We identified participants who violated procedure invariance and compared the mean utilities of violators and satisfiers Participants who rated the states as equally desirable by the two methods were excluded from analyses

Rates of violation of procedure invariance ranged from 19% to 39% Violation of procedure invariance was not associated with age, education level, race, or gender in any of the three studies In all three studies, it was associated with bias in standard gamble utilities for some but not all health states Less bias was observed in visual analog scale ratings Subjects who violated procedure invariance appeared less able to discriminate between health states For these subjects, utility assessments appeared to be randomly distributed about a mean determined by the specific elicitation procedures used in each experiment

Violation of procedure invariance across scaling methods is associated with bias in obtained preference weights Researchers should test for procedure invariance and consider excluding violators from population estimates of utilities

PRELIMINARY VALIDATION OF THE SATISFACTION WITH DECISION SCALE WITH DEPRESSED PRIMARY CARE PATIENTS

CE Wills, M Holmes-Rovner Michigan State University, East Lansing, MI, College of Nursing, College of Human Medicine

A key aim of this study was to assess the psychometric properties of several self-report measures of patient decision making, including the recently-developed Satisfaction With Decision (SWD) scale (Holmes-Rovner et al , 1996) Depressed primary care patients (N = 97) who had recently made a new decision about using antidepressant medication completed a self-report mail survey The SWD scale had high internal consistency reliability (α = 85) Evidence for construct validity was found in a hypothesized pattern of statistically significant relationships that occurred between SWD and other measures Specifically, SWD was positively associated with overall health (SF20), perceived antidepressant medication efficacy, knowledge about depression and depression treatment, perceived physician participatory decision making (PDM) style, and satisfaction with the primary care setting and provider, but not desired amount of decision participation SWD was inversely associated with decisional conflict (DCS) and severity of body pain (SF20), but not depression score (CESD) A sub-sample of participants (N = 11) who discontinued medication soon after starting it reported significantly lower satisfaction with their medication use decision compared to those who started and continued, t (82) = - 4 57, p = 000 SWD is associated with patient education, decision participation, and depression treatment issues that are potential areas for interventions to improve primary care depression treatment Further research is needed to develop and test decision support interventions for these areas for depressed primary care populations

IMPACT OF PROPHYLACTIC SURGERY ON LIFE EXPECTANCY FOR WOMEN WITH BREAST CANCER AND BRCA1/2 MUTATIONS

D Schrag, KM Kuntz, JE Garber, JC Weeks Center for Outcomes and Policy Research, Dana-Farber Cancer Institute, Harvard School of Public Health, Boston, MA

Objective: To determine the survival benefit of bilateral mastectomy (BM) and prophylactic oophorectomy (PO) for BRCA1/2 carriers with early stage breast cancer (BC)

Methods: We constructed a Markov decision analysis comparing gains in undiscounted life expectancy (LE) between prophylactic surgical strategies and breast conserving surgery for women with early stage BC BC specific survival was modeled based on data from the International BC Trialists Group showing 10 year BC specific mortality of 26% and 12% for lymph node positive and negative disease respectively Other estimates from the literature included a 20% contralateral BCA risk at 10 years, (all assumed to be early stage), an effectiveness of BM at decreasing second BCs of 90% and, a lifetime probability of developing ovarian cancer (OVC) of 20% PO was assumed to decrease OVC risk by 50% and had no impact on BC prognosis

Results: Incremental Gains in Life Expectancy Relative to Breast Conserving Surgery 30 year old BRCA1/2 Mutation Carriers with Newly Diagnosed BC

Clinical Strategy		*Node Negative BC*	*Node Positive BC*
Breast Conserving Surgery	(Total LE)	34 7 years	25 7 years
Prophylactic Oophorectomy	(ΔLE)	0 5 years	0 4 years
Bilateral Mastectomy	(ΔLE)	1 7 years	1 1 years
Both Prophylactic Surgeries	(ΔLE)	2 4 years	1 6 years

Results are sensitive to 1) age, such that 50 year old women gain less than a year of life from both surgeries and 2) timing of surgery in relation to primary BCA diagnosis such that 5 year survivors can anticipate greater LE gains from prophylactic surgery than newly diagnosed patients

Conclusion: Young women with early stage BCA and germline BRCA1/2 mutations may gain 1 1-1 7 years of life expectancy from immediate BM rather than breast conserving surgery, however, the life expectancy gains for women over 50 and from PO are less than a year

META-ANALYSIS AND COST EFFECTIVENESS ANALYSIS OF ANTISEPTIC-TREATED CENTRAL VENOUS CATHETERS

DL Veenstra, S Saint, S Saha, T Lumley, SD Sullivan University of Washington, Seattle, WA and University of Michigan Ann Arbor, MI

Purpose Antiseptic-treated central venous catheters are a novel technology for the prevention of catheter-related bloodstream infection (CR-BSI) However, only one of several published studies has shown a statistically significant reduction in CR-BSI with the use of antiseptic catheters, therefore, their cost-effectiveness is uncertain We thus assessed the efficacy of antiseptic catheters based on a quantitative review of available studies and evaluated their cost effectiveness using decision analysis

Methods Randomized clinical trials comparing antiseptic-treated to standard catheters were included in the meta-analysis CR-BSI confirmed by catheter culture was the primary outcome for analysis Summary statistics were calculated using Mantel-Haenszel methods under a fixed-effects model A decision tree was developed to model the potential benefits and risks of using antiseptic catheters Probabilities for the baseline rate of CR-BSI and catheter efficacy were derived from the meta-analysis The probability of death due to CR-BSI was taken from the literature, the probability of hypersensitivity to antiseptic catheters was derived from recent FDA reports Costs were obtained from the literature and the UW Medical Center

Results The meta-analysis summary odds ratio for CR-BSI was 0 56 (95% CI, 0 37-0 84), indicating a significant decrease in CR-BSI associated with antiseptic catheters In the base-case cost effectiveness analysis, the use of an antiseptic catheter was the dominant strategy, resulting in savings per catheter used of approximately $150 and a decrease in the probability of death of 0 004 compared to a standard catheter In a sensitivity analysis with all parameters set to favor use of the standard catheter, using an antiseptic catheter resulted in slightly higher cost ($2) and a decrease in the probability of death of 0 0007, or about $3,000 per death avoided

Conclusion Antiseptic-impregnated central venous catheters are effective in reducing the incidence CR-BSI and are either cost-savings or cost-effective under a wide range of assumptions Our meta-analysis also reconciles the lack of a significant reduction in the incidence of CR-BSI found in the majority of previously reported trials

POWER ASSESSMENTS AND CONFIDENCE INTERVALS FOR COST-EFFECTIVENESS ANALYSES

Joseph C Gardiner, Marianne Huebner, James Jetton Departments of Epidemiology and Statistics and Probability, Michigan State University, East Lansing, Michigan

Economic evaluations of competing health care programs involve estimation of their relative costs and health benefits The cost-effectiveness ratio (CER) of a test intervention or program is defined as the ratio of its incremental cost, over its best available alternative or referent intervention, to the incremental benefit that it produces Assessing the precision of a computed CER is an important aspect of a cost-effectiveness analysis Confidence intervals for CERs provide information on the uncertainty in their point estimates We show the existence of a finite interval under the Fieller method when there is a statistically significant difference in the effectiveness of the two interventions being compared When this difference is not significant the method might yield an unbounded confidence interval A finite confidence interval is guaranteed when the differences in both cost and effectiveness are significant We first assess the sample size required to obtain a confidence interval for the CER of prescribed width Using the correspondence between confidence intervals and hypotheses tests, we derive formulae for statistical power and sample size assessments for tests of hypotheses of the CER The number of patients that should be entered into a clinical trial to have adequate power to identify statistically significant differences between two treatments is compared with the number required to demonstrate cost-effectiveness of the test treatment An expression for the power of the joint test of effectiveness and cost-effectiveness indicates that the sample size requirements for demonstrating cost-effectiveness might be considerably higher than that needed to show effectiveness alone The implications of these results for cost-effectiveness studies are illustrated through numerical examples

MODELING DECISIONS IN THE REPLACEMENT OF RISKY ARTIFICIAL HEART VALVES

EW Steyerberg, M Kallewaard, Y van der Graaf, LA van Herwerden, JDF Habbema Department of Public Health, Erasmus University Rotterdam, Julius Center for Patient Oriented Research, University of Utrecht, Department of Cardio-thoracic surgery, University Hospital Rotterdam - Dijkzigt, The Netherlands

Purpose: Bjork-Shiley convexo concave (BScc) mechanical heart valves have a risk of outlet strut fracture One might consider prophylactic re-replacement as a preventive measure to avert the disastrous consequences of fracture Three decision analyses formally evaluated the life-expectancy with or without valve replacement We compared assumptions and modeling choices in these studies

Methods: We compared the estimation of survival (without strut fractures), strut fracture risk, and surgical risk in three decision analyses (Birkmeyer, Lancet 1992, van der Meulen, Circulation 1993, Supervisory Panel, 1997) Alternative approaches were evaluated on recent follow-up data from 2263 patients with a BScc valve implanted between 1979 and 1985 in the Netherlands, where 50 fractures occurred and 883 patients died after 1 year of follow-up (excluding strut fractures)

Results: Survival was estimated with an additive excess risk model in two decision analyses and a Poisson regression model in another The assumption of a constant excess risk was not supported by our data, where an increase in risk during follow-up was observed This assumption led to overestimation of the life-expectancy of patients currently alive with a BScc valve The Poisson model could adequately estimate the mortality risk by distinguishing the age of the patient from the year of follow-up

Strut fracture risk was estimated with a set of 'classical' risk factors in all analyses (opening angle, size, position), while a recent analysis also included age and production characteristics Incorporation of the latter characteristics was supported by our data and led to risk estimates with better discriminative ability

Surgical risk estimates were either based on the primary implantation data, on a large series of reoperations, or on a limited number of BScc replacements The resulting risk estimates varied substantially, but we had no suitable data available to test the adequacy of the estimates

Conclusions: Considerable variation was noted in essential parameters of previous decision analyses, and some assumptions could not be supported by recent data (e g a constant additive excess mortality risk) This study supports the need of in-depth statistical modeling of individual patient data to improve the estimates of important parameters in a decision analysis

ANGULAR TRANSFORMATION CLARIFIES COMPARISONS OF ADJUSTED COST-EFFECTIVENESS RATIOS

JJ Anderson, S Loveland, DR Miller, Center for Health Quality Outcomes and Economic Research VAMC Bedford MA and Boston University School of Public Health

Bootstrap distributions of marginal cost effectiveness ratios (MCERs) are usually presented in the Cost-Effectiveness (CE) plane A single dimension presentation based on the variance-stabilizing arctan transformation is simpler and facilitates MCER comparisons (Heyse 1997) Evaluations and comparisons of bootstrap distributions are clarified

Adjustment of an MCER may be needed when the assignment to groups is non-random as in an observational study We present unadjusted and adjusted MCERs for specialty versus non-specialty care of 204 Veterans Health Study patients with self-reported knee osteoarthritis and/or chronic low back pain in a VA ambulatory setting Effectiveness is measured as health status improvement, based on change in the SF-36 derived physical component score (PCS), during the time period over which utilization occurs and is costed Patients in these care groups differ in baseline disease burden, pain scores and PCS, and also in age and education so the unadjusted MCER is potentially biased

We standardized the MCER using the ratio of ranges of C and E differences rather than standard deviations, prior to arctan transformation The non-parametric adjustment based on stratification on 2 major imbalance factors, and using the same factors for adjustments of both costs and effectiveness is the preferred method with reduced bias and little loss of precision By comparison, an MCER defined using differences between care group least squares means from risk-adjustment models for each of C and E has reduced bias but is less precise (has wider CI) and there is the potential for overfitting

On the angular scale, MCER medians and 95% confidence intervals are simply described and the relative biases and comparative precision of differently adjusted estimates can be easily appreciated and interpreted

DIFFUSION AND USE OF THE OTTAWA ANKLE RULES IN FIVE COUNTRIES

ID Graham, IG Stiell, A Laupacis, JM Tetroe, LM McAuley, P Durieux, N Simon, M Clancy, M Howell, JI Emparanza, JR Agingaga Clinical Epidemiology Unit, Ottawa Hospital, Canada, Service de Sante publique, Hopital Cochin, St Jacques, France, Département Accui Urgences, Hôpital Intercommunal de Poisey, St Germain, France, Accident and Emergency Department, Bristol Royal Infirmary, UK, Accident and Emergency Department, Queen Alexandra Hospital, Portsmouth, UK, Clinical Epidemiology Unit, Hospital Aranzazu, San Sabastian, Spain, Emergency Service Hosptal de Guipuzcoa, San Sebastian, Spain

Objectives To compare emergency physicians' awareness and use of the Ottawa Ankle Rules (OAR), a highly sensitive decision rule for the use of radiography, to identify factors associated with he use of the OAR Methods A cross-sectional mail survey of random samples of 500 members each of the Canadian, American, UK, and Spanish national association of emergency physicians, and members of the equivalent association in France

	Results				
Question (%)	Canada	US	UK	France	Spain
• Aware of OAR	99	96	91	70	21
• Currently use OAR by those aware of them	98	79	94	78	83
• Would consider using OAR by those currently not using or unaware of them	75	73	75	79	95

A preliminary forward stepwise logistic regression analysis identified the following three factors associated with using the OAR always or most of the time working in Canada or the UK, possessing positive attitudes to decision rules in general, and younger age Factors not found to be related to OAR use were sex, hours worked per week in the emergency department, volume of emergency visits, stress from uncertainty, and risk aversion

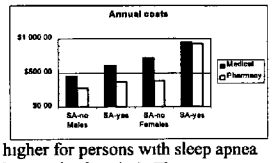
Conclusions Awareness of the OAR was greater in English speaking countries This is likely related to the fact that the OAR have been published predominantly in English language journals Of physicians aware of the rule, greater than 78% in each country report using it some or most of the time As expected, country, attitudes toward decision rules in general, and age were related to frequent use of the rule However, comfort with risk and degree of stress from uncertainty in medical practice were not predictive of use of the rules

DIRECT OUTPATIENT MEDICAL COSTS OF SLEEP APNEA A POPULATION STUDY

DG Fryback, T Young, M Palta, L Finn, P Meek University of Wisconsin, Madison, WI

Background. Based on high prevalence and direct effects of sleep apnea (SA) (e g , accidental trauma) and conditions it may contribute to (e.g , hypertension), treatment of SA may save substantial associated medical costs Our study empirically estimates the direct outpatient medical and pharmaceutical costs associated with undiagnosed sleep apnea in a population.

Data and Methods. A population-based cohort of working adults was evaluated with overnight stay. 3 years of outpatient HMO insurance claims records were used for cohort members (7204 person-calendar quarters (pcq) of medical service and 6164 pcq of pharmacy data). 26,056 instances of medical service provision were priced consistently using the 1996 Medicare fee schedule. Prescriptions filled (n=13,322) were priced with the 1996 *Red Book* and coded by therapeutic class. SA was indicated by apnea-hypopnea index (AHI) from polysomnography and categorized "SA-no"=AHI<5, and "SA-yes"=AHI≥5. Skewed distributions required use of survival analyses, controlling for age and body-mass index, on subsets of the data stratified by sex to test for SA effects on average annual costs and counts of service utilization per person.



Results. Total costs (excluding direct SA treatment) are higher for persons with sleep apnea (see fig.; all differences significant except medical cost difference for females) The cost differentials are not due to any clear source, but seem to stem from accumulated effects across many subtypes of services and costs SA-yes is associated with more referrals and tests than SA-no. Male SA-yes have fewer psychiatric referrals than SA-no; for women this is reversed

Conclusion. Undiagnosed SA is associated with 33% increase in total outpatient medical and drug costs in working adult men and 72% increase (mostly in cost of pharmaceuticals) in working adult women. Treatment of SA may in fact save substantial associated medical costs

THE IMPACT OF INTEGRATED CARE MANAGEMENT ON MEDICAL AND DISABILITY UTILIZATION AND EXPENDITURES

A Comparato, RL Dunn, RJ Ozminkowski, J Schmidt, J Niwa, G Barton The MEDSTAT Group, San Francisco, CA and Ann Arbor, MI

Mental and behavioral health conditions among employees may cause serious medical problems, substantially reduce productivity, and lead to lengthy absences from work These conditions often lead to substantial expenditures for corporate medical and disability programs, in addition to the health and productivity-related problems they cause We studied the impact of a new cooperative program that was designed to jointly manage the medical and disability program services used by employees on short-term disability for mental health problems The program was provided to a large employer in the western U S Its goal was to reduce the duration and recidivism of short-term disabilities due to mental health problems by better coordinating medical and disability program services

A multivariate proportional hazard analysis was conducted to estimate the impact of program participation on the duration of short-term disability program use for 851 program participants and 498 similar program non-participants from June, 1997 – February, 1998

The results from the hazard analysis were used to estimate differences in mean disability duration for program participants and non-participants, adjusted for confounding variables These means were multiplied by a wage-based conversion factor to conservatively estimate the savings associated with the program These dollar savings were then added to dollar savings estimated from a multiple regression analysis of medical expenditures to estimate total program savings These savings were compared to the cost of providing the program to calculate the return on investment in the program

The results of the analysis showed that the adjusted mean disability duration was about 17 1 days less for program participants compared to non-participants This resulted in a savings of about $2,009 per disability case This was offset to a minor extent by an estimated $221 per case increase in medical expenditures Comparing the net savings to the cost of providing the program, we estimated the annualized return on investment to be about $3 5 million, or about $12 96 saved per $1 spent on the program These results suggest that substantial savings may be obtained and that productivity can be increased by the combined, successful management of medical and disability program services

IMPROVING RESIDENTS' COMPLIANCE WITH PRACTICE GUIDELINES RESULTS FROM THE VA COMPUTERIZED-REMINDER MULTI-SITE CLINICAL TRIAL

JG Demakis, C Beauchamp, WL Cull, R Denwood, WG Henderson Hines VA Cooperative Studies Program Coordinating Center, Hines, IL and Durham VAMC, Durham, NC

Purpose To examine whether a computerized-reminder system operating in multiple VA ambulatory care clinics improves practice-guideline compliance for patients seen by resident physicians

Methods Resident physicians (N = 275) at 12 VAMCs were randomly assigned in firms or half-day clinic blocks to either the reminder or the control group Across the 16-month study period, the residents cared for 12,989 patients for whom at least one of the 13 study guidelines was applicable Compliance with the guidelines was tracked following each visit using combined information from encounter forms completed by residents at the time of the visit and from each hospital's computer system

Results Measuring compliance as the number of patients receiving proper care out of all patients with one of the studied health conditions, the reminder group ($\bar{x}$ = 58 8%) had significantly higher rates of compliance than did the control group ($\bar{x}$ = 53 5%) for all guidelines combined and for five of the 13 guidelines examined individually Measuring compliance as the number of visits where the resident provided the suggested care out of all visits where care was indicated, the reminder group ($\bar{x}$ = 17 9%) also had significantly higher rates of compliance than did the control group ($\bar{x}$ = 12 2%) for all guidelines combined and for nine of the 13 guidelines examined individually The benefit of reminders, however, was found to significantly decline throughout the course of the study

Conclusions A reminder system installed at multiple sites can significantly improve residents' compliance to multiple practice guidelines The benefits of the system, however, may deteriorate across time Future research needs to explore methods to sustain better the benefits of reminders

EFFECT OF RACE, GENDER AND AGE ON PHYSICIAN RECOMMENDATIONS FOR CARDIAC CATHETERIZATION

KA Schulman, JA Berlin, W Harless, J Kerner, S Sistrunk, BJ Gersh, R Dube, CK Taleghani, W Ayers, JE Burke, S Williams, JM Eisenberg, JJ Escarse Georgetown University Medical Center, Washington, DC, University of Pennsylvania, Philadelphia, PA and RAND Health Program, Santa Monica, CA

Epidemiologic studies have reported race and gender differences in access to cardiovascular procedures However, the role of physician decision-making leading to this finding has not been determined

We developed a multimedia computer survey instrument to assess physician management of patients presenting for evaluation with chest pain In a full factorial design, physicians were randomized to view a video one of eight patient actors varying by race (black/white), gender (male/female) and age (55 years/70 years) presenting with three different types of chest pain (definite/possible/non anginal pain), and two levels of coronary risk factors (high/low) After viewing the patient actor, physicians classified the type of chest pain and prior probability of disease Physicians then reviewed 1 of 3 possible thallium stress test results and reported posterior probability of disease as well as recommendation for referral for cardiac catheterization The survey instrument was administered to 720 physicians at two national meetings of primary care professional societies Multivariate logistic regression analysis was used to assess the effects of race and gender, controlling for other experimental factors

Prior probability of coronary disease estimates varied significantly by gender, age and type of chest pain Physicians provided lower mean estimates for women (x=64 1% vs 69 2% men, p<0 001), older patients (x=69 5% vs 63 8% 55 year old, p<0 001) and patients with nonanginal pain (x=58 3% vs 64 4% possible angina vs 77 1% definite angina, p=0 001) Race, gender and age were significant predictors of referral for cardiac catheterization Women (OR 0 60, p=0 02), blacks (OR 0 60, p=0 02) and older patients (70 years old, OR 0 62, p=0 04) were less likely to be referred than males, whites and younger patients Analysis of race-gender interactions found that black females (OR 0 40, p=0 004) were less likely to be referred than white males Further analysis found 70 year old black females were least likely (OR 0 2, p=0 002) to be referred for cardiac catheterization Findings were unchanged with sensitivity analysis

Our findings indicate patient race, gender and age independently influence physician management of chest pain This study suggests that physician decision-making is an important factor in explaining differences by patient race and gender in treatment of coronary artery disease

PREFERENCES FOR PROSTATE CANCER SCREENING HEALTH STATES AND THE ROLE OF REASSURANCE

SB Cantor, RJ Volk, AR Cass, SJ Spann
The University of Texas M D Anderson Cancer Center, Houston, TX, The University of Texas Medical Branch at Galveston, Galveston, TX and Baylor College of Medicine, Houston, TX

The purpose of this study was to examine the role of reassurance in determining men's preferences for outcomes of prostate cancer screening using the prostate-specific antigen (PSA) test and digital rectal exam (DRE)

As part of a larger study assessing couples' preferences for outcomes of prostate cancer treatment, we asked 90 asymptomatic male family medicine patients, 45 to 70 years of age, to rank possible outcomes of prostate cancer screening The three screening health states were as follows (A) no prostate cancer screening performed, (B) PSA test and DRE performed with normal results, and (C) PSA test and DRE performed with abnormal results, subsequent ultrasound with biopsy is normal Patients were forced to rank order the health states with no allowance for equivalent preferences

The table shows the number of men who preferred each of the possible screening health state rankings

Ranking	A-B-C	A-C-B	B-A-C	B-C-A	C-A-B	C-B-A
Count	3 (3 3%)	1 (1 1%)	8 (8 9%)	40 (44 4%)	1 (1 1%)	37 (41 1%)

Only 4 5% of the men considered screening health state A (no screening) to be the best, 85 5% considered health state A to be the worst For the men indicating a family history of prostate cancer, 11 of 18 (61 1%) considered health state C, with favorable biopsy result, the best health state Of the men indicating no family history of prostate cancer, 26 of 64 (40 6%) ranked health state C as best (P= 06)

There is variation in men's preferences for possible outcomes of prostate cancer screening A family history of prostate cancer appears to be related to preference for screening health states associated with greater reassurance While the benefits of early detection of prostate cancer in asymptomatic men are uncertain, our sample of primary care patients felt that normal results from prostate cancer screening and evaluation had reassurance value

THE COST EFFECTIVENESS OF PRENATAL DIAGNOSIS: DOES THE 35-YEAR-OLD THRESHOLD MAKE SENSE?

RA Harris, AE Washington, D Feeny, M Kuppermann, University of California, San Francisco; Centre for Health Economics and Policy Analysis, McMaster University, Hamilton, Ontario

Purpose: Current guidelines recommend offering chorionic villus sampling (CVS) or amniocentesis only to women aged 35 or older at the time of delivery. This threshold was chosen, in part, because it is the age that the procedures were found to be cost beneficial in 1970s-era analyses. With the benefit of recent randomized trials and a standard gamble utility assessment, we conducted a cost-effectiveness analysis, comparing CVS, amniocentesis, and no testing for women of all maternal ages

Methods: Cost-effectiveness decision analysis from the societal perspective, using data from the Canadian and European randomized trials of CVS and amniocentesis, large case registries of thousands of patients, and a standard gamble utility assessment on 72 women considering prenatal diagnosis A broad array of multiway sensitivity analyses on all probabilities and utilities was conducted.

Results: When taking the societal perspective and including the cost and utilities of only the mother, amniocentesis costs $5,000 per QALY gained for women aged 35 It remains cost effective for all other maternal ages, ranging from $10,000 per QALY gained at age 20 to cost savings for age ≥ 37 Amniocentesis is less costly and more effective than CVS at all maternal ages Wide variations in the utilities of miscarriage and Down's syndrome do not change the analytic results. However, the model is primarily sensitive to the preference younger women have for fetal chromosomal information: for women under 33, amniocentesis is only cost effective for those individuals who have a marked preference for knowing the chromosomal status of the fetus.

In contrast, the results appear substantially different when taking the societal perspective and including both the mother's costs and utilities, as well as the loss in life years to the fetus from miscarriage In this situation, prenatal diagnosis is both more expensive and less effective than is no testing for all women under 40 For women over 40, prenatal diagnosis is cost saving

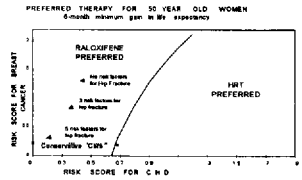
Conclusions: Amniocentesis appears cost effective regardless of maternal age when women are interested in fetal chromosomal information If the lost life years to the fetus from miscarriage are considered in the analysis, however, amniocentesis is only cost effective for women over 40 Whether the life years of the fetus should be counted is a political issue Nonetheless, in neither situation do cost effectiveness arguments support the 35-year-old age threshold for prenatal diagnosis

INDIVIDUALIZING THERAPY TO PREVENT POSTMENOPAUSAL OSTEOPOROSIS A COMPARISON OF HORMONE THERAPY, ALENDRONATE, AND RALOXIFENE

NF Col, SG Pauker, RJ Goldberg, MH Eckman, RK Orr, EM Ross, JB Wong New England Medical Center and Tufts University School of Medicine, Boston, MA

Alendronate and raloxifene were recently approved for the prevention of osteoporosis but data on their clinical efficacy are limited We developed a Markov simulation model to compare the effects of hormone replacement therapy (HRT), alendronate, and raloxifene on the incidence of hip fracture, coronary heart disease (CHD), breast cancer, and on life expectancy, in postmenopausal women with different risk profiles Treatment-induced changes in bone mineral density were used to predict hip fractures, and raloxifene's lipid effects were used to predict CHD Regression models linked individual risk factors to future disease risks We assumed that HRT reduced the risk of CHD but increased the risk of breast cancer, and that raloxifene decreased the risk of breast cancer

HRT, alendronate, and raloxifene have similar predicted efficacies in preventing hip fractures (estimated RR=0 57, 0 54, and 0 58, respectively), while HRT should be over 10-fold more effective than raloxifene in preventing CHD Women at low risk for hip fracture, CHD, and breast cancer did not benefit significantly from any treatment Among women at average risk for hip fracture, CHD, and breast cancer, HRT was the preferred treatment, raloxifene was preferred among such average women only if it could reduce the risk of breast cancer by at least 66% Women at high risk for CHD benefit most from HRT, women at high risk for breast cancer but low risk for CHD would benefit from raloxifene, if raloxifene lowers the risk of breast cancer Among women with multiple risk factors for hip fracture, alendronate should increase life expectancy modestly



Because of significant differences in the projected impact of these drugs on osteoporosis, CHD, and breast cancer, treatment choice depends on an individual woman's risk profile for these diseases

ESTIMATING HEALTH STATE UTILITY FROM A SECONDARY DATABASE

N Sengupta, MB Nichol Pharmaceutical Economics & Policy, University of Southern California, Los Angeles, CA

Assessing health status and utility through different measures has become increasingly important However, relatively little work has been done to relate one method to another The purpose of this study is to investigate the role of the SF-36 to predict Health Utility Index (HUI) score

Method Data were obtained from a longitudinal study of 6,000 randomly selected patients under managed care in Southern California To assess health status, SF-36(version 1 0), and Health Utility Index (MarkII) were used The first instrument was used at baseline, year1, and year2, however HUI was used only in the last cycle The SF-36 scores were used to predict year2 HUI score, with socioeconomic and disease variables as covariates An instrumental variable approach, as well as a reduced form approach, were used in estimation As most of the patients had a high utility value, a censored regression (Tobit) was also performed A 'Bounded Influence Estimation' procedure was also used to capture the effect of any extreme utility value

Results The SF-36 components and sociodemographic covariates accounted for 33%to 50% variation in HUI depending on the model used The magnitude and directions of parameter estimates of regressors were unaffected by the censored nature of the utility value. A split -sample analysis through Monte-Carlo simulation showed absence of any structural break in the model All but two domains of the SF-36 showed statistical significance at $p<0.05$ level Age and Income level showed statistical significance at same level in most of the methods. Chronic disease state of patients also showed statistically significant influence in estimating their health status and health utility scores

Conclusions. This research provides preliminary evidence that the SF-36 can be used to predict HUI, and thus utility scores Estimation procedures presented in this study appear to be quite robust Methodology presented here, will help researchers to cross-translate summary utility score from any kind of previous study using SF-36

RISK ATTITUDE AND TIME PREFERENCE OF ULCERATIVE COLITIS PATIENTS A WWW BASED STUDY

JR Treadwell, MR Bergen, RM Soetikno, LA Lenert, Stanford University Medical Center, Stanford, CA and Palo Alto Veterans' Affairs Medical Center, Palo Alto, CA

It is generally agreed that risk attitude and time preference are logically distinct concepts However, is the distinction of practical importance in medical decision making? The standard assumptions are risk aversion and positive time preference Both of these effects suggest a concave-downward shape of the utility function over life years Therefore, perhaps only one parameter is needed to capture both effects For example, in the reference-case cost-effectiveness analysis, there is no explicit accounting for risk aversion, but the use of a discount rate (r=0 03) might be viewed as a proxy for both risk aversion and positive time preference

We measured both risk attitude and time preference in a group of 200 self-identified patients with ulcerative colitis All patients completed a brief questionnaire on the World Wide Web We recruited 60% of them using email addresses recorded in an earlier study, the remaining 40% were recruited during their search for information on ulcerative colitis using the WWW Based on responses to gamble questions, each patient was classified as either risk averse, risk neutral, or risk seeking In addition, based on responses to time questions, each patient was classified as having either positive time preference, zero time preference, or negative time preference This yielded a 3x3 table in which cells along the main diagonal (e g , risk averse and positive time preference) require only one modeling paramenter, whereas cells off the main diagonal (e g , risk averse and negative time preference) require two parameters because the two effects work in opposite directions

A surprisingly high proportion of patients (54%) had preference patterns that fell off the main diagonal For these patients, it is necessary to have separate parameters for risk attitude and time preference because their choices are impossible to describe with only one parameter Many of them (60%) had negative time preference they preferred to have the most severe ulcerative colitis state sooner rather than later Of the patients who had negative time preference, 53% were even willing to undergo a longer period of time with severe symptoms in order to ensure that it would be finished sooner Preference patterns were largely unaffected by age, sex, educational level, self-rated health status, and time since diagnosis

Preferences regarding risk and time are too complicated to warrant the use of one overall proxy parameter Many people are adamant in their desire to expedite poor health because they believe they can cope better when younger At the same time, however, people are generally risk averse In order for the predictions of decision models to be relevant to patients, risk attitude and time preference must be treated separately

HEALTH RELATED QUALITY OF LIFE IN HIV POSITIVE PATIENTS – VARIATIONS IN OUTCOME ACROSS INSTRUMENTS

SA Call, JC Klapow, MS Saag, KE Stewart, RM Centor University of Alabama at Birmingham (UAB), Birmingham, AL

Purpose Previous studies of health status and health value measures in patients with HIV have shown inconsistencies between the two types of measures and across disease states For example, studies using time trade off (TTO) utilities have shown that values are fairly stable across disease states and that patient rated TTOs are higher than might be expected by society rated values of the same disease states These differences highlight the complexity of identifying appropriate utility values in HIV populations In the present study we employed varying measures of health related quality of life (HRQOL) in a cohort of HIV patients in order to identify convergent outcomes across measures

Methods Four HRQOL measurements were evaluated in a prospective clinical cohort of HIV positive patients recruited from the UAB HIV primary care clinic Demographic, laboratory, quality of life data, prior treatment and disease history were obtained at baseline and at three month intervals thereafter HRQOL instruments included the MOS-SF36, the EuroQol and the EuroQol visual analog scale (VAS) The mental and physical component scores (MCS and PCS) were calculated for the MOS-SF36

Results At time of analysis the cohort consisted of 132 patients 86% male, 64% Caucasian, 66% homosexual, baseline median CD4 count 156, baseline median viral load (VL) 63,465 Median follow up was nine months At entry, mean PCS = 43 7, mean MCS = 43 8, mean EuroQol score = 76, mean VAS score = 75 4 ANOVA analysis for differences in scores across CD4 count ranges found a consistently significant decrease in PCS and VAS scores with decreasing CD4 count at all points of follow up but not in MCS and EuroQol scores ANOVA for differences in scores with patients stratified by VL again showed a significant difference in PCS and VAS scores but no consistent significant difference in MCS or EuroQol

Conclusions As noted in previous studies, there was inconsistency across measures of HRQOL However, the VAS and PCS consistently discriminated across two separate disease state indices (CD4 count and VL) Because the PCS can only be considered a component of HRQOL, the VAS represents the most consistent HRQOL estimate in this cohort Our data and other data suggest that different instruments are likely to elicit very different values, and that the determination of appropriate utility estimates for a cost-utility analysis must be made with caution, particularly in this population

HOW "UTILITIES" ARE MEASURED MATTERS COMPARING VALUES ASSIGNED BY STANDARD GAMBLE, TIME TRADE OFF, AND RATING SCALE APPROACHES

MK Monrose, RB Deber, University of Toronto, Toronto, ON, Canada

How should health outcomes be measured? Although standard gamble (SG) is considered the "gold standard," many studies instead employ time trade off (TTO) or rating scales (RS) SG and TTO measure the value of a health outcome by forcing respondents to "trade off" against perfect health a risk of death (SG) or a loss of life years (TTO) RS approaches, although less well founded axiomatically, evaluate health states without tradeoffs Deber and Goel (1990) suggested that the trade-off based approaches confound assumptions about risk with evaluations of outcomes, and would result in a shift towards high utility values for health states which individuals would be unwilling to risk death to avoid

Because systematic measurement-related variation in utilities assigned would skew economic evaluations, we accordingly reviewed published articles in which the same respondents evaluated the "utility" of the same health state using multiple approaches, of which at least one was trade-off based (SG or TTO) and the other RS-based With the kind cooperation of Drs Bass, Bosch, Deber, deVries, Kattan, Lenert, Stigglebout, Sybolt, and Tsevat, we also constructed a merged data set of 3905 observations from 9 studies, translating all "utilities" to a 0-to-1 scale, and analyzed them using SPSS-PC

The results of paired comparisons (where "Difference" subtracts the 2nd from the 1st metric) were

Comparison	N pairs	Correlation	Difference	Confidence Interval of Difference
RS - SG	599	010 (p= 809)	-0 230	- 254 to - 206
RS - TTO	2880	403 (p= 000)	-0 134	- 145 to - 122
SG - TTO	232	300 (p= 000)	0 042	008 to 076

These results suggest that measures purporting to measure the same health state correlate poorly, even when the same health states are measured by the same respondents within the same study SG generates higher values for health states than does RS, with the mean difference of almost 1/4 of the available 0 to 1 scale (TTO is in between) These findings held across all studies, and a variety of health states

Conclusions as to the value of an intervention are thus highly susceptible to the measurement approach being used In the long run, these results suggest that greater attention must be paid to the meaning and stability of the concept - "value of a health outcome"

PROJECTING THE LONG-TERM CONSEQUENCES OF VARICELLA VACCINATION IN INFANCY

M Rothberg, ML Bennish and JB Wong Divisions of Clinical Decision Making, Informatics and Telemedicine, and Geographic Medicine and Infectious Diseases, Department of Medicine, New England Medical Center, Tufts University School of Medicine, Boston, MA

Introduction Although varicella vaccine is safe and effective in children, some medical providers and parents have chosen not to vaccinate because of concerns that waning efficacy over time may put vaccinated children at risk of chicken pox in adulthood when the disease is more severe

Objective To estimate the lifetime effects of vaccinating healthy children against varicella in infancy

Design A Markov model comparing 1) vaccination for children 18-months-old (**VaxAll**), 2) waiting until age 10 years (**Vax10**), and 3) no vaccination (**NoVax**) The model considers published age-specific incidence, morbidity and mortality rates for varicella and zoster, changing vaccine efficacy over time, and shifting disease burden from childhood to adulthood as vaccine compliance increases

Results Compared to **NoVax**, **Vax10** and **VaxAll** should increase QALE by 0 4 and 4 5 days respectively 48% of the increase in QALE was from preventing zoster, 42% from preventing uncomplicated varicella, and <10% from preventing death Assuming an inception cohort of 3 7 million infants, the projected results compared to the **NoVax** strategy are shown below

	Varicella Outcomes Averted			Zoster Outcomes Averted		
Strategy	Cases	Hospitalizations	Deaths	Cases	Hospitalizations	Deaths
Vax10	261,000	1,480	19	37,000	1,850	7
VaxAll	2,507,000	4,810	54	549,000	29,230	96

With waning vaccine efficacy, chicken pox shifts to older ages when disease is more severe This was offset by the milder nature of breakthrough disease among previously vaccinated individuals In sensitivity analysis, **NoVax** was always inferior to **Vax10** **VaxAll** was always favored over **Vax10** unless the vaccine both 1) failed to decrease varicella severity and 2) resulted in a 23-fold increase in the annual risk of adult infection (increasing the number of adult cases from 230,000 to 2 3 million)

Conclusions Even if efficacy wanes over time, universal vaccination against varicella at 18 months offers significant clinical benefit over waiting until age 10 to vaccinate or not vaccinating at all

HOW DO WOMEN VALUE DOWN-SYNDROME-AFFECTED BIRTH AND MISCARRIAGE? THE THIRTY-FIVE-YEAR-OLD QUESTION

M Kuppermann, RF Nease, LA Learman, E Gates, SF Posner, AE Washington, University of California, San Francisco, CA, Washington University, St. Louis, MO

Purpose: Prenatal diagnostic testing for chromosomal disorders has been available for nearly 35 years Current guidelines reserve such testing (amniocentesis or chorionic villus sampling) for women whose risk of carrying a fetus affected by Down syndrome (DS) is at least as great as the miscarriage risk associated with these procedures. This includes women who will be 35 or older at the time of delivery, or whose risk has been determined via serum screening to be higher than that of an unscreened 35-year-old This policy implicitly assumes that these two outcomes -- experiencing a procedure-related miscarriage and giving birth to a child affected by a chromosomal disorder such as Down syndrome -- are equally burdensome and explicitly ignores the preferences of individual women We sought to determine whether the implicit assumption is correct.

Methods: We assessed time tradeoff utilities for 11 potential outcomes of prenatal diagnostic testing decisions among 383 pregnant women aged 15 to 45 (mean = 33) The cohort reflected a broad racial, ethnic and sociodemographic mix We obtained information regarding subsequent prenatal testing decisions from chart review

Results: On average, these women assigned higher utilities to undergoing testing and then experiencing a miscarriage (mean=0 84, median=0 86) than to foregoing testing and giving birth to a child affected by DS (mean=0 75; median=0 76, p<0 0005 for both statistics). Substantial variation was observed in the utilities for these outcomes (miscarriage SD 0 25, interquartile range. 0 69-0.96, DS SD. 0.25, interquartile range 0 59-0 89) Among women aged 35 and older, higher utility for procedure-related miscarriage without a future birth was a predictor of undergoing prenatal diagnostic testing (OR· 7 68, 95% CI 1.1-56), as was an inclination to have an abortion if the fetus were found to be affected by DS (OR=3.59, 95% CI=1 67-7 73).

Conclusions: Our research reveals two important findings First, pregnant women do not value equally DS-affected birth and procedure-related miscarriage. Second, there is substantial variation in these utilities. These data call into question the "equal risk" threshold for prenatal diagnosis Future guidelines for prenatal diagnosis should incorporate individual preferences, in addition to risk

SUBSTITUTE DECISION MAKING HOW BEST TO ASSESS THE SOURCES OF DECISION UNCERTAINTY?

DN Guerriere, HA Llewellyn-Thomas Clinical Epidemiology Unit, Sunnybrook Health Science Centre, Toronto, Ontario, Canada

The Factors Contributing to Uncertainty (FCTU) Subscale quantifies one dimension of O'Connor's (1995) construct of decisional conflict It assesses 4 postulated factors lack of information, perceived pressure, lack of support, and unclear values Although the FCTU Subscale has been evaluated with individuals making their own treatment decisions, its appropriateness with substitute decision makers, who are also likely to experience decisional conflict, was unknown This study used hybrid quantitative / qualitative data collection and analytic strategies to explore this issue in a pediatric decision context

47 mothers deciding about gastrostomy tube insertion for their children completed the FCTU Subscale, and provided verbal reports about their experience with each of the 4 factors Next, "contributory scores" (indicating the extent to which each factor contributes to overall uncertainty) were derived, and the qualitative data were content analysed Then, for each factor, relatively high, moderate, and low sub-groups were identified using the contributory scores, and the associated verbal reports were examined for across-sub-group differences

There were across-sub-group differences in verbal reports about all 4 factors Mothers with relatively high contributory scores reported that, during their decision making they would have preferred more information (χ^2 = 6 990, p = 0 0082), they experienced greater social pressure (χ^2 = 8 377, p = 0 0038), they received less social support (χ^2 = 5 573, p = 0 0182), and they perceived more short-term losses / fewer long-term gains associated with the decision (χ^2 = 5 76, p = 0 0164 / χ^2 = 3 85, p = 0 0499 , respectively)

This evidence of consistency between mothers' FCTU Subscale scores and their verbal reports implies that the subscale could be used to identify which aspects of a decision situation are generating the highest levels of uncertainty in a substitute decision maker, and which would, in turn, most lend themselves to clinical intervention The results also indicate that this hybrid quantitative / qualitative approach is a useful, flexible strategy for further exploratory research in decisional uncertainty in other, non-pediatric, substitute decision making contexts

HEALTH-RELATED QUALITY OF LIFE (HRQOL) IN MEN WITH LOCALIZED PROSTATE CANCER

DA Katz, D Fryback, CA McHorney, MM Schapira University of Wisconsin Medical School, Madison, WI and Medical College of Wisconsin, Milwaukee, WI

Objective Prior decision analysis has shown that choice of therapy for localized prostate cancer (PCa) is sensitive to the disutility of living with untreated cancer (Kattan, 1996) As HRQOL influences both treatment and screening decisions, this study examines the functional status and utility for current health in men with and without prostate cancer who had recently undergone prostate biopsy for suspected PCa

Methods 120 men with untreated biopsy-proven, clinically localized PCa (Bx+) and 23 men without PCa on prostate biopsy (Bx-) were interviewed in-person (mean 56 and 76 days after dx, resp) at two academic medical centers PSA levels were elevated in Bx+ and Bx- men (mean 7 5 vs 6 5, resp) We compared SF-36 and time tradeoff (TTO) scores of Bx+ and Bx- men to a population sample of 343 age-matched men without known prostate disease who were interviewed in-person (all measures were scaled 0-100) We adjusted for demographic and comorbidity variables using ordinary least squares regression

Results Bx+ men had lower average scores for mental health (MH), role-emotional (RE), role-physical (RP), social (SOC), and physical function (PF), relative to population controls (p≤0 05*) Bx- men also showed reduced scores for RP and MH There was no significant decrement in adjusted mean utility (sd) for current health in Bx+ and Bx- men, relative to population controls 88 (10), 90 (10), vs 85 (8), resp

Health Domain	Average deviation (SE) from population sample Bx+ (n=120)	Bx- (n=23)
Physical function	-5 1 (2 6)*	-2 6 (4 7)
Role-physical (RP)	-15 3 (3 7)*	-23 5 (6 8)*
Vitality	-3 0 (2 3)	-6 1 (4 3)
Social (SOC)	-7 5 (2 0)*	-3 8 (3 7)
Role-emotional (RE)	-14 5 (3 0)*	-7 7 (5 4)
Mental health (MH)	-7 1 (1 6)*	-6 8 (3 1)*

Conclusion In this cross-sectional study, the detection of prostate cancer on biopsy had a significant impact on functional status, esp role function and mental health Even men without PCa on biopsy (but with elevated PSA) showed measurable declines in these areas PCa control programs should account for the effects of early detection on the functional status of men undergoing screening and biopsy

UNDERSTANDING HIV PATIENTS DECISIONS ABOUT EMERGENCY DEPARTMENT (ED) USE

AL Gifford, R Collins, D Timberlake, S Bozzette, M Shapiro, M Schuster, D Kanouse, VA San Diego Healthcare System, University of California San Diego, CA, RAND, Santa Monica, CA

Purpose: Patients often choose the ED for care of both non-urgent and urgent HIV symptoms, however ED care is expensive, and quality may suffer due to poor integration with primary care To better understand patients' choices of venue for HIV care, we interviewed a probability sample of U S HIV patients about how they would seek care in response to several clinical scenarios **Methods:** A nationally representative probability sample of all U S HIV+ adults in care underwent detailed structured interviews Three brief clinical scenarios describing headache, respiratory, and other symptoms were asked of each subject Different scenarios were asked of early stage (CD4≥200, no AIDS) and late stage (CD4<200 or AIDS) patients Scenario response options included "I would go to the emergency room" and "I would go to the doctor's office or speak to the doctor the same day" Sociodemographic (SD), access to care, health status, and knowledge/attitude/belief (KAB) variables were also assessed Weights were used to adjust for the sampling design, and non-response, linearization methods corrected for weights and multistage sample design Proportional logit models to identify independent predictors of propensity to seek ED care were estimated separately for early and late stage patients A hypothesis-driven regression approach was used to evaluate the relative contributions of SD, access, health, and KAB variables to propensity to chose ED care (among those who would seek immediate care) Since unmeasured differences in the organization of care may exist between care sites, dummy variables for each subject's sampling site were entered into the final models **Results:** Data were collected from 1245 early stage and 1612 late stage subjects representing 99,066 and 131,783 patients respectively Coefficients for significant variables in final models are shown Positive coefficients indicate more propensity to seek ED care **Conclusions:** African-American race is a significant independent predictor of propensity to use the ED across stages of disease, even with adjustment for access and other variables Wait times for appointments, and in early disease insurance and income, are also important predictors Cultural preferences, habits, or attitudes may play a role in propensity to use the ED for care

	Early	Late
African-American¶	1 54*	1 02*
Hispanic¶	1 18†	ns
Private Insurance∫	- 58§	ns
Higher Income	- 12§	ns
More Education	ns	- 34§
Longer wait Appt	46§	32§
Time with provider	ns	- 13†
Longer get to Appt	ns	22†
Anxiety-Depression	01§	ns

*p< 001, †p< 01, §p< 05
¶Ref Whites, ∫Ref no insurance

THE RELATIONSHIP BETWEEN PREFERENCES FOR MEDICAL DECISION CONTROL AND THE OUTCOME OF INFORMED CONSENT AMONG ADVANCED CANCER PATIENTS RECEIVING EXPERIMENTAL THERAPIES

CK Daugherty, DM Banik, MJ Ratain University of Chicago Chicago, IL

We hypothesize that there is a relationship between the locus of medical decision control (DC) to the outcome of the informed consent (IC) process for patients (pts) receiving investigational therapies available in early clinical trials In order to describe this potential relationship, we have been surveying advanced cancer pts who have consented to participate in phase I (toxicity defining) trials at our institution using the medical DC instrument of Llewellyn-Thomas and a structured interview regarding specific elements of the IC process Pts eligible for phase I trials at our institution are those who face a life ending cancer diagnosis for whom a standard of care has either failed or is not available, and for whom no chemotherapy is an acceptable alternative as no treatment has been shown to improve survival for these patients Since August 1997, 104 subjects have been studied Median age 60, 61 males and 43 females 87% Caucasian, 26% college educated Regarding DC, 27% of subjects preferred their physician (MD) make decisions of care (with or without patient input) 30% stated they (the pts) should make decisions (with or without MD input), 44% stated DC should be an equally shared process Regarding IC, overall only 22% of all pts were able to state the research purpose of phase I trials and 45% recalled the alternative of no chemotherapy being part of the IC process Of pts who preferred pt-centered DC, 38% were able to state the research purpose of phase I trials whereas only 9% of pts who preferred MD-centered DC were able to state the purpose of a phase I trial This difference was statistically significant (p<0 05) 38% of pts who preferred pt-centered DC recognized the alternative of no chemotherapy, whereas 60% of pts who preferred MD-centered DC recognized this alternative (pNSS) A significant proportion of pts participating in phase I clinical trials prefer pt-centered DC Pts who prefer MD-centered DC may be less likely to give ethically adequate IC for experimental therapies Such inadequate IC contributes to a less than ideal process of medical decision-making for potentially vulnerable pts with life-ending diagnoses Comparative studies of similar pts outside the early clinical trial setting would be enlightening

TREATMENT THRESHOLDS AND POSTDECISION REGRET

A Schwartz, B Djulbegovic, I Hozo University of Illinois at Chicago, Chicago, IL, University of Louisville, Louisville, KY, and Indiana University Northwest, Gary, IN

Purpose To examine the probabilities of disease at which medical students or clinicians would be comfortable recommending for or against a risky treatment and the regret they would feel in ensuing clinical scenarios

Methods A group of 69 medical students and clinicians were given a vignette describing a risky treatment for suspected breast cancer recurrence following a lumpectomy, levels of risk and benefit of the treatment were crossed in a factorial design Respondents indicated the probability of cancer recurrence below which they would recommend against the treatment, and the probability of cancer recurrence above which they would recommend the treatment They also indicated how much regret they would experience if the patient died under six conditions based on whether the clinician was in favor of, against, or neutral about the treatment, and whether the patient chose the treatment or not

Results For both students and clinicians, assessed probabilities differed considerably from predictions of the threshold model (Pauker & Kassirer, 1975) and often identified a wide range of probabilities at which respondents would neither recommend in favor of nor against the treatment In only 3 out of 9 of the conditions did the range of mean probabilities include the normative threshold probability Assessed thresholds for recommending the treatment varied with treatment benefit, but not risk, thresholds for recommending against the treatment varied with risk, not benefit Both thresholds were only mildly correlated with the normative threshold probabilities Respondents reported significantly greater regret when patients died following their recommendations than when patients died from decisions where no recommendation had been made or decisions that were contrary to the respondents' recommendations

Conclusions Normative models specify a threshold probability above which a disease should be treated and below which it should not be treated, this probability depends on the risks and benefits associated with the treatment This study suggests that clinicians may instead consider separate thresholds for treatment and non-treatment, and that the normative probability may lie outside this range Furthermore, the regret associated with the patient's death was greater when the patient died following the clinician's advice than otherwise, this may represent a variation of omission bias

UNSTABLE PREFERENCES A SHIFT IN VALUATION OR AN EFFECT OF THE ELICITATION PROCEDURE?

S J T Jansen, A M Stiggelbout, M A Nooy, E M Noordijk, J Kievit Departments of Clinical Oncology and of Medical Decision Making, Leiden University Medical Center, The Netherlands

Objective. Many studies suggest that health states are valued more positively once experienced, than when still hypothetical We investigated to what extent discrepancies do occur between hypothetical utilities and actual value judgements, and whether such discrepancies are caused by an actual change in valuation of the same health state over time, or by differences between hypothetical scenario descriptions and actual experience

Patients and Methods. Seventy breast cancer patients evaluated their actual health state and a radiotherapy scenario before, during, and after a post-operative radiotherapy treatment A chemotherapy scenario was evaluated as a control scenario All scenarios captured at least those somatic, psychological, and social aspects per health state, that were judged relevant by professional experts and patients Values and utilities were elicited by means of a VAS, a chained TTO, and a chained SG

Results. Fifty-five patients completed all three interviews The utilities for the hypothetical radiotherapy scenario (0 89), evaluated before the radiotherapy treatment, and the actual experience of undergoing radiotherapy (0 92), evaluated during the treatment, were significantly different by means of the TTO ($p<0$ 05) When both health states were evaluated at the same time during radiotherapy, significant differences ($p<0$ 01) were seen by means of the VAS and the SG The utilities for the hypothetical radiotherapy scenario itself remained stable over time, like the utilities for the chemotherapy scenario

Conclusion. Our results suggest that utilities for hypothetical health scenarios remain stable over time, but that utilities obtained through hypothetical health state scenarios may not be valid predictors of the value judgements of real life health states, when actually experienced The implications of these findings for the use of decision analysis and cost-effectiveness analysis on both the micro- and higher levels of health care are discussed

DESIGN AND IMPLEMENTATION OF A COMPUTER-BASED SYSTEM TO ANNOTATE DECISION MODELS FOR USE IN GUIDELINE DEVELOPMENT

GD Sanders, RF Nease Jr, DK Owens Stanford University, Stanford, CA, Washington University, St Louis, MO, and VA Health Care System, Palo Alto, CA

Decision models (DMs) provide a normative framework for representing the evidence, outcomes, and preferences required for guideline development DMs cannot be used directly by computer-based systems to create clinical-practice guidelines (CPGs), however, because DMs do not contain all of the information needed to develop a CPG The goal of this project was to develop a method to enable computer-based development of CPGs from DMs

To determine what information was needed to develop CPGs from DMs in computer-based systems, we developed conceptual models (CMs) for CPGs and DMs The CPG CM specifies the data, knowledge, and functional requirements for development, maintenance, and use of CPGs The DM CM specifies the knowledge contained in DMs, including models that require sequential decisions, dual utilities, or Markov processes We designed and implemented a web-based DM annotation editor to develop annotated DMs that could be used in computer-based systems The DM annotator uses the CPG CM to determine what information not contained in the DM is needed to develop a CPG The system's interface allows the decision analyst to load a DM remotely The system then parses the DM, and dynamically creates web-based DM annotation forms to query the analyst for the required information We used the system to create annotated DMs for staging lung cancer and for the use of implantable cardioverter defibrillators (ICDs)

For the lung-cancer DM, the DM annotator queried the analyst for the objective of the CPG, the targeted patient population, the sensitivity analysis ranges, evidence tables, the underlying assumptions and perspective of the CPG, the methods used for validation of the guideline, the sponsors of the guideline, and any needed definitions of model variables For the ICD DM which is a dual-utility Markov model, the DM annotator also queried the analyst for information regarding the discount rate, methods used for cost inflation, the desired cost-effectiveness threshold, the Markov process cycle length and the time horizon

Our system demonstrates the use of a CPG CM to dynamically collect information needed to create a CPG directly from a DM Because the CM is comprehensive, the system can evaluate a broad range of DMs, and can automatically create a web-based tool that prompts analysts for information needed to use the DM in CPG development Such tools bridge an important gap between DMs and CPGs

IS MORE REALLY BETTER? NOT ALWAYS THE ATTRACTION EFFECT IN PHYSICIANS' DECISIONS ABOUT MEDICATIONS

JA Schwartz, GB Chapman, LY Niedermayer Rutgers University

According to the normative principle of regularity the addition of an inferior alternative to a choice set should not increase the likelihood that one of the original alternatives will be chosen However, in the Attraction Effect (Huber, Payne, & Puto, 1982) the addition of such an alternative (a decoy) increases the likelihood of choosing the dominant alternative (the target) over the non dominant alternative (the competitor) This decision making bias has been examined in consumer decision making, but has not yet been explored in physician decision making In this study we examined whether the attraction effect would occur in physicians' decisions about medications

Forty Internal Medicine residents (57% response rate) participated in a questionnaire study in which they read three patient cases (depression, sinusitis, and vaginitis) and chose the most appropriate medication from a choice set of two or three options In the "two option" condition, one medication was superior on efficacy or length of treatment, whereas the other had fewer potential side effects In the "three option" condition, an inferior third medication (the decoy) was added to the choice set, thereby defining the target as the dominant alternative There were two measures of preference for each drug choice and rating

In two of the three scenarios the percentage of subjects who chose a particular medication was higher when that medication was the target than when it was presented in the two option condition That is, averaged across the three scenarios, choice of the target drug was more frequent when the decoy was present than when it was absent (67% vs 50%) An analysis of medication rating also revealed a significant drug by condition interaction ($p <$ 01), indicating that the appropriateness rating of the target varied depending on whether or not a decoy was present Overall, the mean rating for the "target" when the decoy was present was 4 25, vs 3 74 when the decoy was not present This study is the first demonstration of the attraction effect in physician decision making

WHO WANTS TO KNOW? GENETIC TESTING FOR BREAST CANCER SUSCEPTIBILITY IN THE GENERAL POPULATION

K Armstrong, M.D , Peter Ubel, M D , Barbara Weber, M D , J Sanford Schwartz, M D Division of General Internal Medicine, University of Pennsylvania

Objective To investigate the relationship between interest in genetic testing for breast cancer susceptibility and family history of breast cancer in the general population

Methods Mailed survey of 410 women randomly selected from patients of a faculty General Internal Medicine Practice at the University of Pennsylvania Survey instrument measured perceived lifetime risk of breast cancer, family history of breast cancer, awareness of *BRCA1/2* testing, interest in testing and sociodemographic characteristics Women with breast or ovarian CA were excluded from final analysis

Results 247 women (60%) returned the survey to date 24 women had breast cancer and 4 women had ovarian cancer While women with breast CA were slightly more likely to report interest in testing, the association was not significant (OR 1 7, 95% CI 63-4 74) 113 women (53%) knew of BRCA1 testing, 10 (5%) were unsure and 88 (42%) did not Knowledge about testing was not associated with a family history of breast cancer 121 women were interested in undergoing testing, 47 were unsure about testing and 42 were not interested in testing Table 1 shows characteristics of women who had heard of testing by interest in testing Family history of breast cancer remained a significant predictor of interest (OR 0 32, 95% CI 0 11-0 85) after multivariate adjustment From the women who were interested in testing, 74% wanted the decision about testing to be made with their doctor, 24% on their own and 2% by their doctor

	Interested in Testing	Not Interested in Testing	p-value
Mean Age	52 1	57 6	0 04
Caucasian (%)	0 68	0 62	0 42
College Education (%)	0 61	0 50	0 21
FH Breast CA (%)	0 13	0 32	0 02
Mean Perceived Lifetime Risk of Breast CA	32 0%	33 5%	0 77

Conclusions Although family history is the most important known risk factor for carrying a mutation in *BRCA1* or *BRCA2*, interest in genetic testing in the general population is negatively associated with a family history of breast cancer Because testing low risk individuals is costly and potentially misleading, efforts are needed to target the use of genetic testing in the general population to individuals at risk

USING MULTIPLE-SIGNAL INTEGRATION TO DETECT FALSE ALARMS IN THE INTENSIVE CARE UNIT

CL Tsien, IS Kohane, N McIntosh Harvard Medical School, Boston, MA, Massachusetts Institute of Technology, Cambridge, MA, and Children's Hospital, Boston, MA

The high incidence of false alarms in the Intensive Care Unit (ICU) necessitates the development of improved alarm algorithms Single-signal methods have been attempted with limited success The current study aims to detect patterns across multiple physiologic data signals in order to decide whether or not a false alarm is present on one of those signals Methods Approximately 240 hours of ICU data collected in 1-minute intervals from more than 70 patients were analyzed Available signals included heart rate (HR), blood pressure (BP), and partial pressure of carbon dioxide (CO2) False alarm occurrences in the data streams were visually located and annotated retrospectively by an experienced clinician (N M) For each data stream, the following values were then calculated for every successively overlapping 5-minute interval maximum value, minimum value, range, mean, median, linear regression slope, standard deviation For each 5-minute period, a label of false alarm, not false alarm, or transition, was also assigned for the CO2 signal according to the recorded annotations For each non-transition period, the 21 calculated values along with the classification label were divided randomly by patient number into a training set and test set The training data set was then used to derive a decision tree model for classifying 5-minute data sets as being CO2 false alarms or not Of the 21 available predictors, nine were determined by C4 5, a decision tree generator, to best classify the data These nine predictors were min CO2 value, max CO2 value, CO2 slope, CO2 range, CO2 standard deviation, max HR value, HR slope, min BP value, and max BP value The model was then evaluated on the reserved test set, which consisted of 7055 data periods Of those, the model correctly classified 6690 out of 6746 periods as no false alarm (99 2%), and 258 out of 309 periods as false alarm (83 5%), with an overall accuracy of 98 5% These results indicate that integration of multiple signals may be a viable approach to detecting false alarms Future work should include exploration of other machine learning methods, use of additionally derived values, and prospective trials

SENSITIVITY OF SCREENING MAMMOGRAPHY AMONG WOMEN AGED 40 TO 49 YEARS

S K Plevritis and A M Garber Departments of Radiology and Health Research and Policy Stanford University, Stanford, CA and VA Health Care System, Palo Alto, CA

Purpose: To analyze the sensitivity of screening mammography among women aged 40 to 49 years as a function of tumor growth, screening interval and diagnostic test performance in dense fibroglandular tissue

Method: By enhancing a breast cancer screening model that we've previously presented, we can more closely analyze the factors that may determine the sensitivity of screening mammography The sensitivity of screening mammography is defined here as the number of screened detected cancer patients divided by the number of screened detected cancer patients plus cancer patients clinically detected between screening exams The model combines prior knowledge of breast cancer incidence and stage at detection in an unscreened population with assumptions of uniform and heterogeneous tumor growth rates and has been validated by closely reproducing data observed in randomized controlled trials We apply the model to women ages 40 to 49 years at average risk for developing breast cancer and quantify how the sensitivity of screening mammography may be effected by factors such as tumor growth rates, diagnostic test performance in dense fibroglandular tissue and the screening interval

Results: If mammography detects all tumors 1cm or larger, then for tumor volume doubling times (TVDTs) equal to 2, 4, 6, 8, 10 and 12 months, the sensitivity of annual screening is 6%, 68%, 80%, 85%, 87%, and 88% respectively, and of biennial screening, 3%, 36%, 62%, 73%, 79%, and 82% respectively If mammography detects only 70% of tumors sized 1cm or larger, then the sensitivity of annual screening is 4%, 48%, 66%, 74%, 79% and 82% respectively, and of biennial screening, 2%, 25%, 43%, 54%, 62% and 67%, respectively

Conclusion: The incremental increase in sensitivity of screening every year compared to screening every two years is highest when TVDTs range from 4 to 6 months and mammography detects all tumors sized 1cm or larger If mammography misses some tumors sized 1cm or larger this incremental increase in sensitivity can be substantial even when TVDTs are longer than 6 months

PREFERENCE VALUES FOR CORTICOSTEROID-ASSOCIATED ADVERSE HEALTH STATES IN 107 RHEUMATOID ARTHRITIS PATIENTS

L Merlino, J Bagchi, T Taylor, W Summner, R Nease, K Saag University of Iowa, Iowa City IA and Washington University, St Louis, MO

Preferences for health states associated with rheumatoid arthritis (RA) have not been extensively evaluated Despite common use and definite benefits, corticosteroids (steroid) can cause many serious adverse events (AEs) in RA patients To better understand the relative importance, predictors, and psychometric properties of steroid AE preferences, we developed scenarios depicting 14 commonly reported outcomes of 5 serious steroid AEs (fractures [fxs], infections, cataracts, ulcers, and diabetes) Scenarios were written by clinician-investigators and modified based on input from 10 specialists Fx scenarios were refined in focus groups U-titer (Wash U) was programmed to measure scenario preferences (0=death, 1=current health) by category scaling (CS) and either time tradeoff (TTO) (chronic states) or sleep tradeoff (STO) (temporary states) A single interviewer facilitated all measurements and assigned an ordinal score (from 0=did not comprehend to 10=fully comprehended) for subjects' ability to perform TTO/STO ratings After pilot testing (n=18), preferences were elicited in 107 RA patients Test-retest reliability (n = 15) was assessed 1 to 2 months later As a measure of construct validity, 10 rheumatologists ranked the scenarios in order of severity Characteristics of subjects included (mean or %) 75% women, age 61 yrs, 16 yrs duration of RA, 49% current steroid users, and 54% experienced □ 1 of the AEs of interest Mean preference weights (CS/[STO or TTO]) for the AE scenarios included (n=19 to 51) hip fx requiring nursing home stay (55/ 76), chronic pain from vertebral fxs (58/ 65), wrist fx (85/ 80), septic arthritis (63/ 60), pneumonia (82/ 76), ulcer requiring surgery (72/ 84), new onset diabetes (69/ 83), and cataracts (82/ 98) Only 6 5% of subjects strongly agreed the preference process was confusing, 88% enjoyed the process and 93% were willing to retest Seven subjects had an ordinal comprehension score of < 2 with average time for completion of 42 min compared to 30 min for the other subjects Intraclass-correlation coefficients for test-retest reliability were 59 (VAS), 73 (STO), and 60 (TTO) Ordering of CS, more so than STO/TTO, corresponded with rheumatologist rankings (r= 82 [p< 01]) For scenarios of varying severity within a given health state, logical adherence was generally preserved Increasing age was the only factor significantly associated with low comprehension (OR=1 2, CI=[1 0497, 1 3284]) In summary, RA patients assign low preference values to many steroid-associated AEs Scenarios depicting more serious/chronic outcomes of hip & vertebral fxs had the lowest preferences Preferences measured by CS were slightly lower than TTO/STO for most outcomes Preference assessments for RA patients were in general valid, reliable, and feasible, although more so for younger patients and for CS over TTO/STO

THE EFFECT OF FRAMING BIAS ON THE TIME TRADE-OFF

TG Gamats, T Gilmer, SM Ng, E Kwong Department of Family and Preventive Medicine, University of California San Diego, La Jolla, CA

To evaluate the effect of framing bias on the results of the time trade-off (TTO)

Two sets of time trade-off scenarios involving two health states (wearing glasses and asthma) were developed. One set of scenarios used the standard time trade-off approach of exchanging a shorter life for better health. The second set of scenarios, using the same health states, exchanged longer life for poorer health. For scenarios that exchanged better health subjects were first presented trade offs of their entire life expectancy and zero years For scenarios that exchanged longer life expectancy subjects were given trade offs of each subject's age-adjusted life expectancy and this life expectancy plus 20 years Alternatively high and low options were presented until the subject was unable to choose between the options The scenarios were offered to a convenience sample of 22 adults via U-titer We then calculated each subject's TTO utilities for each condition as measured by the two scenarios

All subjects completed all questions For the condition of wearing glasses, the TTO utility was 0 94 when exchanging better health and 0 58 when exchanging longer life (p = 0 001) For asthma the TTO utility was 0 83 when exchanging better health and 0 69 when exchanging longer life (p = 0 002)

The time trade-off is a standard method for assessing patient utilities, and disagreement with TTO utilities has been used to criticize other health state preference systems If the TTO is to be a gold standard, it should be free of framing bias If the framing of the question significantly affects the TTO estimate of a health state utility, one would not know which framing is "correct " In this study we simply reversed the standard TTO approach. Instead of "buying" health with a currency of life expectancy, we also offered subjects the opportunity to "buy" life expectancy with a currency of health. For example, for the TTO approach to be a gold standard method for assessing the utility of wearing glasses, one would expect that the disutility of wearing glasses would be the same for each approach That it is does not draws into question the TTO methodology for assessing health states for policy purposes In reality both approaches used in this study (buying life expectancy and buying life) are options patients choose between in different clinical settings Therefore, it should be emphasized that for an individual patient in an actual clinical scenario, where the framing for a particular health state is known, the TTO may still be helpful

RELATIONSHIP BETWEEN UTILITY SCORES AND DISEASE-SPECIFIC QUALITY OF LIFE MEASURES IN A BREAST CANCER POPULATION

MA Namjoshi, TN Taylor, J Brooks, R Hohl University of Iowa, Iowa City, USA

Background Previous research assessing the relationship between utility scores and quality of life measures has reported poor to modest correlations These studies, however examined the relationship between generic health-related quality of life dimension scores and utility scores As the use of generic health-related quality of life measures in oncology is very limited, it is important to assess the relationship between utility scores and disease-specific quality of life measures

Purpose To assess the relationship between utility scores and disease-specific quality of life(QOL) measures in a breast cancer population

Methods After a pretest(n=20) to assess feasibility, 75 breast cancer patients participated in a structured interview Patients completed two cancer-specific QOL instruments, the Functional Assessment of Cancer Therapy-G(FACT-G) and the Functional Living Index-Cancer(FLIC) Utilities for current health were assessed using the Standard Gamble(SG), the Health Utilities Index-Mark III(HUI), the Visual Analog Scale(VAS), and the Euroqol-5D(EQ) The order of instrument administration varied for each patient Additional information on patient characteristics was obtained by chart review

Results The mean age was 60(SD=10 8) The majority of patients(96%) had either stage I or stage II disease Mean utility scores were 93 3(SD=11 5) for the SG, 84 6(SD=16 8) for the HUI, 86 6(SD=12 6) for the VAS, and 87 4(SD=13 6) for the EQ Multiple regression results indicated that the FACT-G dimension scores explained 26% of the variation in the SG, 39% in the HUI, 50% in the VAS, and 47% in the EQ, while FLIC dimension scores explained 42% of the variation in the SG, 43% in the HUI, 54% in the VAS, and 56% in the EQ These results were not sensitive with respect to the choice of alternative functional forms The physical, psychological, and social well-being dimensions on the FLIC were significant determinants of utilities The only significant dimension from the FACT-G was physical well-being Presence of comorbid conditions was negatively associated with utility scores

Conclusion Utility assessments for current health among breast cancer patients are feasible, and the scores are significantly related to cancer-specific QOL scores, which may provide a basis for mapping QOL scores into utilities in future studies

COMBINING UTILITY MEASUREMENTS: EXPLORING DIFFERENT APPROACHES.

DE Bonds, KA Freedberg Boston Medical Center Boston University School of Medicine Boston MA

Objective: Using health state utilities from the literature can improve the quality of cost-effectiveness research Our objective was to evaluate the impact of different approaches to combining health states utilities on cost-effectiveness results

Methods: We designed a decision analytic model to test the cost-effectiveness of screening for postpartum thyroiditis (PPT) in pregnant diabetics (IDDM) The model uses a one-year time frame Utility measurements for IDDM and PPT were available in the literature however, no measure of having both IDDM and PPT existed To determine aggregate utilities we used two different methods In the additive model IDDM was assigned a baseline utility of 0 84 (from the literature) and IDDM with PPT determined by adding the disutility of each health state and subtracting this total from 1 0 In the multiplicative model IDDM was again given the baseline utility of 0 84 and aggregate utilities determined by multiplying the utility values of PPT and IDDM together

Results: The additive method resulted in the following utilities IDDM alone 0 84, IDDM with untreated PPT 0 58, IDDM with treated PPT 0 68 Utilities determined by the multiplicative method were higher IDDM alone 0 84, IDDM with untreated PPT 0 62, IDDM with treated PPT 0 71 For a hypothetical cohort of 1000 women the results are

Method	Strategy	Cost ($)	Effectiveness (QALYS)	Incremental C/E ($/QALY)
Additive	No Test	$108,020	798 4	---
	Screening Test	$215 700	810 5	$15 960/QALY
Multiplicative	No Test	$108,020	805 0	---
	Screening Test	$215,700	815 8	$17 730/QALY

Sensitivity analysis showed that the results were most sensitive to the sensitivity of the screening test and the incidence of PPT However, despite substantial variation in the utility estimates, the cost-effectiveness ratio was not sensitive to the method of combining utilities

Conclusion: We conclude that where standard methods of combining utilities are not available, investigators should be explicit about their approach In a decision with a short time horizon, the method of combining utilities may have less impact on cost-effectiveness than other parameters

Health Informatics Distribution for Evidence-Based Decision Analysis. A World Wide Web Based Interactive Clinical Decision Support System.

Aoki N [1, 2], Fearn PA [3], JR Beck [1], Kattan MW [3], Ohbu S [4], Fukui T[2]
1) Information Technology, Baylor College of Medicine, 2) Department of General Medicine and Clinical Epidemiology, Kyoto University Hospital, Kyoto, Japan, 3) Scott Department of Urology, Baylor College of Medicine, 4) St. Luke's international hospital, Tokyo, Japan

Purpose: Although the decision analytic approach is a useful tool, it is a time-consuming process and requires specific knowledge of theory and computer skills This is one reason why many physicians do not currently apply it to the clinical decision making process We have developed the prototype of an interactive clinical decision support system which can be used through the World Wide Web (WWW), called "Web and Evidence-based clinical Decision Support system (WEDS)"

System Overview: WEDS utilizes several technologies, HTML, Microsoft Active Server Pages, Microsoft Visual Basic, Java Script, TreeAge ActiveDATA control, and Decision Analysis By TreeAge (DATA) version 3 0 17 The decision models are created in DATA, and we use the ActiveDATA control to process the form values obtained thorough the web with the DATA engine The ActiveDATA control provides the results in quality adjusted life expectancies according to the user defined probabilities and utilities The results from the ActiveDATA control are displayed on the web using Microsoft Active Sever Pages Clinic-U, an original utility elicitation program, is incorporated into WEDS The ActiveDATA control also provides graphs which can be displayed on the web

Decision Support Process: The clients who visit WEDS can produce decision analysis results in the following five steps 1) Select the decision model (if there is no available model, they can contact us to make a new original model), 2) Review the background of the problem and source literature, 3) Change the probabilities and utilities in the model (default probabilities are set based on the source literature), 4) Perform baseline analysis, 5) Perform sensitivity analyses with output graphs (JPEG files)

Conclusion: WEDS is easy to use, flexible, extendible, and multilingual, which enables any physician to access WEDS worldwide Physicians can perform evidence-based clinical decision making without the time-consuming process, special knowledge, and experience required to master decision theory

EVALUATION OF CHRONIC DIARRHEA IN AIDS A COST-EFFECTIVENESS ANALYSIS

SA Call, CM Wilcox, MS Saag, GR Heudebert University of Alabama at Birmingham (UAB), Dept of Internal Medicine, Birmingham, AL

Purpose Diarrhea is a common symptom in persons with AIDS Prior work has emphasized the importance of endoscopy in evaluation, but no studies have been done in the highly active antiretroviral therapy (HAART) era The purpose of this study was to evaluate the cost-effectiveness of outpatient evaluation strategies in the current era of AIDS care

Methods Three evaluation strategies were compared stool studies first (SS), lower endoscopy first (LE), and empiric metronidazole treatment (EM) Probabilities of specific etiologies of diarrhea were based on data obtained from a retrospective chart review at the UAB HIV clinic Cases were identified if they had a CD4 count <200, diarrhea > two weeks, and underwent evaluation Treatment efficacy and testing performance data were obtained by literature review Cost data was obtained from hospital billing Average Wholesale Price was used for medication cost The primary effectiveness measure was days free of diarrhea (DFD) Analysis was conducted from the third party payer perspective with a time horizon of two months Sensitivity analyses were done on costs and probabilities of specific causes The baseline probability used for opportunistic infectious (OI) causes was 0 125, for infectious (non-OI) causes 0 225, and for non-infectious causes 0 65 Changes in either OI or infectious (non-OI) prevalence affected prevalence of non infectious causes exclusively

Results In the base case analysis EM was dominated by LE

Strategy	Cost ($)	Efficacy (DFD)	Marginal C/E ($/DFD)
Stool first (SS)	973 32	25 89	
Lower endoscopy first (LE)	1016 95	26 17	151 45

The model was sensitive to changes in the probability of an OI etiology of diarrhea (EM preferred at probabilities less than 09) The model was also sensitive to changes in the probability of an infectious (non OI) cause (EM preferred at a prevalence of 0 32 and dominating the LE approach) The model was not sensitive to changes in cost of either procedure or stool culture

Conclusions Despite using prevalence of causes of diarrhea in the era of HAART therapy, endoscopy first remains a reasonable approach However, choice of evaluation strategy is strongly influenced by the prevalence of different etiologies of diarrhea in this population

EVALUATION OF DIAGNOSTIC TECHNOLOGIES APPLICABILITY OF CURRENT STRATEGIES

JG Lijmer, BW Mol, GJ Bonsel, MH Prins, PMM Bossuyt Academical Medical Center, University of Amsterdam, Amsterdam, the Netherlands

Background and Purpose Before a new diagnostic technology is introduced in clinical practice its effectiveness should be assessed In contrast with the evaluation of new pharmaceutical compounds, there is no consensus on how such an evaluation should proceed Several evaluation strategies have been proposed in the literature We wanted to examine their applicability for diagnostic test evaluation

Methods Articles describing strategies for the evaluation of diagnostic technologies were retrieved using the Medline and PubMed database We tried to apply these strategies to research questions from 38 grant proposals of a national MTA programme (1995-1997) in the Netherlands

Results We detected eleven articles that reported a strategy for the evaluation of a diagnostic test from its development until its introduction into clinical practice The phases of test evaluation in these articles appeared very similar Many of them were based on a strategy proposed by Fryback and Thornbury in 1991 The strategies revealed an analogy to the four phases of pharmaceutical treatment assessment (I) Test development, (II) Diagnostic accuracy, (III) Clinical impact, (IV) Societal efficacy Phase III is subdivided by some authors into (1) impact on diagnosis, (2) impact on therapy and (3) patient outcome

Applying this strategy in four phases to the grant proposals revealed some limitations First, in 42% of the proposals there was no reference standard available for the test under examination Many proposals contained new diagnostic technologies resulting in new diagnostic classifications, limiting the possibility of a 'phase II' study Second, the strategy requires that the test is used to obtain a diagnosis In 18% of our proposals the test was used for other reasons (monitoring and prognosis) Third, impact on diagnosis and impact on therapy (proposed in 5%) rather provide information on the behaviour of physicians than on the effectiveness of the test under examination Fourth, this strategy does not take into account the future use of a test, which can either complement or substitute existing diagnostic technologies

Conclusions Current strategies for the evaluation of diagnostic technologies are limited in their applicability A more complete strategy is required, incorporating a proposal for tests that lead to new diagnostic classifications and tests used for other than diagnostic purposes

WHY DON'T PEDIATRICIANS FOLLOW ASTHMA PRACTICE GUIDELINES? APPLYING THE AWARENESS TO ADHERENCE MODEL AND SOCIAL COGNITIVE THEORY

MD Cabana, BE Ebel, L Cooper-Patrick, CS Rand, NR Powe Johns Hopkins University, Baltimore, MD

We conducted a series of focus groups with practicing pediatricians, representing 11 states and one foreign country, to identify barriers to physician adherence to the 1997 National Heart, Lung, Blood Institute (NHLBI) Asthma Guidelines We assigned 22 pediatricians to one of 3 focus groups organized by year since medical school graduation (median year '61, '81, '91) We asked participants open-ended questions to identify barriers to the successful use of different guideline components including prescription of daily inhaled steroids, recommendation of daily peak flow meter use, smoking cessation counseling, and environmental allergen control Two investigators independently analyzed transcripts of the audiotaped conversations, created taxonomies of the barriers, and placed barriers into one of 5 domains (Lack of Awareness, Lack of Agreement, Lack of Self-Efficacy, Lack of Outcome Expectancy, and Lack of Consistency) developed from the Awareness to Adherence Model of Guideline Compliance and Social Cognitive Theory Three additional investigators independently validated the taxonomy organization We identified 170 comments about barriers to guideline adherence, of which 91% could be placed in a specific domain We found that different components of the guidelines (e g prescribing inhaled steroids, smoking cessation counseling, etc) elicited a different domain of barriers Physician characteristics (i e years since graduation from medical school) also affected which domain of barriers was prominent Senior physicians mentioned barriers from the "Lack of Agreement" domain (i e fear of steroid side effects) as barriers to prescribing daily inhaled steroids and barriers from the "Lack of Self-Efficacy" domain (i e lack of familiarity) as barriers to recommending the daily use of peak flow meters, while younger physicians mentioned barriers from the "Lack of Self-Efficacy" domain (i e lack of training) as a barrier to smoking cessation counseling All groups mentioned "time pressures" as a barrier to adherence for all components Focus groups were useful in identifying barriers to guideline adherence This qualitative analysis suggests that organizing these barriers using the Awareness to Adherence Model and Social Cognitive Theory may help select interventions to improve adherence to the NHLBI guidelines by identifying which domain of barriers is prominent, which is influenced by the guideline component in question as well as physician characteristics, such as years since medical school graduation

COST-EFFECTIVENESS OF SCREENING FOR ANAL SQUAMOUS CELL CANCER IN HIGH-RISK MEN

SJ Goldie, KM Kuntz, MC Weinstein, KA Freedberg, JM Palefsky Harvard School of Public Health, Boston University School of Medicine, University of California San Francisco

Objective: The incidence of anal squamous cell cancer (SCC) in men who have sex with men (MSM) exceeds that of cervical cancer in women. Screening for anal squamous intraepithelial lesions (ASIL) in a manner analogous to cervical disease may prevent the development of anal SCC Our objective was to evaluate the cost-effectiveness in dollars per quality-adjusted life year ($/QALY) of screening for ASIL in MSM

Methods: We developed a state transition model to incorporate the natural history of ASIL, screening test performance, competing mortality, and costs of screening, treatment, and follow-up for ASIL Using data from literature, national databases, and a prospective study of the natural history of ASIL in MSM, we evaluated the following strategies (1) No screening, (2) Triennial anal Papanicoalou smears (Pap), (3) Biennial Pap, (4) Annual Pap, (5) Triennial anoscopy, (6) Biennial anoscopy, (7) Annual anoscopy, and (8) Semiannual Pap Selected input parameters annual incidence of low-grade ASIL (4 4%), high-grade ASIL (4%), regression ASIL (25%), sensitivity and specificity of anal Pap (50% and 92%), cost of anal Pap ($26) and anoscopy ($140) Because of the uncertainty of progression and treatment efficacy of high-grade ASIL we intentionally biased the analysis against screening

Results: Screening MSM was associated with QALY gains between 1 3 and 1 8 months Incremental cost-effectiveness of triennial Pap was $3,700/QALY, biennial Pap was $11,900/QALY, and annual Pap was $27,900/QALY Triennial and biennial anoscopy were dominated by semiannual Pap at $119,100/QALY Results were most sensitive to the progression rate and treatment efficacy of high-grade ASIL

Conclusions: The cost-effectiveness of biennial anal Pap screening for ASIL in MSM is comparable to other well-accepted preventive medical interventions In men at higher risk of ASIL progression, such as those with persistent high risk HPV types, annual screening may be a reasonable strategy A policy of screening MSM biennially for ASIL should be considered

A DECISION-ANALYTIC MODEL TO ASSESS THE EFFECTIVENESS OF BONE MARROW TRANSPLANTATION FOR CHILDREN WITH HURLER SYNDROME

LN Venditti, SJ Lee, CP Venditti, C Peters Harvard School of Public Health, Boston, MA,Dana-Farber Cancer Institute and Brigham and Women's Hospital, Boston, MA, Department of Pediatrics, Massachusetts General Hospital, Boston, MA, Department of Pediatrics, University of Minnesota, Minneapolis, MN

Hurler syndrome (MPS1H) is a rare, progressive autosomal recessive inborn error of metabolism that results in the accumulation of heparan sulfate and dermatan sulfate, leading to organ dysfunction and premature death While untreated patients experience rapid neurologic deterioration and death by 5 years of age, bone marrow transplantation (BMT) to normalize enzymatic activity has been shown to significantly ameliorate or halt this course However, the value of BMT is still controversial because of its considerable risks and costs

Data from the Storage Disease Collaborative Study Group, a consortium of 25 institutions in the US and Canada and 100 investigators created to study of the value of BMT for storage diseases, was used to build a model comparing transplantation and no transplantation on the basis of discounted, quality-adjusted life expectancy Using a Markov model, this study evaluates the effectiveness, both unadjusted and quality-adjusted, for the 10-year period beyond the initial decision to undergo transplantation in a hypothetical cohort of 1 year-old children

We conducted threshold analyses to derive the health state utilities required in order to change the optimal decision from not transplanting to transplanting In summary, when using a relatively high estimate of the non-BMT health state utility in the related donor model, the transplant option will usually be optimal given reasonable estimates of the BMT state utilities relative to the non-BMT On the other hand, with the unrelated model, we must assign much higher utilities to the BMT states in order to gain more QALYs with the transplant option then with the non-BMT option

Though transplanting does prolong life, the sensitivity analyses show the quality-adjusted effectiveness is heavily influenced by the quality of life of the post-BMT engraftment state This quality of life captures a state characterized by corrected enzyme levels and significant resolution of disease but potentially, years of multiple corrective surgeries and possible unfavorable neurological development

CLINICAL MANAGEMENT OF ACUTE BACTERIAL SINUSITIS A DECISION AND META-ANALYSIS

EM Balk, EA Engels, DR Zucker, J Lau Tufts-New England Medical Center, Boston, MA

Acute Bacterial Sinusitis (ABS) is difficult to diagnose accurately and the value of therapy is unclear We performed a decision analysis using data obtained from meta-analyses to compare empiric amoxicillin treatment (Emp-Tx), initial symptomatic treatment (Sx-Tx) and sinus-radiography-guided treatment(XR-Tx)

We used a Markov process to evaluate days-to-cure The decision model incorporated ABS prevalence, time-varying probabilities of cure, side effects (SE) rates, and sensitivity and specificity of X-ray We performed meta-analyses to derive estimates of treatment efficacy and X-ray sensitivity and specificity Using data reported in clinical studies, we fit Weibull curves to estimate daily cure rates The outcome measurement estimated symptom (Sx)-free days (whether Sxs were due to ABS or drug SE) The model evaluated a 14 day course From our meta-analyses, we obtained a cure rate of 86% at day 14 for Emp-Tx and 41% for Sx-Tx, and a sensitivity of 0 89 and specificity of 0 62 for XR We assumed a 90% cure rate at day 14 for those without ABS From a review of the literature, we obtained a SE rate of 5% per course of amoxicillin

At a baseline ABS prevalence of 60%, Emp-Tx yielded 1 9 fewer Sx-days compared to Sx-Tx, and 0 1 fewer Sx-days compared to XR-Tx Sensitivity analysis of ABS prevalence revealed that the benefit of Emp-Tx maximized at 3 2 extra Sx-free days over Sx-Tx at 100% ABS prevalence Only at prevalences greater than 33% were Sx-days lessened by at least one day by Emp-Tx Emp-Tx was more effective than XR-Tx at prevalences greater than 10% and XR-Tx was only marginally more effective than Emp-Tx at lower prevalences Our findings were stable over a wide range of SE rates, non-ABS cure rates, and X-ray sensitivities and specificities The benefit in Sx-free days of Emp-Tx over Sx-Tx rose as the cure rate for amoxicillin was increased over placebo from 0 9 fewer Sx-days to 5 2 fewer Sx-days (at 100% prevalence)

While we found that treatment of suspected ABS with amoxicillin (Emp-Tx) is more effective than Sx-Tx over the full range of ABS prevalences, Emp-Tx is a toss-up compared to Sx-Tx up to about 30% prevalence At best, XR-Tx offers only marginal benefit over Emp-Tx

After estimating the likelihood of ABS in a given patient, the patient and clinician should weigh the often limited value of treating with antibiotic against the societal harm of resistance due to antibiotic over-prescription While our analysis implies that obtaining an XR is never of value, a planned cost-effectiveness analysis is needed to clarify this issue

COST EFFECTIVENESS OF THE PNEUMOCOCCAL VACCINE IN THE HEALTHY YOUNG ADULT POPULATION

PV Pepper and DK Owens, VA Palo Alto Health Care System, Palo Alto, CA, and Stanford University, Stanford, CA

Streptococcus pneumoniae infection is the leading cause of bacterial pneumonia resulting in hospitalization in adults Currently, vaccination is recommended for the elderly and immunocompromised hosts, yet pneumococcal pneumonia has an impact on the young adult population We therefore evaluated the cost-effectiveness of the pneumococcal vaccine (Pneumovax) for administration to 22-year old adults

We developed an 8-state Markov model to evaluate the cost effectiveness of the Pneumovax relative to no vaccine in healthy young adults The model incorporated hospitalization and death from pneumonia, age-specific mortality, quality-of-life adjustments, and costs for vaccination and treatment We estimated the incidence of pneumococcal pneumonia at 1 5 cases/1000 persons/year based on published literature We determined cost estimates from published literature of reimbursement data We assumed a vaccine efficacy of 70 percent and lifetime protection from vaccination at age 22

Administration of the Pneumovax resulted in a gain of 1 9 quality-adjusted days and cost savings of $54 05 per person during the lifetime of the subjects Our estimates of change in QALYs were sensitive to the quality-adjustment for treatment and hospitalization, the incidence of pneumonia, and the cost of the pneumococcal vaccine Vaccination remains the dominant strategy with an efficacy as low as 30 percent and a cost per vaccination twice that of our base case ($48) In addition, vaccination is cost-saving until the pneumonia incidence is less than 0 5 cases/1000 persons/year At an incidence of 0 25 cases/1000 people/year, vaccination spends $13,000 per QALY gained Benefits of vaccination are reduced if vaccine protection is not lifelong

Vaccination for S pneumoniae in the healthy young adult population would reduce overall expenditures and produce modest long-term health benefits, assuming lifelong protection A vaccine program should evaluate vaccine efficacy, the frequency of severe adverse effects, and the duration of protection in young adults Based on currently available data, routine vaccination of healthy young adults is warranted

A FEASIBILITY STUDY OF KNOWLEDGE BASED SYSTEMS FOR MEDICAL REHABILITATION

MDonald DScuse The University of Manitoba, Winnipeg, Manitoba

The purpose of this study was to investigate the feasibility of applying knowledge based system applications (KBS) to support clinical decision making in physiotherapy The primary goal was to determine if physiotherapy expertise could be implemented into a computer program A secondary goal was to gauge the acceptance of KBS when integrated into a clinical setting If feasible, KBS could help physiotherapists to make better clinical decisions by efficiently accessing case relevant expertise and knowledge

To study the primary objective, case management of post-poliomyelitis sequelae (PPS)was selected as the problem to be addressed by the KBS application Knowledge extracted from medical journals and interviews was encoded with rapid prototyping and the Expert-2 shell The resultant application was tested with 4 real cases The knowledge base was validated by 3 physiotherapists with relevant expertise Physiotherapists and students interacted with the first prototype, were tested and evaluated the interface The knowledge base was ported to Level 5 Object for refinement To study the secondary goal, questionnaires were distributed to senior year physiotherapy students and physiotherapists Items included attitudes towards using computers for decision support, and concerns and perceived needs regarding KBS

The knowledge base was deemed comprehensive and accurate by the physiotherapy experts The KBS treatment recommendations for all 4 real cases were comparable with those made by the physiotherapists who had managed the cases Using the KBS, practicing physiotherapists were able to find the applicable information 100% of the time and students 82% of the time Practicing physiotherapists (100%) found the program difficult to use, however students found it easy to use (56%) Regarding the questionnaires, 1/3 of the students believed that KBS would be especially helpful for students/new graduates and 45% that they would be an efficient tool to access information 100% of practicing physiotherapists who used the KBS felt that such programs could benefit all physiotherapists Other physiotherapists expressed concerns regarding job security decision making autonomy, and computer training costs

Physiotherapy expertise for case management of PPS can be encoded into a functional computer program More research is needed to determine how to facilitate KBS acceptance and use in the clinic

VALIDATION OF A NOMOGRAM FOR PREDICTING DISEASE RECURRENCE FOLLOWING RADICAL PROSTATECTOMY FOR CLINICALLY LOCALIZED PROSTATE CANCER

MW Kattan, M Witte, RA Morton, T Cangiano, J deKernion, JA Eastham, F Meyer, Y Fradet Baylor College of Medicine, Houston, UCLA, Los Angeles, LSU, Shreveport, l'Universite Laval, Quebec City

Over 180,000 men are expected to be diagnosed with prostate cancer during 1998, and 40,000 men will die from the disease this year The most common aggressive therapy for clinically localized prostate cancer is radical prostatectomy However, radical prostatectomy fails to prevent disease progression in 25-50% of men who receive it Because radical prostatectomy can result in substantial morbidity such as impotence and incontinence, clinicians and patients would like personalized estimates of the probability that radical prostatectomy will be effective prior to performing the operation

A nomogram was recently published (JNCI, 90(10) 766-71, 1998) which claimed to predict the probability of cancer recurrence following radical prostatectomy by using the patient's pre-treatment PSA, biopsy Gleason grade, and clinical stage The authors performed some internal validation analyses using their data from Baylor College of Medicine and presented the confidence intervals for the nomogram's predictions However, the authors of the nomogram failed to test it on data from another institution The purpose of this study was to apply the nomogram on patients from several different institutions and determine whether the nomogram was as accurate as that claimed by the authors

Baseline and follow-up patient data were obtained from Quebec City (N=249), UCLA (N=424), and several VA institutions (N=592) Each patient had his probability of recurrence computed using the published nomogram, and the predictions were compared with actual outcome by computing a concordance index (the same measure presented in the JNCI publication) This measure is the fraction of patient pairs where the patient with the worse prognosis recurs prior to the other patient

Application of the nomogram to the data from Quebec City, UCLA, and the VAs produced concordance indices (0 61, 0 68, and 0 64, respectively) apparently lower than that reported by the authors in their original publication (0 74), but the 95% confidence intervals of these indices (0 50-0 75, 0 55-0 81, and 0 53-0 75, respectively) overlap the original index that was reported We conclude that the JNCI prostate cancer nomogram may predict as accurately as previously reported, though larger series and/or longer patient follow-up is needed to achieve greater statistical power The nomogram may ultimately prove valuable in decision making for patients who are otherwise candidates for radical prostatectomy

PATIENT AND POPULATION PREFERENCES FOR OUTCOMES IN PROSTATE CANCER

Krahn M, Ritvo P, Naglie G, Irvine J, Bezjak A, Trachtenberg J University of Toronto

Background Modelling studies have shown that outcomes in prostate cancer (PC) (sexual and urinary dysfunction) are key variables in clinical policy regarding screening and treatment

Methods We measured preferences in 177 ambulatory PC patients (45 with metastatic disease), from the practices of urologists, radiation oncologists, and medical oncologists at two tertiary treatment centers Population-weighted preferences were elicited using the Health Utilities Index (HUI) and the Quality of Well Being (QWB) Scale Standard gamble (SG) and rating scale (RS) preferences for patients' own health state were elicited using a disease specific utility instrument (PORPUS), which includes a multiattribute health state classification system for PC, a structured interview protocol, SG and RS (feeling thermometer) props, trained interviewers, and standardized clinical marker states

Results Mean utilities for the entire group were directly elicited SG 84 9 (95% C I 82 9-87 7), RS 78 2 (77 1-83 3), HUI 80 2 (77 1-83 3), QWB 63 3 (60 6-66 1) For most subgroups, SG-weighted preference scores (SG, HUI) were higher than rating scale (RS, QWB) scores, and directly elicited preferences (SG, RS) were higher than population-weighted preferences (HUI, QWB) Preferences were not predicted by age, educational level, or other demographic variables There was no difference among subgroups with previous radical prostatectomy, radiation therapy, or hormonal therapy, but small differences were observed in patients with and without metastatic disease (SG 86 4/80 5, RS 79 0/72 8) Statistically significant differences were detected across the range of sexual dysfunction (range 13 2, SG) and urinary control (range 16 8, RS) No differences were observed for sexual interest, urinary frequency/urgency, or bowel function A multivariable analysis adjusting for 8 non-sexual attributes showed a mean SG and RS utility difference across the range of sexual dysfunction of 6 0 (SE 7 0) and 2 7 (SE 5 1) Across the range of sexual interest, the difference was SG 2 2 (SE 3 9) and FT 2 6 (SE 2 9) Differences of similar magnitude were detected for urinary and bowel dysfunction

Conclusions: Directly elicited RS and SG scores are higher than scores from generic instruments When preferences are derived from patients, decrements in health state preferences for PC outcomes are much smaller than estimates reported in the literature for scenario-based utilities Utilities for policy models for screening and treatment of prostate cancer should be estimated directly from patients, and should consider outcomes in the context of existing co-morbidity

FEASIBILITY AND COST-REDUCTION ANALYSIS OF AN EMERGENT/URGENT TELE-CARDIOLOGY CONSULTATION PROGRAM EMPLOYING TELE-ECHOCARDIOGRAPHY

JB Perlin, KO Akosah, DC Butler, WG Harman. Virginia Commonwealth University, Richmond, VA, McGuire Veterans Affairs Medical Center, Richmond, VA.

Purpose To provide a feasibility, economic, and educational needs analysis of implementing a tele-cardiology consultation program (TCP) employing transmission of echocardiographic images to provide emergent and urgent specialty consultation from the Richmond VA Medical Center to referring generalist clinicians at the Beckley VA Medical Center approximately 250 miles away

Methods Case-series and cost reduction analysis using historical controls Outcome variables included 1) adequacy of transmitted echocardiography, 2) comparison of health system resources utilized by TCP versus consultation by traditional telephone referral and subsequent patient transfer, 3) disparity in perception of urgency between referring clinicians pre-consultation and consulting clinicians post-consultation, 4) frequency of diagnoses for which consultations were requested, and 5) relationship between disparity in perceived urgency of consultation and diagnostic categories Referring physician intent to request urgent (<24hrs) or emergent (<4hrs) consult was the single criterion for patient selection

Results Twenty-three consultations were requested during the study Pre-consult intentions would have resulted in emergent and urgent transfers costing $77,500 TCP consultation cost $5,463 to provide and resulted in transfers costing $15,500 for a cumulative savings of $56,537 ($p<0.001$, paired t-test) Congestive heart failure (CHF) was the diagnosis for which consultation was most frequently sought as well as the diagnostic category with greatest disparity in pre- and post-consult assessment of urgency The pre-consult "overestimation" of urgency was significantly different ($p< 0.001$, Mann-Whitney) for CHF than for the diagnoses of arrhythmia, coronary artery disease, endocarditis/thrombus, and valvular/anatomic lesions, but not significantly different than for aortic aneurysms or pericardial effusion

Conclusions This study suggests that an urgent/emergent TCP using tele-echocardiography is technically feasible, can reduce costs by avoiding unnecessary transfer, and may identify educational opportunities and facilitate improved management and communication between referring and consulting clinicians Diagnoses with greater disparity in pre- and post-consult urgency may reflect the relative comfort (or discomfort) with specific diagnoses at the referral site, or the fact that echocardiography is a diagnostic modality particularly useful in the management of CHF

UTILITY OF SCREENING FOR LEIDEN FACTOR V AND PROLONGED PROPHYLAXIS. A DECISION ANALYSIS

M Marchetti, G Barosi
Laboratory of Medical Informatics I R C C S Policlinico S Matteo, Pavia Italy

Leiden mutation of coagulation factor V is highly prevalent in patients with venous thromboembolism (VTE) and is associated with double to fourfold the risk of recurrent events Nevertheless, management of carriers is still debated and screening for the mutation is recommended only for selected patients Five European experts in the field of Thrombosis could not reach a consensus on the target population for screening

We, therefore, built a Decision Analysis comparing the effectiveness on quality-adjusted life expectancy (QALE) of current practice with a screening program, followed by prolonged prophylaxis of the mutation carriers A four-states Markov model was framed to account for the natural history of patients with a first episode of VTE and for Leiden mutation of factor V, it included well-being, chronic anticoagulation, disability after a hemorragic stroke and death

The analysis showed that screening added patients younger than 40 years up to three quality-adjusted months of life, but supplemented less than one quality-adjusted month to patients older than 40 Screening outcomes were most sensitive to the risk of recurrence among Leiden carriers, consequently screening benefit was higher for the subset of patients presenting with idiopathic VTE patients younger than 40 years gained up to 5 quality-adjusted months of life from screening On the opposite patients with secondary VTE would get at most one quality-adjusted month Young carriers of the mutation would get up to 14 quality-adjusted months of life if a two-years long prophylaxis with warfarin were initiated soon after screening This benefit is even higher if the patients presented with an idiopathic episode of VTE 21 quality-adjusted months Moreover, even older patients found carriers by screening would prolong their quality-adjusted expected life longer than one month if following a prolonged prophylaxis

Screening for Leiden mutation factor V and prolonged anticoagulation of carriers of venous thromboembolism is advantageous especially for patients younger than 40 years or with idiopathic thromboembolism

RANDOM VS FIXED ATTRIBUTE ORDERING IN PREFERENCE ASSESSMENT COMPARING THE VISUAL ANALOGUE SCALE TO THE STANDARD GAMBLE

JJ Gonzales, K Weinfurt, K Bajwa, L Cai, K Schulman, Departments of Psychiatry and Medicine, Clinical Economics Research Unit, Georgetown University Medical Center

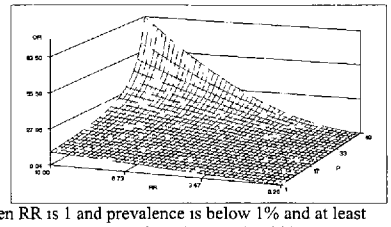
Background and Purpose: The validation of health state descriptions used in preference assessment can be affected by design issues such as framing and format presentation This study examined whether attribute order—fixed vs random—has an impact on preferences using visual analogue (VA) ratings **Methods**: Using a within-subjects design, 68 volunteer subjects rated a set of 5 health states (holistic presentation) with attributes in a fixed order and a set of 5 states with attributes in random order The 5 health states described different major depression scenarios, but the 2 sets were identical except for attribute ordering Order of set presentation was counterbalanced Subjects rated both sets using health state rank ordering, VA and SG tasks **Analysis** VA ratings were transformed to 0 - 1 metrics Each respondent's profiles of health state ratings under fixed and random conditions were compared along three dimensions *Level* the average rating, *scatter*, or variability in ratings, and *shape*, the ranks of the five ratings Agreement for level and scatter was calculated as a difference score, and agreement in shape was measured by Spearman's *rho* **Results**: The agreement between profile ratings under fixed versus random presentation varied across patients for level (range - 12 - 22), scatter (range - 33 - 33), and shape (range 29 - 1 00) The average difference between the levels (02, *SD* = 06) was marginally different from 0 ($p< 10$), suggesting that ratings were higher in the fixed than the random condition The average difference in scatter between conditions was 0 (*SD* = 11), suggesting no systematic difference between conditions Finally, the median correlation was 98, indicating that the relative ordering of the health states was highly consistent in both conditions **Conclusion**: Random versus ordered attribute presentation in health state descriptions does not impact VA ratings, except for producing small differences in the average rating given This is in contrast to earlier work with SG measures showing greater variability in agreement across respondents and lower correlations between profiles of ratings These findings suggest that the VA may be a more robust measure than the SG across different research conditions

APPROXIMATING ODDS RATIO WITH RISK RATIO CLINICAL IMPLICATIONS AND AFFECTING FACTORS

AM Mutgi, SA Khuder Medical College of Ohio, Toledo, OH

Published case-control studies in clinical medicine frequently approximate the (OR) odds ratio (association) with (RR) risk ratio (risk magnitude) This assumption has been made for rare diseases without defining "rare " We conducted series of simulations to evaluate the assumption of OR= RR at different prevalences and risk ratios

Methods Assuming a perfect case-control study that follows a cohort distribution we developed series of 2x2 tables with prevalences varying from 1 to 50% and risk ratio from 0 2 to 10 We calculated OR for a given RR and plotted against prevalence and RR We assessed its impact on a recently published case-control study evaluating the effects of smoking on breast cancer in families with BRCA 1 and 2 gene mutation Calculated OR was 0 46 in smokers with estimated lifetime breast cancer prevalence of 90% and the study concluded benefit from smoking



Results OR were close to RR at prevalence of 1% OR consistently overestimated RR and this was significant at a prevalence of 10% and above OR approximated to RR when RR = 1 but overestimated RR above and below 1 In the given study the calculated RR would be 0 93 suggesting no protection from smoking

Conclusion OR approximates RR when RR is 1 and prevalence is below 1% and at least under 10% Both the magnitude of the OR and prevalence of the disease should be considered when extrapolating OR to RR

COST-EFFECTIVE MANAGEMENT OF NONPALPABLE BREAST ABNORMALITIES USE OF A COMPUTERIZED MODEL TO DETERMINE NEED AND TYPE OF BREAST BIOPSY

Richard K Orr, Marshfield Clinic, Marshfield, WI

Introduction Because the costs of surgical biopsies represent a significant proportion of the total cost of mammographic screening, it is desirable to select appropriate mammographic abnormalities for surgical procedures In lieu of biopsying all lesions, various surgeons have advocated careful clinical stratification based on defined mammographic criteria, and stereotaxic core biopsy (SCB) has been proposed as a noninvasive, inexpensive test which may be a cost-effective alternative to surgical biopsy (Hillner Acad Radiol 1996 3 351) Methods Using a modification of Hillner's decision analysis model we calculated the cost-effectiveness of 4 different management strategies surgical biopsy of all patients (BX), SCB for all patients, a clinical stratification strategy , and a computer based strategy In the clinical strategy abnormalities were classified as to low, medium, or high risk by a surgeon (based on written criteria) The computer based strategy used a previously validated neural network which included the surgeon's impression, patient age, type of lesion, and change from prior mammograms as input variables For clinical and computer strategies, low risk patients were observed with follow-up mammograms, medium risk patients received SCB, and high risk patients received BX For calculations, an actual series of mammographic referrals (1312 women, prevalence of cancer = 14 1%) was utilized

Results Areas under ROC curves for the clinical strategy (surgeon's impression alone) and the computer strategy were 0 84 and 0 91 respectively, indicating good discrimination in both cases Biopsying all patients missed no cancers, but was expensive The computer strategy represented a practical compromise as the additional cost of open biopsy per missed invasive cancer was $227,000

	Cost($) / pt	SCB / 1000 pts	BX / 1000 pts	Missed invasive Ca/ 1000 pts	Missed in situ ca / 1000 pts
Clinical	1554	330	251	29	11
SCB (all pts)	1866	1000	186	11	4
Computer	1976	450	346	4	1
Open Bx (all)	2847	0	1000	0	0

Conclusion Using a computer model to aid clinical decision making is a cost-effective way of decreasing unnecessary breast biopsies for mammographically detected abnormalities

ACCESS TO THE ICU UTILIZATION OF COMPUTER MODELING FOR STRATEGIC PLANNING.

W Cahill, J Tsevat, ML Render Veterans Administration Medical Center – Cincinnati and Pulmonary Critical Care Div, Dept of Neurology and Medicine, University of Cincinnati College of Medicine

Introduction: Determining appropriate numbers of critical care beds balances patient care needs against cost Insufficient ICU beds leads to diversion of patients to other facilities The purpose of the study was to predict the most efficient bed allocation to assure appropriate ICU bed availability at our facility

Methods: Using dynamic simulation software (Arena), we analyzed ICU bed availability with changes in capacity (ICU beds/telemetry) Dynamic simulation imitates a system using a probabilistic computer model to give an approximate answer to an exact problem Parameters of the model were determined by data collected over 1 year at our facility These included number of admissions per year to the ICU, the telemetry unit (TELE) and the wards (Floor), average length of stay (LOS) in each unit, and rates of transfer between units Three scenarios were examined based on existing beds, planned construction of additional telemetry and ICU beds, and construction of a Heart ER and a Respiratory Step-down unit (RCU) to decrease ICU admissions and LOS Ten 40-day simulation were performed

Results The model predicts inadequate ICU bed availability using the present bed configuration Diversion of patients was reduced by adding TELE and ICU beds and eliminated by adding Heart ER and RCU beds

Model Assumptions	*Present*	*Add TELE/ICU*	*Heart ER/RCU*
# ICU/TELE/Floor beds	10/15/70	16/24/70	20/24/70
ICU LOS (ave /max)	3/10	3/10	3/5
Results			
Ave #ICU patients diverted/40days	29 6	3 4	0
Ave bed occupancy ICU/TELE/floor (%)	100/75/76	96/51/81	52/52/82
Max bed occupancy ICU/TELE/floor (%)	100/100/92	100/81/98	85/88/97

Conclusions: The most effective allocation of resources to avoid diversion of critically ill patients at our facility is an increase in beds with a reduction in length of ICU stay Simulation models such as this more efficiently promote strategic decision making for medical facilities, identify potential limitations of strategic plans, and allow testing of alternative solutions

STENT PLACEMENT COMPARED WITH PTA FOR ILIAC ARTERY OCCLUSIVE DISEASE A COST-EFFECTIVENESS ANALYSIS PERFORMED IN THE UNITED STATES AND THE NETHERLANDS

JL Bosch, C Haaring, MF Meyerovitz, KA Cullen, WPThM Mali, MGM Hunink Erasmus University Rotterdam and University of Utrecht, The Netherlands, Keio University, Tokyo, Japan, and Brigham and Women's Hospital and Harvard School of Public Health, Boston, MA

Purpose To evaluate costs of percutaneous transluminal angioplasty and stent placement for iliac artery occlusive disease from various perspectives in the United States and in the Netherlands, and to evaluate the cost effectiveness of these treatment strategies in the United States and the Netherlands

Methods Cost identification was performed in a teaching hospital in the Netherlands and the US from the perspective of the interventional radiology department, hospital, patient, and the society Effectiveness data, including patency results, risk reduction of failure, complication and mortality rates and quality-of-life data were derived from a randomized controlled clinical trial and a meta analysis A cost-effectiveness analysis was performed from the societal perspective Using a Markov decision model a hypothetical cohort of patients with life-style limiting claudication caused by an iliac artery stenosis for who a percutaneous intervention was indicated was followed over time

Results Costs assessed from the various perspective were significantly different ($p<0$ 005) Total procedure cost were lower in the Dutch setting than the US setting (PTA $ 2 566 (SD $ 833) and $ 4 866 (SD $ 1 560), PTA+STENT $ 4 402 (SD $ 848) and $ 6 332 (SD $ 1 573) in the Netherlands and US respectively) In a sensitivity analysis assuming an equivalent duration of the procedure, the departmental costs were very similar across countries In both countries angioplasty with selective stent placement was a cost-effective treatment strategy (i e , incremental cost-effectiveness ratio < $ 20 000)

Conclusion Cost analyses may yield different results depending on the chosen perspective and country Comparison of the various cost components between countries may identify cost-saving strategies In both the US and Dutch setting the results suggest that in the treatment of intermittent claudication angioplasty with selective stent placement is a cost-effective treatment strategy compared with angioplasty alone

AGREEMENT BETWEEN PATIENT AND PROXY RESPONSES TO THE FUNCTIONAL STATUS QUESTIONNAIRE

KP Weinfurt, L Cai, S Trucco, RJ Willke, KA Schulman Clinical Economics Research Unit, Georgetown University Medical Center, Washington, DC , Pharmacia & Upjohn, Kalamazoo, MI

Background When patients cannot provide responses to health-related quality of life (HRQOL) measures, family or friends may be asked to respond We sought to examine agreement between head injury patients and their proxy responders, and the relationship between agreement and injury severity **Method** 475 head injury patients in a multi-national clinical trial of tirilazad mesylate and their proxies completed the Functional Status Questionnaire (FSQ) at six months post-randomization **Analysis** Each respondent's ratings on five FSQ scales (0 - 100, 100 = Best) comprised their "profile" Patient and proxy profiles were compared along three dimensions *Level*, the average ratings score, *scatter*, or variability in ratings, and *shape*, the ranks of the five ratings Agreement for level and scatter was calculated as a difference score, and agreement in shape was measured by Pearson's *r* Finally, we evaluated agreement covariation with the Glasgow Outcome Severity rating **Results** Patient and proxy agreement varied significantly for ratings of HRQOL in terms of level (range -50 62 - 60 4), scatter (range 22 83 - 32 97), and shape (-0 73 - 1 00) There was a significant ($p<$ 0001) tendency for patients to have higher levels than their proxies (mean difference = 1 68, *SD* = 9 41) Mean difference in scatter was -0 95 (*SD* = 7 67), reflecting a significant ($p<$ 01) tendency toward proxies having more scatter than patients The median *r* of 88 indicated good agreement in the shapes of the profiles Greater severity of head injury was related to greater disagreement in both the level ($p<$ 01) and scatter ($p<$ 05) of profiles Agreement in profile shape was not related to severity To summarize, proxies agreed with patients in reporting which domains of HRQOL were impaired relative to other domains However, disagreement in actual ratings assigned was due to proxies providing lower more variable ratings for the profile of FSQ subscales Such disagreement was greatest for patients with more severe head injury **Conclusion** This study illustrates the benefit of examining different components of agreement via psychometric profile similarity analysis and provides some insight into the ways in which proxy responses are appropriate for various applications

NEONATAL CIRCUMCISION A COST-UTILITY ANALYSIS

RS Van Howe Marshfield Clinic -- Lakeland Center, Minocqua, Wisconsin

Purpose In the seven years since cost-utility analysis has been applied to neonatal circumcision, several studies have been published that drastically alter the data on which the previous calculations relied A new cost-utility analysis was undertaken utilizing this new information

Method A cost-utility analysis between boys circumcised at birth and those whose prepuce was left intact was undertaken Financial impact was measured in full-value dollars and dollars discounted at inflation rates of 3% and 5% The effect of morbidity was calculated using the General Health Policy Model and Quality of Well-being scale

Results When compared to leaving the prepuce alone, neonatal circumcision and its sequelae are cost-ineffective, costing $488 (5% discounted dollars) more per patient over a lifetime Neonatal circumcision and its sequelae also have an overall negative impact on health when compared to those not circumcised, resulting in a difference of 29 83 well-years lost per 1,000 people The financial costs combined with the well-years lost indicate that neonatal circumcision is not good health policy Even if the initial procedure of neonatal circumcision was cost-free and complication-free, its sequelae are more costly than leaving the prepuce alone To make neonatal circumcision cost-effective noncircumcised males would need to undergo a medically indicated circumcision at rate of 39% (currently 1-2%), have urinary tract infections at a rate of 23% (currently 1%), develop penile cancer at a rate of 39% (currently 0 08%), or contract AIDS at a rate of 1 6% (currently 0 014%) Despite the great variability in the epidemiological data used for the basis of the calculations, it is impossible to manipulate the data to make neonatal circumcision cost-effective

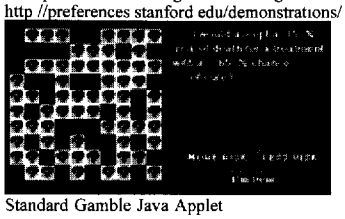
Conclusions While neonatal circumcision has never been conclusively proven to have medical benefit, it still persists as the most commonly performed surgery in the United States With the marked cost-inefficiency, the overall negative impact on health, the possible lack of consent, and the violation of the child's right to choose, support of neonatal circumcision as a medical procedure can no longer be justified and should be discouraged It would be more cost-effective to pay physicians $200 to $300 not to perform neonatal circumcisions than continue the current reimbursement for the procedure

PUBLIC WWW TOOLS FOR PREFERENCE ELICITATION

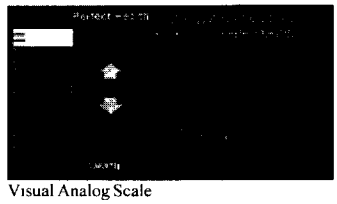
CM Chong, LA Lenert, Stanford University, Stanford, California

The World Wide Web (WWW) is a rapidly growing medium that can be used both for research in preferences and for physician and patient decision support While it is relatively easy to create descriptions of health states using pictures, text, and sounds, one factor inhibiting dissemination of preference elicitation methods over the WWW is the difficulty in construction of tools for assessment of preferences

Java applets are small computer programs that can be embedded WWW pages We developed a publicly accessible suite of Java applets for measurement of health values that other investigators can access and integrate into their own WWW pages These platform-independent Java applets are compatible with Netscape Navigator or Microsoft Internet Explorer Use of Java applets in Web pages requires only minimal hypertext markup language (HTML) authoring skills Applets support display of custom scale labels, custom search patterns for indifference points and can save results either by email or to databases Applets can be added to HTML pages by referencing their universal resource locator (URL) on our server Current Java applets include a standard gamble, single visual analog scales, and multiple communicating visual analog scales Software tools can be accessed at http //preferences stanford edu/demonstrations/



Standard Gamble Java Applet



Visual Analog Scale

REDUCING RESOURCE-UTILIZATION IN PATIENTS WITH DYSPEPSIA AND/OR UNEXPLAINED UPPER ABDOMINAL PAIN: EARLY GASTROSCOPY WITH THE ADDED INFORMATION OF ENDOSCOPIC ULTRASOUND MAY REDUCE COSTS (AS OPPOSED TO GASTROSCOPY ALONE)

AV Sahai, I Penman, D Williams, W Knapple, G Mishra, A vanVelse, RH Hawes, and BJ Hoffman Digestive Disease Center, Medical University of South Carolina, Charleston, SC, USA PURPOSE To determine if the endoscopic and ultrasound capabilities of endoscopic ultrasound [EUS] could reduce costs by replacing gastroscopy [EGD] and additional abdominal imaging tests (upper GI barium studies [UGI], transabdominal ultrasound [US] abdominal CT [CT]) in patients with dyspepsia and/or unexplained upper abdominal pain [D&/orUAP] METHODS D&/orUAP patients undergoing EUS were interviewed to determine dyspepsia type, pain site, symptom duration [ΣD], and the number of supplemental negative or non-contributory abdominal imaging tests completed in addition to EGD EUS and EGD were performed using the Olympus EU-M20 radial scanner by operators aware of the history, but blinded to previous test results Potential cost savings were calculated using facility direct and indirect costs EGD $96, UGI $165, US $141, CT $238, MRI $376, and EUS $338 RESULTS 149 patients with D&/orUAP studied to date (18% pain only, 3% dyspepsia only, 79% both) (study ongoing) Mean age 47 3+/-14 1, 72% female, 72% white, 52% status post-cholecystectomy Mean ΣD 14 3+/-14 6 months (median 9 months) Previous testing EGD 113(76 9%), UGI 49(33 3%), CT 118(80 3%), and US 102(69 4%), and MRI 11(7 5%) Mean number of imaging tests per patient 2 5+/-1 3 (5 tests 2%, 4 tests 22%, 3 tests 34%, 2 tests 24%, 1 tests 10%, 0 tests 9%) There was no correlation between ΣD and the number of tests (r=0 015) EGD was performed using the echoendoscope in 62(42%) Adequate visualization of the distal esophagus, gastric cavity, and duodenum was obtained in all cases Endoscopic findings gastric ulcer 3%, erosive gastritis 10%, erosive esophagitis 8%, food retention 3%, non-erosive gastritis 19% "other" 10%, and normal 48% Adequate ultrasound visualization of the pancreas, and gallbladder was obtained in all cases Mean proportion of common bile duct visualized 93 3+/-18% Clinically-relevant EUS findings gallstones 3%, sludge 4%, chronic pancreatitis (≥5 EUS criteria) 32%, and "equivocal" for chronic pancreatitis (3-4 EUS criteria) 23% Estimated avoidable costs by EUS total hospital cost (147 pts) $11713 or $79/patient CONCLUSIONS 1) EUS provides a comprehensive endoscopic and ultrasound examination of the upper gastrointestinal tract and surrounding structures in one sitting 2) Patients with dyspepsia and/or unexplained abdominal pain frequently undergo multiple negative or noncontributory abdominal imaging procedures -- regardless of the duration of symptoms 3) On a large scale, early EUS (instead of gastroscopy alone) could reduce hospital costs significantly, since dyspepsia and unexplained abdominal pain are extremely common problems

RECURRENT DEPRESSION IN POOR, YOUNG WOMEN A DECISION ANALYSIS OF ACUTE PHASE PHARMACOTHERAPY

ME Singer Case Western Reserve University at MetroHealth Medical Center, Cleveland, Ohio

Background: Poor, young women are at particularly high risk for depression Medicaid reimbursements for antidepressants nearly quadrupled from 1992-1996 Recent studies suggest high rates of unreported sexual dysfunction in patients taking selective serotonin reuptake inhibitors (SSRI) This could lead to poor compliance and an increased number of treatment dropouts

Purpose: To determine the drug strategy that maximizes the proportion of poor, young women completing acute phase pharmacotherapy for recurrent depression

Methods: A decision tree was created to represent the four most common drug strategies for depression, based on three classes of antidepressants tertiary tricyclic antidepressants(TCA), SSRIs, and heterocyclic antidepressants(HCA) Each strategy consisted of initial drug choice, and a second drug choice in the case of initial treatment failure The strategies were SSRI/HCA, SSRI/SSRI, TCA/SSRI and TCA/TCA Success was defined as three consecutive prescriptions(Rx's) for the same drug Patients who did not succeed with the first drug and did not take any other antidepressants were considered to be treatment dropouts All model parameters were based on race adjusted logistic regression models using Ohio Medicaid claims data for 7/91-6/96 Sensitivity analyses were performed for all parameters using 95% CI limits Patients had to be continuously enrolled from 1 year prior to 1 year following index Rx Included were females aged 25-44, AFDC eligible, with diagnosis of unipolar, recurrent depression <30 days prior to index Rx Patients with other mood disorders, or any antidepressant Rx in the last year were excluded

Results: Although patients taking SSRIs had better initial success than TCAs among non-dropouts(702 vs 550, p= 0005), they were far more likely to drop out of treatment(200 vs 058, p= 0001) Baseline results showed that TCA/SSRI(747) was the best strategy, followed by TCA/TCA(721), SSRI/SSRI(697) and SSRI/HCA(643) The model was sensitive to the rate of SSRI treatment dropout (SSRI/SSRI becomes optimal), and success of TCA and SSRI after initial TCA failure (TCA/TCA preferred)

Conclusions: In this population of poor, young women with recurrent depression, prescribing a TCA first maximized the proportion of patients completing acute phase treatment due to fewer treatment dropouts Pharmacotherapy selection must be based on strategies, and not solely on the success of the initial drug

IS BEDSIDE BAYES' GOOD ENOUGH OVER A BROAD SPECTRUM OF PATIENTS A SIMULATION STUDY

BM Singal, E Somoza, P Succop VA Medical Center, Cincinnati, OH and University of Cincinnati, OH

It is now recognized that the sensitivity (sen) and specificity (sp) of a diagnostic test are not fixed but may vary with clinical spectrum Application of Bayes' rule to estimate the post-test probability of disease in an individual patient relies on the average sen and sp for the test over the whole clinical spectrum of diseased and non-diseased patients and the physician's estimate of the pretest probability of disease (prior) In practice, however, sen and sp are often determined in the face of other biases such as verification and case mix bias It is the purpose of this simulation to determine how well Bayes' theorem operates over the clinical spectrum of patients when the average sen and sp are determined without other biases

This simulation is based on a real, albeit simplified, clinical scenario where asymptomatic men ages 50 to 80 are screened for prostate cancer using prostate specific antigen (psa) Patient age is related to prevalence of disease and determines the prior Log psa level is independently related to three deciles of age The parameters used were based on studies published in the medical literature Sen and sp for the standard cutoff of 4ng/ml were determined to be 85 and 90 via a simulation of a case control study The sen varied between 83 and 88 and the sp between 97 and 77 for the extremes of age Disease prevalence was set at 05

In this scenario, Bayes' theorem predicted the probability of disease in each group within a few percentage points except in the test positive, low decile group (prior = 01), where Bayes' underestimated the probability of disease (08 vs 19) and the test positive, high decile group (prior = 1) where it overestimated (49 vs 29) Similar findings were noted when disease prevalence was changed to 1

In this simulation, Bedside Bayes' operated reasonably well over a broad spectrum, approximating the magnitude of the change in probability given the test result

MEASURING THE FINANCIAL IMPACT OF SPONSORED TECHNOLOGY ASSESSMENT ON CORPORATE VALUATION

FJ Papatheofanis University of California San Diego, La Jolla, CA

Technology-oriented, problem-oriented, and project-oriented health technology assessment [HTA] is sponsored by a myriad of public and private sector organizations The emphasis of HTA reflects the needs or purposes of the sponsoring organization Undoubtedly, however, the results of HTA overlap, and seemingly disparate organizations leverage the conclusions of HTA for potentially unintended aims This investigation measured the short [1-3 wk]- and long [≥ 12 wk]-term impact of negative and positive sponsored HTA decisions on respective corporate valuation when calibrated by fluctuations in common stock pricing [publicly-traded US corporations]

Source organizations monitored by the University HealthSystem Consortium were queried to identify HTA projects released between 1993 to 1997 These HTA were primarily produced by payer [BlueCross BlueShield TEC, etc] and provider [American College of Physicians, etc] organizations The stock market valuation of corporations producing the devices undergoing HTA were tracked for 12-18 weeks before and after the release of HTA conclusions Market variables and discounting were adjusted,and, marketing and promotional factors were adjusted

Monitored organizations produced 442 HTA during the study interval Payer-sponsored negative assessments resulted in short- and long term decreased [p< 05] corporate valuation in 57% [< 001] whereas positive decisions resulted in increased [< 05] valuation in 32% [< 01] of cases The greatest decrease in valuation resulted in technologies which underwent payer-sponsored HTA in 1996-97 No significant stock price fluctuation resulted from provider-sponsored HTA decisions

These results demonstrate that conclusions from payer-sponsored HTA may have far reaching results beyond guideline formulation and coverage decisions Corporate valuation may increase or decrease as a result of such decisions This represents a serious departure from commonly understood intentions of HTA, and may represent an indirect mechanism for payer-triggered adoption or rejection of a new technology

WHAT INFORMATION DO PATIENTS NEED TO PARTICIPATE IN TREATMENT DECISIONS? IT DEPENDS ON HOW YOU ASK.

D Feldman-Stewart[1], MD Brundage[1], BA McConnell[1], RH Cosby[1], CRR Hayter[1], JC Nickel[2], WJ Mackillop[1] Radiation Oncology Research Unit[1], Dept of Urology[2], Queen's University, and Cancer Care Ontario

Patients with curable prostate cancer often have three treatment options (surgery, radiotherapy, or "watchful waiting") This study compared two methods for identifying the information items that such patients need in order to decide which treatment they prefer presenting items as unanswered questions (e g , Would I die from the radiotherapy?) or as facts (e g There is no chance that you would die from the radiotherapy) All 69 men diagnosed with early stage disease in a regional urology practice within an 18 month period were surveyed They were asked to consider the treatment decision for a specific hypothetical patient and to judge whether each of 59 items was "necessary" for them to know in order to make the decision Recipients were randomized to receive the items either as questions or as facts

Response rate was 78% (N=54) 30 in the question group and 24 in the fact group There were no significant differences between the groups with respect to mean age [t(48)=-1 54, p=0 13], proportion with high school education ($\chi^2_{(1)}$= 78, p=0 38), proportion receiving each of the three possible treatments ($\chi^2_{(1)}$=1 75, p=0 42), or proportion offered each treatment (all $\chi^2_{(1)}$ <1 05, p>0 3) All questions were judged "necessary" by at least 13% of respondents and all facts by at least 4% Considering responses overall, questions were rated as "necessary" more often than were facts (Kruskal-Wallis χ^2=4 14, p<0 05) even though the percentages of respondents judging each item as "necessary" were correlated between groups (r(57)=0 69, p<0 001) Considering each item individually, responses differed significantly between the two groups on 14 (24%) of the items (all p<0 05) 12 of the 14 were deemed "necessary" by more respondents when presented as questions than as facts Using 67% as a threshold for group consensus, the 2 groups agreed on 20 items as being necessary but did not agree on 16 the question group identified 13 items that the fact group did not and missed 3 of the items identified as necessary by the fact group

We conclude that investigating what information patients need by asking them to choose which questions they would like answered produces spurious results Decision aids should be constructed based on patients' judgments of the facts rather than of unanswered questions

CONFLICTS IN HEALTH VALUES BETWEEN PATIENTS AND PROVIDERS MAY IMPAIR THERAPEUTIC DECISION MAKING IN SCHIZOPHRENIA

LA Lenert, J Zeigler, T Lee, R Mahmoud, Stanford University School of Medicine Stanford CA, USA and the Janssen Research Foundation, Titusville, NJ

The objective of this study was to compare the health values of patients with schizophrenia, health care providers (HCP's), and family members of patients for different combinations of disease symptoms of schizophrenia and adverse drug effects and identify differences in values than may impede the optimization of therapy To study this issue, we developed a computer program that described health outcomes in schizophrenia and measured health values Using this program, we presented patients, HCP's, and family members with descriptions states with mild and moderate symptoms of schizophrenia depicted with and without pseudo-parkinsonism depicted using brief video presentations After each presentation we measured participants health preferences using a visual analog scale and a standard gamble Values were compared using analysis of variance methods for multiple observations after excluding persons with evidenced failure of the utility assessment process Study subjects were convenience sample of 148 Patients, 92 family members, and 99 HCP's drawn from 4 geographically and clinically diverse environments 39% of patients, 10% of family members of 0% of HCP's had evidence of failure of the utility elicitation process Failures were associated with educational level and, in patients, the level of thought disturbance In both visual analog scale and standard gamble ratings, HCP's placed greater weight on control of psychiatric symptoms than patients Patients valued freedom from extrapyramidal symptoms more highly than HCP's More over, patients preferred states with moderate disease symptoms over states with milder disease symptoms and a the extrapyramidal syndrome HCP's had the opposite preferences All comparisons were statistically significant at p values of less than 0 05 Family members had values intermediate between patients and HCP's There are meaningful differences between patients and HCP's in preferences for health outcomes in schizophrenia These differences may underlay problems with patients "compliance" with drug therapy

BUILDING UNCERTAINTY INTO COST-EFFECTIVENESS RANKINGS PORTFOLIO RISK-RETURN TRADE-OFFS AND IMPLICATIONS FOR DECISION RULES

BJ O'Brien, MJ Sculpher McMaster University, Hamilton, Ontario, Canada, Centre for Health Economics, University of York, Heslington, York, UK

Conventional decision rules for resource allocation using data from incremental cost-effectiveness analyses comparing independent programs indicate that the decision maker should rank-order programs according to their expected economic return (e g , QALYs per dollar) and fund programs in this order until the budget is exhausted But recent advances in the measurement and standardization of uncertainty for incremental cost-effectiveness ratios - particularly from trial-based studies with observed data on costs and effects - mean that decision-makers will increasingly be faced with two moments of the cost-effectiveness distribution -- return and risk (uncertainty) Characterizing such resource allocation as a (risky) investment problem we show how concepts of portfolio analysis from financial economics can be adopted as a conceptual framework for presenting cost-effectiveness data from multiple programs Two specific propositions emerge from this line of reasoning (1) portfolios will reflect cost-effectiveness risk-return preferences of health care payers and therefore will only be equivalent to rank-ordering by expected return under special circumstances, and (2) the degree of joint risk or covariation in cost-effectiveness between programs will create incentives to diversify an investment portfolio We present a numerical example of program portfolio diversification illustrating how the incentive to diversify is a function of program risk correlation Some interesting issues emerge from this conceptual framework first, what are the implications for decision rules in CE analysis if decision makers are asymmetrical in their attitude to risks in costs and in effects (ie the risk of budget over-run versus QALY shortfall)? Second, how might the equity characteristics of programs be reflected in portfolio analysis? We conclude that (1) the portfolio approach to thinking about cost-effectiveness is a logical extension of developments in the measurement of cost-effectiveness uncertainty, (2) it yields some interesting insights into the decision rules of cost-effectiveness under uncertainty and suggests rank-ordering by expected return as a special case of the general framework, and (3) it suggests an empirical research agenda on cost-effectiveness risk-return preference measurement and preferences decision makers have for cost-risk versus outcome-risk

COST-EFFECTIVENESS OF STENT VS PTCA IN PATIENTS WITH SINGLE-VESSEL CORONARY DISEASE IN JAPAN

S Ikeda, T Uemura, JL Bosch. Keio University School of Medicine, Tokyo, Japan

Purpose· To compare the cost-effectiveness of coronary stent implantation (CSI) to percutaneous transluminal coronary angioplasty (PTCA) in patients with single-vessel coronary disease in the Japanese clinical setting.

Method We used a computerized decision analytic model over a three year period to compare both the outcomes and costs of CSI vs. PTCA. Probabilities of the events were based on specific data from three hospitals and 1,095 patients in Japan when available. Other data sources (published data and expert opinions) were used when the specific data were not available Societal viewpoint was used for cost estimation. Discount rate of 3% was used for both outcomes and costs.

Results: The initial procedural costs for CSI were higher than PTCA, however this cost difference was decreased during the inhospital period due to avoided procedural complications in the CSI group and associated bail-out treatments in the PTCA group By the end of year one, the incremental cost of CSI was virtually recovered through savings in reduced revascularisation procedures in the CSI group, and the results were maintained over subsequent years. This favorable cost combined with a higher rate of event-free patients, also improved the cost-effectiveness of CSI, which has a lower cost per event-free survivor than PTCA at the 1, 2, and 3 year follow-up periods. The base-case results of cost per event free survivor at the end of three years were 6.24 million yen for PTCA and 2 79 million yen for CSI.

Conclusion. This model analysis which applied to the Japanese health care system and used the latest clinical and economic data, demonstrated CSI in single-vessel coronary disease to be clinically and economically superior to PTCA

COST-EFFECTIVENESS OF PRAVASTATIN IN PATIENTS WITH CORONARY HEART DISEASE AND AVERAGE CHOLESTEROL LEVELS

PP Glasziou, SE Mulray, JP Hall, A Martin, P Harris, P Thompson, A Tonkin, RJ Simes on behalf of the LIPID Study Group University of Queensland, University of Sydney, and the National Heart Foundation of Australia

Introduction The LIPID Trial is a double-blind, randomised placebo-controlled trial which commenced in 1990 and aimed to evaluate the long-term effects and cost-effectiveness of pravastatin in 9,014 patients with established coronary heart disease but average cholesterol levels (range 4 0-7 0 mmol/l)

Methods The design of the prospective economic evaluation (Cont Clin Trials, 1997 464-76) incorporated some novel features including the use of extensive sensitivity analysis on a preliminary model to determine optimal subsampling requirements

Results Interim analysis showed a 23% reduction in all-cause mortality (p<0 003) and the trial closed in late 1997 with an mean follow-up of 6 1years Hospital admissions and coronary revascularisation procedures were similarly reduced The trial included quality-of-life and cost-effectiveness sub-studies In a blinded analysis of all patients, average quality-of-life measured by the Rosser Index was high, 0 98 (on a 0-1 scale) with lack of energy, difficulty sleeping, and breathlessness being the most frequent causes of distress Pravastatin cost $22 6M but reduced hospitalisation costs by $4 3M, and medication costs by $1 9M Absolute reduction in all-cause mortality was 3 2% and, depending on the extrapolation of long-term survival the cost-per-life year saved is between $8-12,000

Conclusions Pravastatin after myocardial infarction or unstable angina and "normal" cholesterol levels reduces all-cause mortality and appears acceptably cost-effective Several important lessons were learnt in the design of prospective economic analysis including the need to consider sampling of "person-time" and the use of an internal life-table for extrapolation beyond the end of follow-up

EFFECT OF FUTURE COSTS ON THE COST-EFFECTIVENESS OF LIFE EXTENSION AND QUALITY OF LIFE IMPROVEMENT AMONG THE ELDERLY

D Meltzer[1],G Hazen[2],M Johanesson[4], I Abdalla[1],R Chang [2],A Elstein[3],A Schwartz[3] [1]U Chicago,[2]Northwestern U,[3]U Illinois,[4]Stockholm Sch Econ

Purpose The inclusion of future medical costs for "unrelated" illnesses and future non-medical costs in cost-effectiveness (C/E) analyses is controversial Recent theroretical work suggests that these costs should be included in C/E analyses and that failure to include them will bias analyses to favor interventions that extend life expectancy (LE) over interventions that improve quality of life (QOL) However, some analysts have been uncomfortable including future costs because they think that this would prevent all interventions in the elderly from being cost-effective To assess this, we contrast the effects of including future costs on the C/E of three common medical interventions among the elderly that differ in their effects on QOL versus LE hip replacement for osteoarthritis, treatment of hypertension, and treatment for prostate cancer

Methods We modified published decision and C/E analyses to include future medical and non-medical costs and calculated the cost per QALY Due to uncertainty about the C/E ratio based on current and related costs (CRC) of treating prostate cancer, we report those estimates contingent on CRC

Results

INTERVENTION	ΔLE	ΔQALY	ΔLE/ ΔQALY	ΔCost/ΔQALY w/o future costs	ΔCost/ΔQALY w/ future costs
Hip Replacement, women age 60	-0 03	6 9	-0 005	Cost-saving	Cost-saving
Hip Replacement, men age 85	-0 02	2 0	-0 01	$9,177	$9,042
Treatment Mild HTN, men age>70	0 06	0 05	1 25	$5,000	$32,000
Radiation Therapy MDPC, age 65	0 8	0 4	2	CRC	CRC+$32,000
Prostatectomy MDPC, age 65	0 7	0 2	3 5	CRC	CRC+$56,000

HTN=Hypertension MDPC=Moderately differentiated prostate cancer CRC=Current + related costs

Conclusions Including future costs does not prevent interventions for the elderly from being cost-effective However, including future costs does improve the C/E of interventions that improve QOL relative to those that increase LE Some interventions with small gains in QALY compared to LE may not be C/E even if CRC are zero Resource allocations based on analyses that exclude future costs may spend too much on interventions that extend life compared to interventions that improve QOL The effects of including future costs should be reported in all C/E analyses

INTERNATIONAL COMPARISON OF COST-EFFECTIVENESS IN DIAGNOSIS OF PULMONARY EMBOLISM

WB van den Hout, AR van Erkel and PMT Pattynama Dept of Medical Decision Making and Dept of Radiology and Nuclear Medicine, Leiden University Medical Center, The Netherlands

Purpose To estimate costs of diagnostic tests by questionnaire in different countries and to compare the cost-effectiveness of diagnostic strategies for suspected pulmonary embolism

Method A decision tree was used to compare 18 diagnostic strategies, consisting of combinations of conventional and spiral CT angiography, perfusion and ventilation scanning, ultrasound (for diagnosis of deep vein thrombosis) and D-dimer assay The disease and diagnostic characteristics were obtained from the literature and assumed to be identical for all countries National cost estimates were obtained from a cost questionnaire, sent to 13 centers in Europe and the USA A small fee was offered for returning the questionnaire In a top-down approach costs were calculated for different types of staff, equipment, material, housing, overhead and anticoagulant treatment Exchange rates were used to convert currencies to ECUs (1ECU=1 08$)

Results Questionnaires were returned from 6 centers in England, France, The Netherlands, Austria, Switzerland and the USA Dubious results were checked, but none proved incorrect Due to differences in wages, utilization of equipment and material, the calculated cost differences between countries were large The variation was smallest for ultrasound (45 to 85 ECU) and largest for anticoagulant treatment (1 700 to 22 000 ECU) However, the influence of these differences on the optimal diagnostic strategy was very limited The most effective strategy was ultrasound combined with spiral CT angiography It strongly dominated all other strategies, except for D-dimer strategies in The Netherlands The CE-ratio ranged from 8 000 ECU (5 000 £) per life in England to 68 000 ECU (73 000 $) per life in the USA The average life expectancy of these patients is about 25 years

Conclusions It is possible to obtain detailed data on costs using a questionnaire, provided the respondents are motivated and have sufficient economic training The differences in estimated costs are large, but the optimal policy is insensitive to these differences Ultrasound with spiral CT angiography is most effective and cost-effective in all countries

ASSESSMENT OF METHODOLOGIES TO COMBINE SURVIVAL CURVES IN META-ANALYSIS

CC Earle, B Pham, G Wells University of Ottawa, Ottawa, Canada

Purpose To assess the accuracy of different methods of combining published survival curves into a summary curve that can be used in disease models or to make treatment comparisons in meta-analysis

Methods We identified five methods for combining survival curves 1) iterative generalized least-squares (IGLS, Dear 94), 2) meta-analysis (MA) of failure-time data with adjustment for covariates (MFD, Hunink 94), 3) non-linear regression (NLR, Shore 90), 4) log relative risk (LRR, Voest 89), and 5) weighted LRR (w-LRR) We scanned published survival curves into a graphical package, identified coordinates along each curve, and extracted required data elements semi-automatically Two separate analyses were done 1) we used each method to combine the survival curves from eight single arm Phase II trials of chemotherapy in 918 patients with advanced non-small cell lung cancer (NSCLC), 2) we used each method to combine the survival curves from seven randomized trials in order to reproduce a MA of chemotherapy in advanced NSCLC We compared the resulting summary curves using the Kolmogorov-Smirnoff goodness of fit test to a "gold standard" curve calculated from the corresponding individual patient data (IPD)

Results Data extraction from the published survival curves was reliable All methods were able to produce reasonably accurate summary survival curves, and no method was consistently more accurate than the others Maximum discrepancies of the calculated summary curves from the IPD curves ranged from 1 8 – 4.1% for the first analysis, and up to 11% in the reproduction of the meta-analysis None of these differences were statistically significant All methods were able to differentiate between the two treatment arms of the meta-analysis, which had 6 month survival rates of 33 5% and 51 8% in its two treatment arms Only one method, IGLS, preserves randomization for theoretically justified comparisons of intervention arms It also does not require provision of unusual detail in published reports Characteristics of the component trials that affected the accuracy of the different techniques were 1) a high proportion of censored observations (MFD), 2) variability in the length of follow-up (IGLS, NLR, LRR, w-LRR), and 3) the heterogeneity of the treatment results (NLR, w-LRR)

Conclusions All methods were able to accurately reproduce summary survival curves from the published literature and to detect treatment differences Dear's method is preferred for comparing intervention groups because it preserves randomization

COST-EFFECTIVENESS OF BLOOD PRESSURE SCREENING IN ADOLESCENCE

AM Cheung, MC Weinstein, NR Cook, KA Freedberg, ANA Tosteson, LW Williams, L Goldman, MW Gillman Harvard School of Public Health Harvard Medical School, Boston University School of Medicine MA, Dartmouth Medical School, NH, and San Francisco School of Medicine, CA

Purpose To assess whether blood pressure screening at age 15 (with lifestyle interventions such as weight reduction, exercise or dietary sodium reduction for adolescents with blood pressure at or above the 95th percentile) or population-wide life style interventions for all 15-year-old adolescents is more cost-effective

Methods We estimated the intervention effects from meta-analysis of published studies, and we assumed that even with yearly maintenance interventions the effects would diminish by 5% per year We used a combination of standard life-table analysis and longitudinal tracking correlations to model the effects and costs of blood pressure screening from childhood until age 35 To simulate subjects from age 35 to death, we used the Coronary Heart Disease Policy Model, a computer simulation model of the morbidity, mortality, and costs of coronary heart disease and other blood pressure-related diseases

Results Most blood pressure screening strategies for 15 year-old children are more cost-effective than population-wide interventions ($76,000-$131,000 for boys and $120,000-$216,000 for girls compared with $120,000-$187,000 for boys and $212,000-$330,000 for girls per discounted year of life gained) In general, boys have lower (more attractive) cost-effectiveness ratios than girls Overall, when compared to no screening or intervention, the most cost-effective strategy for the control of blood pressure is screening plus dietary sodium reduction for those who are hypertensive ($76,000 and $120,000 per discounted year of life gained for boys and girls, respectively) The next most cost-effective strategy among the ones evaluated is screening and weight reduction for the obese hypertensives and dietary sodium reduction for the non-obese hypertensives with incremental cost-effectiveness ratios of $122 000 and $213,000 per discounted year of life gained for boys and girls, respectively Our estimates of cost-effectiveness are most sensitive to our assumptions of the costs of intervention the costs of maintenance programs, the annual decrement of intervention effects, the discount rate, and the proportion of hypertensive adolescents with secondary hypertension They are not very sensitive to the magnitude of longitudinal tracking correlation coefficients of blood pressure and the proportion of secondary hypertension with curable causes

Conclusions Both blood pressure screening and population-wide interventions are not very cost-effective Future research efforts in the childhood prevention of essential hypertension should focus on finding inexpensive strategies that have prolonged effects

DEVELOPMENT AND VALIDATION OF EMPIRICALLY DERIVED RISK-ADJUSTMENT INDEXES FOR A DEMENTED POPULATION

JF Ricci, BC Martin Department of Clinical and Administrative Sciences University of Georgia, Athens GA

The objectives of the study were to develop and validate risk-adjustment indexes for patients with dementia or Alzheimer disease (A/D) for two outcomes total cost utilizations and mortality

A 7-year retrospective review of Georgia Medicaid claims data from 1990 to 1996 was used to detect first time A/D patients Comorbid conditions were measured from medical and pharmaceutical claims submitted during 12 months prior the first A/D event A/D was defined by selected ICD-9-CM codes from the 290 XX, 294 XX 331 XX and 797 XX series The risk-adjustment dementia-specific weighted-indexes were derived from the Charlson index and Chronic Disease Score Multivariate stepwise logistic and linear regressions were performed on one half of the cohort to predict respectively mortality and total cost utilizations All regression models controlled for age gender race eligibility type and follow-up time We then tested the indexes on the second half of the cohort

We identified 4 160 A/D patients with a mean age of 73 years (range 40 – 102) Of all patients 32% died within the two-year follow-up 73% were women and 49% white A total of 2 080 patients were randomly assigned to each of the two cohorts The 2-year disease-specific mortality index (8 comorbidity covariates) and the one-year total cost index (10 comorbidity covariates) were more concise than either the generic Charlson or CDS index alone Also, the dementia-specific mortality index showed a better stepwise increase in the cumulative mortality (p log rank χ^2<0 001) than the Charlson index on the validation sample Finally, the correlation between predicted and actual total costs in the validation sample was 0 41 with a difference between mean predicted and actual total costs inferior to 6%

These shorter A/D-specific indexes allow for the development of more highly discriminant comorbidity models for risk adjustment They help better predict future resource utilizations of demented patients In the future, the indexes will be tested on the North Carolina Medicaid population

A BAYESIAN APPROACH TO DIAGNOSTIC TEST EVALUATION PERFORMANCE OF A WHOLE BLOOD NEAR PATIENT TEST FOR *HELICOBACTER PYLORI* IN PRIMARY CARE

BC Delaney, RL Holder, TF Allan, JE Kenkre, FDR Hobbs The Dept of General Practice, The University of Birmingham, Birmingham, UK

Background Spectrum bias, describing the variation in performance of a diagnostic test between populations on account of an imperfect reference standard, is a particular problem for evaluations of near patient tests (NPTs) in the primary care setting In addition, there is no single accepted gold standard for the identification of H pylori infection, tests usually being compared against a range of invasive tests at gastroscopy This method is not available where patients are not undergoing gastroscopy and only laboratory-based serological tests can be used as a comparison The potential for spectrum bias does not allow the evaluation of NPTs for H pylori in selected populations to be applied to the primary care setting This study aims to evaluate the performance of a NPT for H pylori in this setting using a novel statistical method to allow for the lack of direct reference standard

Method 311 patients including both those with known dyspepsia and asymptomatic patients were recruited from primary care centers in Birmingham, UK The Helisal Rapid Blood test (Cortecs diagnostics, Clywdd) was performed in the centers, and a concurrent venous sample taken for two ELISA tests (Helisal and Helico G, Porton, Cambridge UK), performed at the laboratory, blind to the NPT result Analysis was conducted using a Gibbs sampling procedure β-prior distributions were constructed using data from evaluations of the same NPT in the secondary care setting, identified from a systematic review of the literature The study data was used to revise these priors to obtain posterior distributions, using the software Bayesian Utilities and Graphics Software (BUGS, Medical Research Council Biostatistics Unit, Cambridge, UK) Seven variables underlying eight data values were simultaneously computed to form 95% credible intervals for the performance of the NPT and ELISA tests and the prevalence of H pylori

Results The estimate of prevalence of H pylori in the study population was 64% (95% Credible Intervals, 59-70%), the sensitivity and specificity of the NPT were 89% (94-94%) and 84% (77-91%)

Conclusions The NPT performs as well as laboratory-based tests and would be expected to perform satisfactorily in primary care in selecting dyspeptic patients under the age of 50 for endoscopy The method could be used to enhance the evaluation of diagnostic tests in settings where gold standards do not exist or cannot be used for ethical or practical reasons

CHOICE OF THROMBOLYTIC IN MYOCARDIAL INFARCTION IDENTIFYING PATIENTS MOST LIKELY TO BENEFIT FROM T-PA,

DM Kent, S Yuan, RA Hayward, University of Michigan, Ann Arbor, MI, J Griffith, J Beshansky, HP Selker, New England Medical Center, Boston, MA

Thrombolysis is a critical therapy for acute myocardial infarction (AMI) The GUSTO trial demonstrated a 1% absolute mortality reduction for those treated with tissue plasminogen activator (t-PA) over those treated with streptokinase (SK) Primarily because of expense, however, t-PA has not been universally adopted and there is controversy about which patients warrant the more expensive therapy Therefore, we designed a prognostic tool to identify patients likely to gain the most benefit from t-PA based in part on the Thrombolytic Predictive Instrument (TPI)

The TPI is a prognostic instrument developed and validated on a large data set to predict patient-specific likelihood of thrombolytic benefit This instrument, which is incorporated into EKG machines for convenient use in real time, predicts mortality risk with and without thrombolytic therapy, based on easily obtainable clinical and EKG variables To develop a model that can inform the choice of thrombolytic agent, we assumed that the likelihood of benefit from t-PA over SK in a given patient is proportional to the likelihood of mortality benefit from any form of thrombolysis in that patient Thus, we derived a prediction of patient-specific relative benefit from t-PA over SK using TPI-predicted benefit of thrombolysis (i e the difference in TPI-predicted mortality with and without thrombolytics) and patient-specific life expectancy (as determined by the patient's age and gender)

We studied 823 patients who met EKG criteria and received thrombolytic therapy for AMI at 28 US hospitals According to the model, fully 70% of the total life-years gained by using t-PA (instead of SK) in all patients could be achieved by using t-PA in only 25% of patients, incremental costs of t-PA treatment in this group would be less than $20,000 per life-year saved However, those in the lowest quartile of predicted benefit received less than 5% of the total life-years gained, incremental costs of t-PA treatment in this group would be well in excess of $200,000 per life-year saved

This model predicts that most of the incremental benefit of t-PA accrues to a small, definable sub-set of patients presenting with AMI, in whom it is unjustifiable to use SK Further, a substantial proportion of patients are highly unlikely to get incremental benefit from t-PA, the more expensive agent If validated, this model can be used at the bedside to tailor thrombolytic therapy to individual patient characteristics

AT WHAT AGE IS ONE-TIME COLONOSCOPIC SCREENING FOR COLORECTAL CANCER MOST COST EFFECTIVE FOR MALE PATIENTS?

RM Ness, RW Klein, AM Holmes, RS Dittus Indiana University, Indianapolis, IN and Vanderbilt University, Nashville, TN

One-time colonoscopy has been suggested as a method for screening the general population for colorectal cancer (CRC) We compared the cost-utility (C/U) of performing one-time colonoscopic screening at various ages using decision modeling.

We constructed a discrete-event simulation model of the natural history of colorectal neoplasia in male patients Probability estimates for the risk of colorectal cancer (CRC), the incidence of colonic neoplasms, the mortality from CRC, the diagnostic performance of colonoscopy, the complication rates of colonoscopy and surgery, and the direct costs of CRC diagnosis and treatment were summarized from the literature Quality adjusted life years (QALYs) using utilities as the weights were computed as the measure of effectiveness We derived utility estimates using an instrument that we specifically designed to measure utilities for outcome states of CRC A hypothetical cohort was created consisting of 40 year old male adults not previously screened for CRC We modeled one-time colonoscopic screening at alternative ages to evaluate the effect of age at screening on cost-utility The screening strategy included the initial removal of all identified polyps and post-polypectomy surveillance as currently recommended for men in whom more than one polyp or a large polyp was found We discounted both costs and QALYs at a 3% annual rate

Screen age	Cost ($)	QALYs	Marginal cost($)	Marginal QALYs	Marginal C/U ($/QALY)
58	534	20 8645	-	-	-
54	550	20 8756	16	0 0111	1392
50	601	20 8791	51	0 0035	14651

The mean per-person results are shown above Screening at age 58 had the lowest cost and dominated screening at all older ages (including never screen), while screening at 50 was most effective and dominated all younger ages The fewest cancers and cancer deaths were incurred when screening was done at age 54 The marginal C/U ratio of moving to earlier screening at 50 years of age is well within generally acceptable limits and decreases further as the sensitivity of colonoscopy increases

MANAGEMENT OF ANTIRETROVIRAL THERAPY FOR HIV INFECTION MODELING WHEN TO CHANGE THERAPY

Rebecca M D'Amato, Richard A D'Aquila, Lawrence M Wein MIT, Cambridge, MA Massachusetts General Hospital, Charlestown, MA and MIT, Cambridge, MA

Objective -To evaluate different strategies for monitoring plasma HIV RNA levels and/or resistance genotypes to decide when to change antiretroviral therapy

Design and Setting -A Monte Carlo simulation model is developed which tracks patients' viral loads and presence of opportunistic infection during the course of therapy The model uses clinical and virological data and statistical variation in patient parameters This computer modeling evaluates strategies intended for use outside controlled trials in clinical practice

Main Outcome Measures -The objective of the modeling is to determine which strategies minimize the delay in detection of viral rebound while maintaining a low (prespecified) probability of switching therapy before viral rebound occurs

Results -Guidelines published in 1997 and all strategies using rebound of viral load as the trigger for switching drug regimens involve lengthy delays in detection of rebound, particularly when patients are drug-naive and the detection limit of the viral load assay is 500 HIV RNA log copies/mL An assay for plasma HIV RNA with a detection limit of 20 HIV RNA copies/mL decreases delay of detection of viral rebound substantially Genotyping achieves only minor additional reductions in detection delay A policy based on switching drugs at a predetermined time after initiation of therapy if viral load has not yet rebounded (proactive switching) leads to the smallest detection delays and lowest virologic assay costs of any of the strategies tested.

Conclusions The modeling indicates that prolonged periods may be required for viral load to rebound to detectable levels Policies for switching antiretroviral therapies after viral load rebound is detected will likely be ineffective at preventing selection of drug-resistant HIV during drug failure A proactive policy, which periodically changes combination regimens prior to detecting evidence of drug failure, may significantly reduce the duration of viral replication under selective pressure of a failing regimen This requires testing in clinical trials

PHYSICIANS' DIAGNOSIS OF CHILDHOOD ASTHMA

LN Werk, S Steinbach, W Adams, H Bauchner Boston Medical Center, Boston University, Boston, MA

Introduction It is unclear how and when physicians make the diagnosis of childhood asthma Expert guidelines from sources like the NIH present extensive lists of possible features of the disease but lack clinical utility Purpose We undertook this study to determine which factors primary care physicians believe are necessary to make the diagnosis of childhood asthma Methods Pediatricians in MA were surveyed using a self-administered questionnaire Selection criteria were designed to identify respondents who care for children in a primary care setting The questionnaire included a Likert scale of twenty factors used in establishing the diagnosis of asthma, inquiries concerning obstacles that may delay or discourage the diagnostic labeling of asthma, and vignettes with diagnosis and treatment options Results A total of 451 of 681 eligible participants (66 %) completed surveys The three factors considered most critical to the diagnosis of asthma were recurrent wheezing, symptomatic improvement with a bronchodilator, and recurrent cough No single factor among 20 listed was cited by a majority of clinicians as necessary to establish the diagnosis Some beliefs inconsistent with asthma guidelines included 28 % of respondents identifying a minimal age requirement of two years as important to establishing the diagnosis and 19 % requiring the absence of fever during acute episodes 20 % of respondents reported they were likely to delay the diagnostic label of "asthma" due to concern for future problems in entering the military, parental restriction of activities, and a negative response from parents Diagnoses physicians share with parents in place of asthma include reactive airway disease, asthmatic bronchitis, and recurrent bronchiolitis In a series of vignettes including presentations of children with asthma, 67 – 96 % of clinicians correctly identified the diagnosis However, inappropriate use of antibiotics was as high as 25 % and 38 % failed to consider prescribing corticosteriods Further, antibiotics were more commonly prescribed if the clinician considered asthmatic bronchitis as a likely diagnosis and corticosteriods prescribed if the clinician considered asthma as a likely diagnosis Conclusion Pediatricians uniquely select and weigh criteria to establish the diagnosis of asthma – some inconsistent with expert guidelines Obstacles are cited that delay "labeling" a child with the diagnosis In addition, euphemisms such as asthmatic bronchitis are used in place of the term "asthma" and may lead to suboptimal therapy A specific set of criteria (perhaps weighted as major and minor) or creation of a decision tree is necessary in order to prevent delay in recognition of, improve diagnostic labeling of, and promote appropriate therapy for childhood asthma

STATISTICAL MODELING OF EXPERT PANEL RATINGS IN THE DEVELOPMENT OF APPROPRIATENESS CRITERIA FOR MEDICAL PROCEDURES: AN APPLICATION TO LAMINECTOMY IN SCIATICA PATIENTS.

V Wietlisbach, JP Vader, B Burnand, F Porchet, MC Costanza
Institute of Social and Preventive Medicine, University of Lausanne, Lausanne, Switzerland, Department of Neurosurgery, Centre Hospitalier Universitaire Vaudois, Lausanne,Switzerland, Medical Biostatistics Unit, College of Medicine, University of Vermont, Burlington, VT

The widely used RAND/UCLA method to develop appropriateness criteria for medical procedures consists in identifying relevant factors through literature review, building a set of clinical conditions (CC) by combination of the factor levels and asking a panel of experts to rate whether the procedure is indicated or not in each specific CC However, the factorial design often leads to a high number of CC that increases the variability in the expert judgments and obscure the logic linking the factors to the rating decisions In this respect, the performances of two statistical techniques, stepwise logistic regression (SLR) and classification trees (CT) were tested in modeling the relationship between seven factors, defining 720 CC, and the verdicts issued by two nine-expert panels (USA-1992 and Switzerland-1995) on the appropriateness of performing laminectomy in patients with sciatica The panelists of both countries rated the procedure as inappropriate in a large part of the CC (60% and 70% respectively) and, particularly, in nearly all 240 CC not presenting any imaging evidence of disk herniation For both panels, the SLR selected the same four factors as being the most influential, i e in decreasing order the level of imaging evidence, the degree of disability, the severity of neurological findings and the intensity of previous conservative treatments The CT resulted also in the same hierarchy of factors, the first split diverting the CC without imaging evidence of hernia into a terminal group of inappropriate indications The other CC were further divided according to the other three factors, in a cascade of similar binary splits for the US panel and in a more balanced branching for the Swiss panel When using only the four selected factors, CT was more efficient than SRT in the US panel (correct classification rate 93% vs 85%) and less efficient in the Swiss panel (89% vs 93%) Understanding how the panelists have taken into account the different factors may be critical in outlining practice guidelines best adapted to the local pathways of medical thinking and in designing more parsimonious sets of CC for future expert panels (or for the second rating round of same panel)

COST BENEFIT AND COST EFFECTIVENESS ANALYSIS OF METHOTREXATE AND GOECKERMAN THERAPY FOR PSORIASIS A STUDY REVISITED

S Chen, AM Garber Stanford University, Stanford, CA, and VA Health Care System, Palo Alto, CA

We previously performed cost benefit (CB) and cost effectiveness (CE) analysis of Goeckerman (Gkm) and liquid methotrexate (MTX) therapies for psoriasis, using the visual analog scale (VAS) to assess utilities and willingness to pay (WTP) methods to assess benefits (in dollars) among patients at the Stanford Dermatology Clinic The CE analysis implied that Gkm should be chosen over liquid MTX at all levels of severity of psoriasis if the cutoff CE ratio exceeds $35,000 The cost-benefit analysis (CBA), in contrast, implied that the net benefits of MTX exceeded those of Gkm in the treatment of severe psoriasis, and that the costs of both therapies exceeded their benefits in moderate and mild psoriasis

The apparent discrepancy between the CE and CB analysis could reflect the use of an inappropriate CE cutoff or inaccurate utility assessment Dartmouth investigators measured utilities for psoriasis using the time trade-off (TTO) and standard gamble (SG) as well as VAS Although their definition of disease severity differs slightly from ours, we sought to determine whether using their utilities would eliminate the apparent discrepancy between the results of our CB and our VAS-based CE analyses

Methods For the CE analysis, we considered a ratio of costs to effectiveness of less than $35,000 to be "cost-effective" We also analyzed the results using a $50,000 cutoff

Results Using Dartmouth VAS scores and the $35,000 cutoff, Gkm remains CE against liquid MTX across all levels of disease severity, but it is not CE compared to MTX when utilities are assessed by TTO or SG At a $50,000 CE cutoff, Gkm remains CE when VAS is used to assess utilities However, when utilities are assessed by TTO or SG, Gkm is not CE against MTX for mild or moderate psoriasis

Conclusion If the appropriate CE cutoff is used, the results from CB and CE analyses should lead to similar rankings of alternatives Using the TTO and SG utilities and the $35,000 cutoff, the results from the CE analysis are consistent with our previous CBA Gkm is generally not CE against MTX This finding suggests that VAS-obtained utilities for the state of ill health are too low when compared to WTP, SG, and TTO methods, leading to an exaggeration of the value of quality-improving interventions

Rankings are inconsistent when a $50,000 CE threshold is used Since the threshold is arbitrary, it may not correspond to the willingness-to-pay for QALYs in the population we studied

A RANDOMIZED, COMPARATIVE TRIAL OF SHARED DECISION MAKING FOR PROSTATE CANCER SCREENING ONE-YEAR FOLLOW-UP

RJ Volk, SJ Spann, AR Cass Baylor College of Medicine, Houston, TX, and The University of Texas Medical Branch, Galveston, TX

Recent guidelines from the American College of Physicians and American Academy of Family Physicians recommend individualized decision making regarding prostate cancer screening This study evaluated a patient education, shared decision making approach to prostate cancer screening

Male family medicine patients, 45 to 70 years of age, with no history of prostate cancer, were randomly assigned to view the Shared Decision-making® Programs videotape "The PSA Decision What YOU Need to Know", or to a control group Baseline assessments included a measure of prostate cancer screening knowledge, past screening behavior, and intention to be screened in the future The same knowledge measure was readministered at a 2-week follow-up assessment, at which time control group patients received also a brief educational brochure on prostate cancer The one-year follow-up assessment included the knowledge measure and questions about prostate cancer screening over the past year

One-year follow-up assessments were completed for 131 of 160 patients (82% retention) Over 70% of patients in either group reported having a general physical examination during the past year Patients having viewed the PSA videotape demonstrated greater knowledge of prostate cancer than controls (P< 001), although these differences were less than those observed at the two-week follow-up assessment Most men (over 90%) in both groups felt they were fairly or very well informed about the harms and benefits of screening for prostate cancer Regarding testing behavior, 33 8% of patients in the videotape group reported having a PSA test in the past year, compared to 56 1% of controls (P= 011) Similarly, 44 6% of videotape group patients indicated an intention to be screened in the future, compared to 71 2% of controls (P= 002)

Patient education for prostate cancer screening appears to improve patients' knowledge, although these gains are modest and decline over one year Yet, most patients consider themselves fairly well informed about the potential harms and benefits of prostate cancer screening Furthermore, the intervention appears to impact screening behavior and intention to be screened in the future Periodic patient education may be needed to ensure informed prostate cancer screening decisions are being made

COST-EFFECTIVENESS OF RALOXIFENE AND HORMONE REPLACEMENT THERAPY IN POSTMENOPAUSAL WOMEN

Tze-Ming Chen, Katrina Armstrong, Daniel Albert, Janie Hrung, and J Sanford Schwartz University of Pennsylvania, Philadelphia, PA

Hormone replacement therapy (HRT) reduces the long term adverse effects of menopause However, many women decline HRT because it may increase breast cancer (BCA) risk Raloxifene, a selective estrogen receptor modulator, increases bone mineral density, decreases total cholesterol and low density lipoproteins, and may decrease BCA incidence To date, no study has examined the effect of raloxifene on long-term postmenopausal clinical outcomes

We performed a decision analysis to compare the impact on life-expectancy and the cost-effectiveness of HRT and raloxifene in healthy 50 year old postmenopausal women without coronary heart disease (CHD) risk factors A Markov model was constructed consisting of 78 health states ranging from perfect health to death and representing the five diseases affected by the alternative strategies CHD, vertebral fracture, hip fracture, endometrial cancer, and BCA Baseline incidence and mortality transition probabilities were derived from data from the Framingham Heart Study, from the Rochester, Minnesota population study, and from SEER databases In the base-case, raloxifene was not modeled to prevent BCA

The life expectancy of healthy 50 year old women receiving no intervention in the model is 78 4 years Under base-case conditions, both HRT and raloxifene increase life-expectancy relative to no treatment, however, while HRT is both cost- and life-saving, raloxifene costs $340,000 per year of-life saved One- and two-way sensitivity analyses demonstrate that only CHD and BCA incidence and mortality rates altered the optimal life-saving option from the base-case HRT is the optimal cost-effective therapy except in populations at increased risk for BCA and at low risk for CHD relative to base case conditions, raloxifene becomes cost-effective if it decreases BCA incidence or if the cost of the medication is lowered

For postmenopausal women, the optimal life-saving treatment alternative is dependent upon patient CHD and BCA risks For women unwilling to consider HRT, raloxifene will prolong life expectancy but will be cost-effective only under certain conditions

PATIENTS WILLINGNESS TO PAY FOR DIAGNOSTIC IMAGING ACCURACY

FJ Papatheofanis, University of California San Diego, La Jolla, CA

In recent years, contingent valuation has been used to elucidate economic measurement of patient preferences Diagnostic radiology is currently perceived to represent a significant cost burden for health care delivery organizations However, there exists a poor understanding of patient preferences regarding the attainment of diagnostic accuracy through the utilization of diagnostic imaging This investigation employed contingent valuation methodology to measure the willingness of patients undergoing diagnostic imaging to pay for incremental improvement in diagnostic accuracy in the evaluation of suspected benign and malignant pulmonary disease

Two hundred patients completed the self-administered questionnaire which included items on demographics, risk perception, trade-off valuation, and willingness-to-pay valuation Parametric and nonparametric analysis estimated median willingness to pay Regression analysis [logit model] was used to analyze answers to binary questions Diagnostic performance increments were defined as 50,75, 85, 95, 98, 99, and 100% accuracy in the evaluation of suspected primary pulmonary carcinoma and tuberculosis

Median willingness to pay for diagnostic imaging evaluation of suspected pulmonary carcinoma was approximately $15,400 per patient, and $4,300 for tuberculosis Multivariate analysis demonstrated a direct correlation of willingness to pay with perceived risk of disease or death

Patients who are undergoing diagnostic imaging studies for suspected benign and malignant disease value diagnostic accuracy highly These patients are willing to pay significant amounts out of pocket to gain incremental improvement in accuracy

THE LEARNING CURVE FOR SENTINEL NODE BIOPSY IN BREAST CANCER - PRACTICAL CONSIDERATIONS WHEN INTRODUCING NEW TECHNOLOGY

Richard K Orr, Nananda F Col, James L Hoehn Marshfield Clinic, Marshfield, WI and New England Medical Center, Boston, Ma

Introduction Sentinel node biopsy (SNB) has been proposed as a low morbidity alternative to formal axillary lymph node dissection (ALND), the gold standard for early breast cancer staging Although SNB obviates the need for complete ALND in those patients with histologically negative nodes, it is associated with a significant learning curve and additional "start-up" costs, including the need for a purchasing a radioisotope scanner (gamma probe) Initial efficacy studies of SNB required full ALND to assure appropriate sensitivity of the procedure, but it has become obvious that many surgeons are eliminating the controlled learning phase and performing SNB instead of ALND

Design To determine the effect of performing SNB instead of ALND, we created a decision anlysis model simulating a randomized trial of 1000 women in each treatment arm We compared accuracy and outcome during the surgeon's learning phase (40-80 procedures) and thereafter We used published literature estimates for sensitivity of SNB

Main outcome measures Percentage of patients with inaccurate staging, overall survival, cost-benefit of SNB

Results Performance of SNB instead of ALND results in inability to locate a sentinel node in 40% of attempts during the learning phase (vs 20% in later cases) and understaging in 12% of patients in the learning phase (vs 0% in later cases) This understaging is associated with a small decrement in survival (1-2%) and an increased risk of axillary recurrence SNB is cost-effective only when the sensitivity for detecting postive lymph nodes exceeds 95% and the cost of SNB is less than 50% of the cost of ALND

Conclusion To ensure accurate staging of breast cancer patients, all surgeons should perform formal axillary dissection while learning SNB techniques After documentation of accuracy (sensitivity for detecting positive nodes >95%), axillary dissection may be avoided for those patients with histologically negative sentinel nodes Because of the high start-up costs and steep learning curve, SNB will not be cost-effective in hospitals with limited numbers of breast cancer patients

ARE TROPONIN I OR SESTAMIBI ALONE ADEQUATE STRATEGIES TO RULE OUT ACUTE CORONARY SYNDROMES? A FORMAL DECISION ANALYSIS

Andrew Sucov, MD
University of Rochester School of Medicine and Dentistry, Rochester, NY

Introduction 1 5-2 million US patients/year are admitted to rule out acute coronary syndromes (ACS) at a cost of $10-20 billion EPs are reluctant to admit fewer patients, as 2-4% of all MIs are currently missed Decision analysis can model strategies to maintain or increase sensitivity and limit costs Methods Literature review to identify sensitivity and specificity of the following strategies CK mass plus either stress echo or sestamibi if CK was negative, Troponin I alone, Troponin I plus either stress echo or sestamibi if Troponin was negative, or sestamibi alone Costs used were 1998 $US reimbursement from either the dominant local HMO or inpatient DRGs Sensitivity analysis was performed on all variables, two- and three-way sensitivity analyses were performed on variables that resulted in changes in optimal strategy Outcomes reviewed were the sensitivity, the positive predictive value (PPV) and $/ACS patient identified for each strategy Disease prevalence was varied from 1-25% Data 3 0 from TreeAge (Williamstown, Ma) was used for analysis Results See table for base-case results Results are sensitive to the sensitivity and specificity of Troponin I and stress echo, but not to costs or disease prevalence Troponin I + stress echo is the most sensitive strategy, but the marginal cost-effectiveness is very high ($77,000/ACS patient) compared to Troponin alone If Troponin I sensitivity is <94%, CK mass + stress echo are more sensitive, if <74%, they are more cost effective Troponin I + sestamibi might be more sensitive than Troponin I + echo, but is never cost-effective Conclusion Troponin I alone may be an adequate strategy to rule out ACS if sensitivity is >94% More data on Troponin I's sensitivity and specificity for ACS is needed to confirm these results

	Sensitivity (%)	PPV (%)	Cost ($/ACS pt identified)
Troponin alone	94 0%	46 5	16,751
Troponin + Echo	99 3%	31 2	21,056
Troponin + Sestamibi	98 7%	33 6	26,076
CK mass + Echo	94 0%	40 0	20,592
CK mass + Sestamibi	86 7%	43 6	29,666
Sestamibi alone	80.0%	42.6	26,318

THE RATING SCALE CANNOT BE SUBSTITUTED FOR THE TIME TRADE-OFF IN PATIENTS WITH CLINICALLY LOCALIZED PROSTATE CANCER

MW Kattan, ME Cowen, PA Fearn, D Cahill, JR Beck, BJ Miles Baylor College of Medicine, Houston

A recent analysis (MDM,17(4) 542,1997) illustrated that applying a published decision analysis to the individual using group utilities, rather than a personal utility profile, was often suboptimal for the patient This was illustrated by comparing the results of group vs individual time trade-off utilities While the time trade-off is often advocated due to its theoretical appeal, it is expensive to measure Measuring utilities with the rating scale method is much simpler and faster yet may not yield the same results The purpose of this study was to learn whether use of the rating scale in the prostate cancer treatment model would result in the same treatment recommendations as those provided by using the time trade-off

Health state utilities were assessed for 63 men using both the time trade-off and rating scale methods The health states measured were those of the aforementioned prostate cancer treatment model watchful waiting, impotence, incontinence, hormonally responsive disease, and hormonally refractory disease We calculated the quality-adjusted life expectancies for radical prostatectomy and watchful waiting using each set of utilities We did these calculations three times once assuming the patient was diagnosed with well-differentiated disease, and again with moderate and poorly-differentiated disease All calculations were tailored to the patient's age and number of comorbidities We then compared the treatment recommendation using the rating scale utilities with the treatment recommendation using the time trade-off utilities to determine whether the rating scale utilities provided comparable answers

The results suggest that the rating scale treatment recommendations often disagreed with the time trade-off treatment recommendations Assuming well, moderate, and poorly differentiated disease, the misclassification percentages of the rating scale utilities (assuming time trade-off to be the gold standard) were 49%, 52%, and 35%, respectively The Kappa statistics for well, moderate, and poorly differentiated disease were 0 013 (95%CI -0 231 to 0 257), -0 065 (-0 3 to 0 171), and -0 052 (-0 273 to 0 170), respectively Our hypothesis was that in cases where the rating scale suggested that one treatment very strongly dominated another (e g , was associated with >5 QALYs of benefit), the rating scale would be a suitable substitute However, cutoffs for this purpose were not readily apparent

In conclusion, personalized rating scale utilities are not substitutes for personalized time trade-off utilities in the decision analysis for the treatment of clinically localized prostate cancer

A COST-EFFECTIVENESS MODELING EFFORT FOR ARTHRITIC JOINT REPLACEMENT DECISIONS

GB Hazen, JM Pellissier, RW Chang Northwestern University, Evanston IL, Merck Research Laboratories, Blue Bell PA, Northwestern University, Chicago IL

We performed cost-effectiveness analyses for total hip replacement and total knee replacement decisions due to osteoarthritis, using stochastic tree modeling techniques

We constructed stochastic tree models for total hip or knee replacement using the Excel spreadsheet add-in StoTree StoTree is a graphical model formulation and solution tool for stochastic trees Operating in a spreadsheet environment, the StoTree user can access all the usual features of that environment in addition to StoTree Using the factoring capabilities of StoTree, separate models for osteoarthritis progression, joint-replacement outcome, prosthesis loosening, and background mortality were constructed and subsequently linked We have found it convenient to use a novel extension of influence diagrams to depict links between these models We incorporated quality of life into the model by including the patient's functional status on the American College of Rheumatology (ACR) scale Data on quality of life was elicited from subjects using a computer-assisted utility assessment protocol Cost data was obtained from Northwestern Hospital's cost accounting system and from the literature

For a white male age 85 in ACR functional class III, our results indicate that on average, total hip replacement cost (discounted at 3% per year) exceeds cost of conservative medical management by approximately $10,200, but improves average discounted quality-adjusted life expectancy by 2 0 years The resulting CE ratio of $5100 per quality-adjusted life year compares favorably with such well accepted procedures as cardiac bypass and renal dialysis For a white female age 60 in class III, hip replacement actually reduces costs by $120,000 and improves quality-adjusted life expectancy by 7 1 years The key insight is that hip replacement reduces the time spent in severe functionally disabled states whose high associated custodial costs outweigh the initial cost of the procedure

Hip and knee replacement procedures seem particularly vulnerable to scrutiny in economic terms due to their elective nature, their geriatric target population, their high short-term cost, and the fact that they do not extend life Nevertheless, when quality of life is included, our results show that these procedures can be very cost-effective, comparing favorably to most accepted procedures

EVALUATING THE EFFECT OF TREATMENTS ON FUNCTIONAL HEALTH AND WELL BEING AN ANALYSIS OF ONE HUNDRED AND SEVENTY EIGHT SF-36 STUDIES (1988-1997)

M Manocchia, JE Ware Jr Health Assessment Laboratory, New England Medical Center, Boston, MA

Growing concerns over the costs, quality and outcomes of medical treatments have led to the widespread use and standardization of patient based assessments to assist in clinical decision making As documented in over 750 published studies through 1997, the MOS 36 Item Short Form Health Survey (SF-36) has become the most widely used standard for comparing outcomes In this report, we summarize 178 published treatment studies We standardized each of the SF-36 scales reported in the studies and computed physical [PCS] and mental [MCS] summary scores if they were not published The SF-36 Treatment Effects Database consists of 69 (38 8%) surgeries, 56 (31 5%) pharmaceutical trials, 30 (16 9%) clinical practice follow ups, 14 (7 9%) behavioral/educational interventions and 9 (5 1%) physical rehabilitations These studies were concentrated in the following therapeutic areas 77 (43 3%) musculoskeletal, 24 (13 5%) cardiovascular, 19 (10 7%) endocrine, 16 (9 0%) mental illness, 13 (7 3%) viral, 12 (6 7%) respiratory, and 8 (4 5%) other vascular Our analysis indicates significant differences in average changes in PCS between medical interventions for surgeries showing the greatest average improvements in physical health (+5 6, 56 s e) followed by physical rehabilitations (+2 6, 1 4 s e) and pharmaceutical trials (+1 6, 51 s e) Significant differences in average changes in MCS showed that behavioral/educational interventions had the greatest average improvements in mental health (+5 4, 1 4 s e) followed by pharmaceutical trials (+3 4, 0 6 s e) and surgeries (+3 3, 0 6 s e) Interestingly, surgeries outperformed other medical interventions in improving vitality scores (+4 6, 0 5 s e) and social functioning scores (+4 8, 0 7 s e) on the SF-36 The largest average change in any SF-36 scale due to medical interventions was reflected in average improvements in bodily pain scores due to surgeries (+7 8, 0 5 s e) These data further reflect the validity of the SF-36 and lend valuable information to clinical decision makers on how medical interventions can be compared under a similar metric and how patients are responding to their treatments in ways meaningful to them

A FUNCTIONAL MEASUREMENT APPROACH TO ASSESSING THE SUBJECTIVE VALUE OF TWO DIMENSIONS OF PATIENT SATISFACTION

Gregory J Zarow, Beth A Jaworski VASDHS, UCSD Dept of Psychology, San Diego California

Purpose the purpose of these experiments is to utilize Functional Measurement [FM] to measure patient satisfaction across differing doctor/patient visitation lengths and across differing information dissemination levels in two different hypothetical outpatient scenarios so as to determine the relationship between these two important dimensions of patient satisfaction Overview FM has been proposed as a valid methodology for assessing valuation However only a handful of examples exist in medical decision making literature By measuring, on a single scale, the subjective valuation of each level of one attribute paired with each level of a second attribute, FM allows for inference of the underlying mathematical operators influencing a multiply-determined decision Method Experiment 1 Subjects (Ss) rated satisfaction by placing a mark on a line which ranged from "extreme dissatisfaction to "extreme satisfaction' Following a vignette which posed a hypothetical scenario in which "Bob" presented with chest pains, hypertension, job stress and a family history of cardiovascular illness, Ss performed this rating task 15 times once for each of 3 levels of visit length [VL], once for each of 3 levels of information [Info] from a hypothetical doctor, and one for each of the 9 possible VL Info pairs Experiment 2 was identical to Experiment 1 except that outpatient "Jim" has a prolonged, worsening cold Results Results from experiment 1 indicate that longer-than-expected VL improves satisfaction, while shorter-than-expected VL decreases satisfaction and that satisfaction improves with Info Results from experiment 2 essentially replicate experiment 1, but with slightly less polarity of ratings, possibly due to the severity of the illness in each scenario More importantly, factorial graph of VL and Info versus satisfaction produce parallel lines, these factors appear to average together when generating a satisfaction rating Conclusion The parallel lines in the factorial graph validate that this response scale is a linear function of the underlying psychological constructions and that this response scale is an equal interval scale The elevation of these curves indicates that patient satisfaction depends, in part, on VL, with longer visits increasing satisfaction and on Info, with more Info increasing satisfaction The psychological law, or mathematical operator, which describes the relationship between VL and Info in generating the subjective value of satisfaction appears to be an averaging of the two attributes FM is a valid methodology for assessing the satisfaction of patients in medical scenarios and will facilitate decision making at the individual and the policy levels

COMPARISION OF EMERGENCY ANGIOPLASTY WITH FIBRINOLYSIS FOR THE TREATMENT OF MYOCARDIAL INFARCTION BY DECISION ANALYSIS

J Kellett Department of Medicine, Nenagh General Hospital, Nenagh, County Tipperary, Ireland

Purpose: The risks and benefits of fibrinolysis and primary angioplasty (PTCA) for the treatment of suspected acute myocardial infarction (AMI) were modelled by decision analysis

Methods: Data for the model's construction were obtained from MEDLINE searches and bibliographies of all studies reporting efficacy and side effects of fibrinolysis and primary PTCA for AMI

Results: a 65 year old patient with an "average" sized AMI gains 129, 193 and 202 days of life from streptokinase, accelerated tPA and PTCA, respectively Below 65 years of age PTCA was unlikely to produce any benefit over fibrinolysis, except for patients with very large infarctions with a high predicted mortality PTCA was usually preferred when predicted AMI mortality exceeded 10%, and aspirin when it fell below 2% At all ages the analysis was most sensitive to the chances of death from infarction, failed PTCA, procedural death from PTCA, and the relative effectiveness of PTCA over fibrinolysis The analysis was insensitive to the chance of stroke, bleeding, death from other causes, and quality of life adjustments If emergency coronary artery bypass surgery were unavailable PTCA was only preferred for the elderly or those with unusually large infarctions for their age

In Conclusion: probably primary PTCA for AMI should only be performed if predicted infarct mortality exceeds10%, and only if facilities for emergency bypass surgery are available

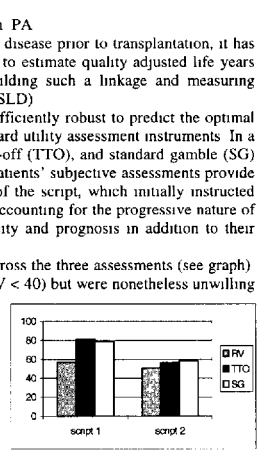
UTILITY ASSESSMENTS OF PATIENTS WITH CHRONIC LIVER DISEASE VARY AS A FUNCTION OF THE TIME HORIZON STUDIED

CL Bryce, J Switala, M Roberts University of Pittsburgh Pittsburgh PA

Although quality of life has been examined for patients with liver disease prior to transplantation, it has never been linked to numerical scales of utility, making it difficult to estimate quality adjusted life years (QALYs) The research reported here offers one approach to building such a linkage and measuring changes in utility throughout the course of end-stage liver disease (ESLD)

As part of a larger effort to build a simulation of ESLD that is sufficiently robust to predict the optimal timing of liver transplantation, we are assessing QALYs using standard utility assessment instruments In a pilot study (n=69) we administered relative value (RV), time trade-off (TTO), and standard gamble (SG) assessments to patients with ESLD using the UMAKER program Patients' subjective assessments provide quantitative measures of overall health We also tested variations of the script, which initially instructed patients to evaluate their immediate, current state of health without accounting for the progressive nature of their disease A second script allowed patients to consider chronicity and prognosis in addition to their current state of health

The latter script resulted in more consistent valuations of health across the three assessments (see graph) With the first script some individuals rated current health as low (RV < 40) but were nonetheless unwilling to take risks in the standard gamble (SG > 99) because the script assured them of remaining in their current state of health "forever without worsening As a result, the two scripts produced similar ratings using relative value scales to assess health, but exhibited significant differences when using scales that involved risk ($p < 0.01$ for SG and TTO)



All subjects completed the NIDDK Quality of Life survey Together, QOL and utility measures generate numerical values for every health state defined in ESLD and provide a means by which quality-adjusted survival can be evaluated These findings demonstrate the feasibility of using subjective utilities to measure QALYs in patients with progressive illnesses However, the scales must define an explicit time horizon since responses vary depending on whether or not patients consider long-term prognosis

COMPARED TO WHAT? HOW THE CALCULATION OF COST-EFFECTIVENESS RATIOS CAN CHANGE STUDY CONCLUSIONS

Kathryn A Phillips, Susan Fernyak, University of California, San Francisco

The purpose of this study is to present three case studies that demonstrate various approaches to calculating cost-effectiveness (CE) ratios and their impact on study results, conclusions, and practices and policies

Case Study #1 Some studies have found that mammography screening for women ages 40-49 is relatively cost-effective, while others have found that it is not cost-effective A key explanation for these discordant results is the choice of comparator program For example, Salzmann, Kerlikowske, Phillips (1997) calculated the incremental CE of screening women ages 40-49 compared to women ages 50-69 and concluded that mammography in younger women is relatively not cost-effective because of the delayed benefits of screening Conversely, Lindfors et al (1995) calculated the average CE of screening women ages 40-79 and concluded that this approach could be as cost-effective as screening of women ages 50-79

Case Study #2 Paltiel and Kaplan (1997) demonstrated that differential participation rates in HIV testing programs could greatly influence CE results These findings illustrate how misleading results can be obtained when using incremental CE ratios to compare the population currently receiving services to the population expected to receive services after an intervention For example, Eddy 1990 found that increasing Pap smear screening among high-risk, unscreened women was more CE than annual screening (vs every 3 years) for average-risk women Despite findings such as this, annual Pap smear of average-risk women continues to be recommended and practiced

Case Study #3 Phillips, Fernyak, et al (1998) found that two policy options, (1) routine, voluntary HIV testing of all new primary care patients or (2) using risk histories to identify individuals for counseling and testing had similar CE ratios when each was compared to current testing practices Yet, incremental analyses that compared the most effective option (routine testing) with the next most effective option (risk histories) showed that the option of risk histories was dominated through extended dominance (i e , it had a higher incremental CE ratio) and therefore it should be eliminated as an option These results illustrate that the choice of research question and the resultant calculation of CE ratios can change study conclusions

Conclusion The choice of comparator programs – a decision that is often not explicitly discussed in CEAs is often a complex decision that can greatly influence study results and conclusions

IMPROVING MATERNITY SERVICES IN THE FORMER SOVIET UNION USING DECISION MAKING TOOLS DOING MORE WITH LESS

BC Else and BA Else TEO Resource Managers, Lake Placid, NY, Merck-Medco Managed Care, L L C , Montvale, NJ

Identifying areas where practical and low cost improvements in the clinical process are possible and implementing those improvements is a common challenge in international development work The application of pragmatic decision making tools in international development projects can facilitate changes in health policy and promote sustainable improvements in core clinical processes and practices The purpose of this study was to evaluate the financial and clinical impact of integrating Family Centered Maternity Care (FCMC) into traditional maternity services offered in the Ukraine Traditional maternity services include long standing state mandated uniform health policy that often requires all patients in labour to receive the same procedures and processes of care regardless of evidence-based clinical information or patient preference

This study used decision tree and cost-effectiveness analyses to evaluate the costs of care and link this information to specific clinical indicators The methodology and approach was chosen not only because it was appropriate for the study questions but because it a) provided policy decision makers with more than one implementation path promoting implementation flexibility, b) positioned the on-site host country clinical team to continue the health services research at a local level, c) promoted more state-of-the-art decision making tools at a patient level, and, d) facilitated cross-cultural communication and language translation Clinical data were obtained directly from patient charts Cost data were obtained from the on site facility finance offices

Results showed variable cost savings ranging from $-2 50 to $ 7 00 per patient The FCMC approach resulted in a decrease in cesarean section rates with a corresponding decrease in anesthesia and pain medicine costs, a reduced rate of induced labours, increased levels of ambulation, an improvement in APGAR scores,and an improvement in patient satisfaction The study methodology and results promoted sustainability by demonstrating cost savings and suggesting opportunities for clinical improvements The decision model provided a tool suggesting several ways for health planners and clinical decision makers to integrate improved health processes and procedures as well as manage costs

COST SAVINGS FROM PHARMACEUTICAL THERAPEUTIC SUBSTITUTION: ESTIMATES FROM THE 1995 NATIONAL AMBULATORY MEDICAL CARE SURVEY

SS Rathore, WL Boyko Jr , KP Weinfurt and KA Schulman Clinical Economics Research Unit, Georgetown University Medical Center, Washington, DC

Pharmaceutical costs are a significant component of health care expenditure in the United States One method for controlling pharmaceutical costs is therapeutic substitution – the replacement of high cost pharmaceutical agents with clinically comparable, lower cost agents We sought to estimate potential cost savings associated with therapeutic substitution of eight often prescribed, high cost medications

We used data from the 1995 National Ambulatory Medical Care Survey (NAMCS) to evaluate prescription medication use in older Americans NAMCS is a nationally representative sample of ambulatory care visits in the United States and contains clinical information concerning each visit, including medication We evaluated patients 50 years or older for use of eight prescription medications A lower cost substitute agent of similar safety and clinical efficacy was then identified for each medication Therapeutic substitution savings were calculated based on a 30 day course of initial treatment Medication costs for initial and substitute agents were obtained from 1997 *Red Book* average wholesale prices Savings associated with therapeutic substitution for each indication are reported below

Initial[1] Sub[2]	Lozol HydroPar	Mevacor Lescol	Vasotec Lotensin	Prozac Zoloft	Glucotrol Glipizide	Procardia Diltiazem	Proscar Hytrin	Clinoril Ibuprof
Scripts[3]	1,232	3,316	2,168	2,109	5,689	3,209	691	985
Initial[4]	25,330	223,838	66,956	152,958	126,682	218,544	40,478	36,761
Sub[4]	647	120,375	43,596	136,061	104,706	124,195	26,363	19,815
Saving[4]	24,683	103,463	23,360	16,897	21,576	94,349	14,115	16,946

[1]Initial medication [2]Therapeutic substitute [3]Scripts reported in thousands [4]Costs in US$1,000

Cost savings range from $103 5 million in the Mevacor/Lescol therapeutic substitution to $14 1 million for Proscar/Hytrin Full therapeutic substitution offers potential cost savings in excess of $315 4 million

While therapeutic substitution may not be feasible in all cases due to differences in side-effect profile, dosing preferences or off-label use, preliminary estimates indicate a sizeable cost saving may be realized by the use of therapeutic substitution

HOW TO QUANTIFY PHYSICIAN CONSISTENCY WITH PRACTICE POLICIES? THE CASE OF ASTHMA MANAGEMENT

LC Cicutto, HA Llewellyn-Thomas University of Toronto, The Toronto Hospital and the Clinical Epidemiology Unit, Sunnybrook Health Science Centre, Toronto, Ontario.

To foster evidence-based practice, audits are used to evaluate the extent to which health professionals' diagnostic and therapeutic decisions are consistent with practice guidelines. However, the development and testing of a systematic approach to quantify the degree of consistency is particularly problematic when there are areas of clinical uncertainty ("gray zones") in the guidelines themselves The study aim was to use the context of asthma care -- in which there are notable gray zones -- to explore various approaches to gauging consistency.

163 Ontario physicians caring for asthma patients (response rate 81%) considered 6 asthmatic episodes ranging from mild to severe. For each episode, the physician selected a first priority action. Then each selection was compared to the Canadian asthma guidelines' recommended first line therapy, and scored for guideline consistency in 3 different ways. For all 3 scoring systems, a maximally consistent physician could obtain a total score of 6 However, the systems ranged from most to least stringent in terms of the degree of penalty ascribed to selections that fell into a gray zone. (Gray zones were operationally defined by the content of the guidelines themselves.)

Physicians' mean total scores ranged widely depending on the penalty rule (most stringent 2.19 ±2.3; moderate 3.45 ±1.9; most lenient 5 12 ±1.3). At the same time, all 3 systems were sensitive to sub-group differences; pulmonologist/family physician mean differences ranging from 2.3 to 1.7 to .58 were detectable (higher pulmonologist consistency; p< 0 004 in all 3 tests).

These results imply that, in this context, audits of practitioners' consistency with practice guidelines will be affected by the extent to which a scoring strategy not only allows for a full range of clinical severity but also takes the presence of gray zones into account Since misleading conclusions can be drawn from audits in which such issues are not addressed, the development of explicit, systematic approaches to quantifying the effects of uncertainty may be a matter of some urgency.

"REASSURANCE COST" MAY AFFECT ESTIMATES OF RESOURCE UTILIZATION IN DYSPEPSIA IF GASTROSCOPY IS INSUFFICIENT TO PREVENT FURTHER TESTING

AV Sahai, I Penman, D Williams, G Mishra, W Knapple, A vanVelse, RH Hawes, and BJ Hoffman Digestive Disease Center, Medical University of South Carolina, Charleston, SC PURPOSE Cost-analyses of dyspepsia management do not account for additional resource-utilization that may occur when a negative gastroscopy[EGD] is insufficient to reassure patients and/or physicians that no serious disease is present Therefore, we aimed to determine 1) What is the frequency of negative and/or non-contributory abdominal imaging tests in dyspepsia patients (pts)? 2) Do pts who undergo EGD have less additional testing than those who do not undergo EGD? 3) What is the cost of reassurance? METHODS All dyspeptic patients seen for endoscopic ultrasound completed a structured physician-administered questionnaire on previous dyspepsia-related negative and/or non-contributory testing Facility direct and indirect costs for barium meal [UGI] $165, abdominal ultrasound [US] $141, abdominal CT [CT] $236, and abdominal MRI [MRI] $376 were used to calculate the "reassurance cost"/patient RESULTS 121 pts studied to date (study ongoing) Dyspepsia types pancreatic 55%, non-specific 14%, biliary 13%, IBS 8%, ulcer 6%, reflux 4% Ninety-three(76%) had EGD, 29(34%) did not "EGD" and "no EGD" gps were not significantly different for age, sex, race, and dyspepsia type, but differed in the mean duration of symptoms (ΣD) (18+/-14mo vs 9+/-14mo, p=0 003) However, the # tests did not correlate with ΣD (EGD or no EGD pts) Total # tests/pt (including EGD) 5 tests 3%, 4 23%, 3 33%, 2 25%, 1 10%, 0 7% Frequency of tests in EGD and no EGD pts

	ALL (n=121)	EGD (76%)	NoEGD (34%)	p
Previous UGI	33 1%	38 0%	17 2%	0 04
Previous US	65 2%	68 5%	55.2%	NS
Previous CT	80 2%	84 8%	65 5%	0 03
Previous MRI	9 1%	8 9%	10 3%	NS
# tests (excluding EGD)	1 9+/-1 1	2 0+/-1 0	1 5+/-1 1	0 02

Mean "reassurance cost" per pt EGD pts $394, no EGD pts $298 CONCLUSIONS 1) Dyspepsia pts undergo multiple negative and/or non-contributory abdominal imaging tests -- with no relation to symptom duration 2) EGD pts actually had more tests than no EGD pts, suggesting that EGD is insufficiently reassuring to prevent further testing 3) On a large scale, "reassurance cost" may be substantial (since dyspepsia is a common problem) It should therefore be considered in economic analyses of dyspepsia management 4) In dyspepsia pts who are referred for EGD, an imaging test that can comprehensively examine the upper GI tract and surrounding structures in one sitting (e g endoscopic ultrasound) may reduce overall costs and resource utilization if its cost is sufficiently low and it prevents further testing

THE TRANSLATION PROCESS IN SURVEY RESEARCH A COST ANALYSIS

SB Cantor, T Byrd, JY Groff, Y Reyes, PD Mullen The University of Texas M D Anderson Cancer Center, The University of Texas-Houston School of Public Health, and The University of Texas-Houston Medical School, Houston, TX

It has been demonstrated that careful translation of documents is critical for research with populations whose primary language is not English To ascertain a major component of the costs of outcomes research on non-English-speaking populations, we conducted a formal cost analysis of a survey-instrument translation process

We conducted a process flow analysis of the steps required to create and validate a 30-page survey instrument in Spanish that already existed for English-speaking populations Following the translation procedure described in Marin and Marin (1991), the components included back-translation, translation by committee, "decentering," pre-testing on available personnel, and pilot testing on potential subjects The perspective taken in the cost analysis was from the perspective of the institution conducting the research Identification of and time required for steps in the process were based on interviews with research personnel Overhead costs were based on the institutional indirect cost rate (49 5%) as a proxy for the true costs of facilities and administration Sensitivity analysis was conducted to determine the robustness of the total estimate to variability in the parameters of the cost model

We identified 11 steps of the translation process The total cost of the translation process cost $10,426 The largest component of costs was the 272 personnel hours spent developing the survey, which accounted for 61% of monetary resources Research assistants' time accounted for 76% of total personnel hours but only 58% of personnel costs Overhead costs accounted for 33% of monetary resources, a direct consequence of the assumed institutional indirect cost rate Smaller costs included travel, phone/fax, and supplies The sensitivity analysis indicated that the results were robust to reasonable variation in the salaries of the investigators and time required to complete the components of the task, but were sensitive to the indirect cost rate

The methodology and results of this study will be useful for the preparation of future studies that involve populations whose native language is not English Investigators should be aware of the significant costs that are involved when budgeting for studies that require survey development in a second language

UTILITIES FOR THE EFFECTS OF DEPRESSION ON HEALTH STATUS

LA Lenert, C Sherbourne, and K Wells, M D, Stanford University School of Medicine, Stanford CA and the RAND Corporation, Santa Monica, CA

Background Relatively little is known about the significance of health changes caused by depression and the cost-effectiveness of programs that prevent or treat it Objective Develop a tool for mapping between measures of health status and utilities for depressive illness to facilitate the application of cost-effectiveness in this domain

Methods Using data from the Medical Outcomes Study (MOS) Depression panel, we identified a set of 6 principal health states summarizing the typical effects of depression on health status as measured my the Short Form 12 (SF-12) health questionnaire We measured the preferences of a group of depressed patients (n=149) for the 6 states and created a mapping function to convert between SF-12 measurements and utilities We applied this function to estimate utility weights for changes in health status with onset or remission of depression from one-year follow up data from the MOS study (n=266)

Main findings Patients standard gamble utility ratings for the six health states ranged from 0 946 (0 11 s d) to 0 667 (0 26 s d), visual analog scale ratings ranged from 0 90 (10 s d) to 0 47 (0 14 s d) Patients experiencing remission of depression on one-year follow-up had a gain of 9 0 in the SF-12 Mental health score (p< 0001) and no change in the Physical score, this corresponded to an increase in utility of 0 055 (p=0 002) Patients without initial depression who developed an new episode during the one-year follow up period had a net change of -5 0 in SF-12 Mental health score (p= 0 024), no change in Physical score and decrease in utility of 0 035 (p=0 076)

Conclusions Based on patients' preferences, treatment of occult cases of depression in general medical populations affords about a 35% greater gain in utility than prevention of incipient cases Both treatment and prevention of depression appear to were associated gains in utility that could support substantial resource allocations

USE OF AN ANGULAR TRANSFORMATION FOR RATIO ESTIMATION IN COST-EFFECTIVENESS ANALYSES

JF Heyse, JR Cook Merck & Co , Inc , Blue Bell PA

Purpose. Several methods are now available to estimate the cost effectiveness ratio (CER) and its 95% confidence interval (CI) These methods typically have difficulty when the incremental effectiveness (health outcome) is small, resulting in extreme instability in the CER estimate

Methods. We propose using an angular transformation of the standardized mean difference in cost and effects to stabilize the variance of the CER The average cost-effectiveness angle and its 95% clockwise and counter-clockwise limits can be estimated using either the jackknife or bootstrap resampling method An inverse transformation was applied to the angle and its limits to convert back to the original units of the ratio A simulation study was conducted to assess statistical properties of the method The method was applied to data from a large clinical trial to estimate the incremental cost per additional survivor

Results. The estimated cost-effectiveness angle using the jackknife method was unbiased and consistent Lack of coverage in the 95% CI was close to 5% (range 3 6 to 7 0%) In the analysis of the clinical trial data, use of the angular transformation reduced the bias in the estimated CER based on the bootstrap 95% CIs were similar for the two resampling methods (bootstrap -66,908 to 8,197, jackknife -63,382 to 11,000)

Conclusions. The jackknife and bootstrap methods for estimating the CER and its 95% CI perform well when an angular transformation is used

A DYNAMIC PROGRAMMING APPROACH TO THE EFFICIENT DESIGN OF CLINICAL TRIALS

K Claxton* and KM Thompson Department of Economics, University of York, UK and Harvard School of Public Health, Boston, MA * Supported by the Commonwealth Fund of New York

Objectives To estimate the net benefit of sample information in sequential clinical decision problems, and to establish optimal sample size and optimal allocation of trial entrants in technically efficient research designs

Methods The equal allocation of trial entrants to alternative arms of a clinical trial (although often used) has little justification There is a body of literature addressing optimal allocation in sequential clinical trials but here the more fundamental problem of optimal allocation in a fixed sample design was considered The value of sample information, optimal sample size and the allocation of patients at each stage must be established before any sample information is available A Bayesian approach to the value of sample information was combined with dynamic programming and applied to numerical examples of sequential decision problems The results when optimal patient allocation was used were compared to using an equal allocation rule

Results The dynamic programming approach to optimal patient allocation provided higher expected net benefits of sampling in all circumstances, and optimal sample size and sample allocation differed considerably from the analysis assuming equal allocation In some circumstances it was not efficient to allocate any of the sample to a particular arm of the trial and in these cases the strategy could be safely ruled out as irrelevant The full enumeration of every feasible allocation of every sample size considered in a sequential decision problem would require A^S estimates of net benefit where A=((N+1)2+(N+1))/2, N is the largest sample considered and S is the number of stages In a four stage decision problem (N=500) it would take 792 billion years to complete (at 10 estimates per second) The Dynamic programming approach provided a feasible solution by reducing the number of alternatives to A*S or under 14 hours of computing time

Conclusions The analysis demonstrates that Bayesian approaches to the value of information can be applied to more complex sequential decision problems It shows that optimal allocation is a fundamental design problem and not simply an issue for sequential clinical trials Arbitrary fixed allocation rules will lead to technically inefficient research design, the expected net benefits of sampling will be underestimated and there will be a danger that cost-effective research proposals may be rejected Finally this approach provides a consistent way to define and identify which of the competing alternative strategies are relevant and should be included in any evaluative study design

FACTORS INFLUENCING THE DECISION TO BUY SUPPLEMENTARY HEALTH INSURANCE

DRM Timmermans, PP Wakker, I Machielse Dept of Social Medicine, Vrije Universiteit Amsterdam, The Netherlands, Leiden University Medical Center, The Netherlands, Zorg & Zekerheid Health Insurance, Leiden, The Netherlands

Background Restriction of health care expenses is becoming increasingly important in many countries In the Netherlands, the introduction of compulsory contributions for the use of health care services is being considered, so as to reduce inefficient use of health care services and to enhance people's awareness of expenses Insurance companies are offering their clients the opportunity to buy supplementary insurance to cover this risk Clients now face the decision whether they should buy this supplementary insurance

Purpose To determine whether expected health care consumption, psychological factors and/or demographical factors influence people's decision to buy supplementary health insurance

Methods Intention to buy supplementary insurance was asked to N=474 clients of a Dutch health insurance company In addition, they were asked to estimate their expected health care consumption of the forthcoming year, to estimate their relative chance of health care consumption as compared with people of the same age and sex, and to indicate their perceived control of health, worry about health, subjective health perception, anticipated regret when not buying supplementary insurance, and risk attitude They were also asked to provide demographical information about income, education level, and family size

Results 52% of the participants indicated to prefer buying supplementary insurance, 16% indicated no preference and 32% indicated they would not buy supplementary insurance Multiple linear regression analysis shows that 41% of the variance of the reported intention to buy supplementary insurance could be explained by all factors, seperate analysis of the 3 categories shows that only 7% could be explained by expected consumption, 22% by the psychological factors, and only 8% by demographical factors The most important psychological factors are relative chance of health care consumption, subjective health perception , risk attitude and especially anticipated regret The most important demographical factors are family size and educational level

Conclusions Clients' (intended) decision to buy supplementary health insurance is mostly based on psychological factors and not on expected health care consumption or income It is therefore doubtful whether the introduction of compulsary contributions will affect people's attitude towards the use of health care services

COMPARING FACILITIES ON QUALITY OF CARE A HIERARCHICAL BAYES LINEAR MODELING APPROACH

BH Chang, JJ Anderson, AK Rosen Center for Health Quality, Outcomes, and Economic Research, Bedford VA Medical Center, Bedford, MA

The importance of case-mix adjustment in comparing facilities on quality of care is well-recognized in health services research However, statistical techniques for making case-mix adjusted facility comparisons are still in developmental stages Therefore, developing and adapting more sophisticated techniques is needed In this study, we demonstrate the use of a Hierarchical Bayes Linear Model (HBLM) for such comparisons

Administrative inpatient and long-term care databases from the Department of Veterans Affairs (VA) are used for analyses We are interested in comparing all 133 long-term care facilities on 6-month mortality rates of residents who were present on April 1, 1996 The facility case-mix adjusted mortality rates are estimated by a regression model which adjusts for individual patient's medical conditions and treatments The difference between observed and expected (case-mix adjusted) mortality rates is used as the outcome We model this outcome and estimate its posterior distribution using the HBLM

Results from the HBLM analysis show that facilities are not homogeneous in mortality rates even after case-mix adjustment The between facility variance (τ^2) is significantly different from zero ($p< 0001$) Facilities whose 95% central posterior interval of the posterior mean does not include zero are identified as outliers (those with significantly higher or lower than expected mortality rates)

We conclude that the Hierarchical Bayes Linear Model is useful in identifying outlier facilities Future studies can utilize this model to identify facility-level characteristics explaining variation among facilities after case-mix adjustment

DECISION ANALYSIS AND PRACTICE GUIDELINES IN THE SETTING OF LOW-QUALITY EVIDENCE DEVELOPMENTAL DYSPLASIA OF THE HIP (DDH)

HP Lehmann, MJ Goldberg Johns Hopkins University, Baltimore MD, and Tufts University, Boston MA

Objective DDH is an important morbidity in infants Universal screening by ultrasound (US) has been recommended Is this recommendation evidence based?

Method· An influence diagram of 31 nodes was created, representing the universe of screening strategies for DDH in the first year of life 613 articles were gleaned from the literature and review of ancestor articles, 101 were deemed useful, only 4 of which were comparative, prospective studies The articles generated 200 evidence tables for a minority of arcs in the diagram, including risk factors (gender, obstetrical presentation, and family history) All evidence was ascertained by two readers and discussed in committee Bayesian hierarchical meta-analysis was used to integrate evidence for individual arcs Strategies were valued as to the number of expected excess neonatal treatments, avascular necrosis (AVN) of the hip, and post neonatal cases Marginal "cost"-effectiveness ratios (MCER) were calculated, using AVN or neonatal cases in the numerators, and total late cases in the denominator Uncertainty ranges were calculated through monte carlo simulation Variable importance was calculated as the product-moment correlation of the variable with an MCER

Results The meta-analytic estimates for positive screening rates for pediatric, orthopaedic, or US screening were 8 5, 11 5, and 25 per 1,000 children screened, respectively Female breech infants had the highest absolute risk of 120/1,000 The uncertainties in MCERs are so large and on both sides of zero, that no conclusion can be made The MCERs were most sensitive to the current knowledge about AVN rates, followed by the neonatal orthopaedic and US rates

Conclusion The current evidence cannot support one strategy over another We need further (better) data regarding AVN rates and neonatal orthopaedic and US rates

SETTING PERFORMANCE STANDARDS FOR A NATIONAL HIV PREVENTION PROGRAM

DR Holtgrave, SD Pinkerton Centers for Disease Control and Prevention, Atlanta, GA, Center for AIDS Intervention Research, Medical College of Wisconsin, Milwaukee, WI

Purpose Policymakers responsible for national allocations of resources across health service areas (e g cancer, heart disease, and HIV prevention services) need to know whether their mixture and levels of investments yield optimal health benefits Important information for them is whether or not their investment of fiscal resources for a given health service (e g , HIV prevention) could be considered cost-effective or even cost-saving to society We sought to determine the number of HIV infections that would have to be prevented in order for the $616,790,000 (fiscal year 1997) investment in the national HIV prevention program funded by the U S Centers for Disease Control and Prevention (CDC) to be considered cost-effective or cost-saving to society

Methods Standard methods of economic evaluation and threshold analysis were employed A societal perspective and a 3% discount rate were used The cost-saving threshold number of HIV infections that would have to be prevented by the program was estimated by dividing the program budget by the discounted lifetime medical costs of treating one case of HIV disease The cost-effective threshold number was calculated by dividing the program budget by these medical costs plus the product of the number of quality-adjusted life years lost to one case of HIV infection times the societal willingness to pay to save one quality-adjusted life year

Results The CDC-funded program (for fiscal year 1997) could be considered cost-saving if 3,995 HIV infections were averted (in other words, the CDC-funded program would have to modify the course of the epidemic in the U S by about 9% -- relative to the epidemic in the absence of the national program -- in order to be cost-saving to society) The program could be considered cost-effective if 1,255 infections (3% of the total) were prevented

Conclusions It is possible to use economic evaluation techniques to rigorously set performance thresholds for a national HIV prevention program The results obtained for the national CDC-funded HIV prevention program are very attainable, they suggest that if (relative to the absence of the national program) the CDC-funded efforts avert only a small fraction of the HIV infections that would otherwise occur then the program is cost-saving (or cost-effective) to society

ESTROGEN THERAPY AND WEIGHT GAIN

Diane M Nederostek, IMS America

The purpose of the study was to examine the effects of estrogen replacement therapy (ERT) on body weight noted in post-menopausal women

The study was a retrospective, non-experimental study with a quantitative correlational design Its sample was an availability sampling of all women over the age of 34 years whose medical records were contained in the IMS America LifeLink™ Medical Record Solutions Database Body weight measurements taken three months prior to estrogen therapy and six to thirteen months post-therapy were evaluated for a test group of 735 women Similarly, weight changes recorded for a control group of 46,149 women in the same time period were examined Changes in body weight from each of these groups were tested for statistical significance using chi square and an independent samples t-test

Results of the chi square analysis showed no statistically significant relationship between estrogen therapy and an increase in body weight Similarly, the results of an independent samples t-test revealed no statistically significant difference in weight change between the test group (women on ERT) and the control group (women not taking ERT supplements) In general, the women taking estrogen supplements experienced an average weight gain of 0 4 pounds while the women not on therapy experienced an average weight loss of 0 2 pounds

These findings suggest that estrogen replacement therapy has no impact on the increase in body weight often experienced by post-menopausal women As a result, the supplements can continue to be prescribed to women for the alleviation of menopausal symptoms, as well as the provision of therapeutic benefits relative to cardiovascular disease and osteoporosis, without the complication of weight gain

COST-UTILITY ANALYSIS OF DETECTING AND TREATING DIABETIC RETINOPATHY IN DIABETIC PATIENTS IN JAPAN

I Ohsawa. Research Center of Health, Physical Fitness and Sports, Nagoya University, Nagoya, JAPAN

PURPOSE: To assess the efficiency of early detection and treatment of retinopathy in non-insulin dependent diabetic patients (NIDDM) in terms of their quality of life

METHODS: A Markov decision analytic model was constructed to compare four detecting strategies complete dilated eye examination by ophthalmologists at 6-, 12-, 24-, 36-month intervals. The clinical effect on quality of life was expressed as quality-adjusted life years (QALYs). Incorporated utilities for no-retinopathy, nonproliferative retinopathy, proliferative retinopathy or macular edema, visual loss, and death were 1.0, 0.99, 0 9, 0 6, and 0 0 respectively The costs were estimated from medical charges paid by the health insurance system in Japan and it included eye examination for detection, photocoagulation and vitrectomy for prevention and treatment of visual loss. Incremental cost-effectiveness ratio (incremental C/E) was calculated to evaluate the efficiency of four strategies Because lack of available data in Japan, we used clinical and epidemiological data from both USA and Japan All represented data were not discounted

RESULTS

Strategies	Costs (US$)	Effectiveness (QALY)	Average C/E (US$/QALY)	Incremental C (US$)	Incremental E (QALY)	Incremental C/E (US$/QALY)
24-month	7220	26 89	269			
12-month	7304	27 01	270	84	0 12	720
36-month	7431	26 78	277	127	-0 22	dominated
6-month	8071	27 07	298	767	0 06	12274

CONCLUSIONS. Although the periodic examination of detecting diabetic retinopathy by ophthalmologists reduces visual loss and improves quality of life in individuals with NIDDM, the incremental cost-effectiveness ratio of 6-month interval strategy is very high compared with 12-month interval strategy The estimation not only from insurer's viewpoint but also from societal viewpoint is necessary to the appropriate evaluation of the efficiency

STAKING A CLAIM DO ADMINISTRATIVE DATA STAND UP TO CLINICAL REVIEW?

DA Ollendorf, AM Fendrick, G Oster Policy Analysis, Inc , Brookline, MA, and Consortium for Health Outcomes, Innovation, and Cost-Effectiveness Studies (CHOICES), Department of Internal Medicine, University of Michigan, Ann Arbor, MI

While estimates of disease incidence are often based on health-care claims data, only infrequently are they validated against clinical records In this study, we sought to examine the sensitivity of a definition of septicemia that was developed for use in health services research by the Agency for Health Care Policy and Research (AHCPR), using data for a sample of patients for whom both detailed clinical and administrative data were available

The study sample consisted of 71 hospitalized patients participating in a clinical trial for severe sepsis of presumed infectious origin for whom a Uniform Bill (UB-92) also was available All patients had severe sepsis defined by (1) microbiological isolation of infectious organism, or organ infection diagnosed by clinical criteria, or likelihood of infectious etiology, (2) hyper- or hypothermia, (3) tachycardia, (4) tachypnea or mechanical ventilation, and (5) hypotension or evidence of systemic toxicity We sought to ascertain whether these patients also met the AHCPR definition of septicemia, as reflected in the ICD-9-CM diagnosis codes on their UB-92 bills

A total of 49 patients (69%) had one or more of the AHCPR-specified diagnosis codes for septicemia Examination of the final hospital bills for the 22 patients who did not have such a code revealed no clear patterns that would explain the omission of the disease-specific codes

Our findings suggest that incidence estimates based on the AHCPR definition of septicemia may substantially underestimate the actual magnitude of this problem Precise estimates of the burden of illness are critically important for resource planning, surveillance, and allocation of research funds A discrepancy of the magnitude observed in this study suggests a need for routine validation of definitions of clinical conditions that are employed in research involving administrative data

IMPROVING RATIONAL DECISIONS BY PROVIDING RISK INFORMATION

PP Wakker, DRM Timmermans, I Machielse Leiden University Medical Center, Leiden, The Netherlands, EMGO Free University, Amsterdam, The Netherlands, Zorg en Zekerheid, Leiden, The Netherlands

Background Restriction of health care expenses is becoming increasingly important in many countries In the Netherlands, the introduction of deductibles for national health insurance is being considered, so as to reduce inefficient use of medical care and to enhance people's awareness of expenses Clients will now face a decision concerning the level of deductible to choose and the possible buying of supplementary insurance

Purpose Investigate for various forms of risk-information if they improve rational decisions and customer satisfaction

Method Demand for supplementary insurance was asked to N=474 customers before and after the receipt of risk-information Various characteristics, such as degree of risk aversion, were measured Descriptive and psychological results will be reported elsewhere (Timmermans et al , in preparation) Here we analyse normative implications, i e , to what degree do various forms of information improve rational decisions and social welfare

Results

(1) Information about individual expenses, specified for various health services, had most effect
(2) Not surprisingly, risk averse clients want more insurance than risk seeking clients
(3) Surprisingly, risk information enhances demand for insurance among risk averse clients and decreases it among risk seeking clients This is opposite to regression to the mean
(4) Customer satisfaction was improved
(5) Risk information enhances demand for insurance among clients with high expenses and decreases it among clients with low expenses ("adverse selection")

Discussion (3) and (4) are normatively desirable (5) is desirable from the client's individual perspective In the context of insurance, it is unwarranted from the societal perspective

Conclusion The decision whether or not to provide risk information requires a tradeoff between (5) versus (3), (4), and another advantage reduction of health expenses due to people's increased awareness

ECONOMIC EVALUATION OF DONEPEZIL FOR TREATMENT OF ALZHEIMER'S DISEASE IN CANADA

B O'Brien, R Goeree, M Hux, M Iskedjian, G Blackhouse, S Gauthier, M Gagnon Dept Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Innovus Research Inc, Centre for Studies in Aging, McGill University, Canada

Background Donepezil (Aricept) is a new drug for the treatment of Alzheimer's disease (AD) We estimated the cost-effectiveness of donepezil 5mg daily as an adjunct to usual care in the management of persons with mild-to-moderate AD defined as a Mini-Mental Health State Examination (MMSE) score in the range 10-26

Methods Treatment effect data as MMSE change-over-baseline scores were obtained from a 30-week placebo-controlled trial of donepezil MMSE scores beyond observed trial data were estimated using a Markov model with 10 cycles of 24 weeks based upon the placebo MMSE progression observed in the trial Data from AD subjects in the Canadian Study of Health and Aging were used to estimate costs of nursing home care, community services, medications and caregiver time as a function of MMSE score A clinic-based cohort study from Alberta was used to estimate the distribution of AD patients by MMSE score presenting for treatment The effectiveness measure for the economic model was expected time (over 5 years) spent with non-severe AD (MMSE □ 10)

Results Over 5 years of treatment, donepezil is predicted to reduce health care costs by CA$929 per patient but increase caregiver time costs by CA$48 per patient for an overall cost saving to society of CA$882 per patient Patients not receiving donepezil are predicted to spend 2 21 years of the 5 years in non-severe AD compared to 2 41 years for treated patients (a gain of just over 2 months) Sensitivity analysis reveals that cost savings per patient increase if more AD patients are assumed to survive to 5 years, however, if donepezil treatment continues when patients' MMSE falls below 10, the incremental cost is higher for treatment at CA$1,554 per patient

Conclusion Under a wide range of assumptions, the use of donepezil for mild-to-moderate AD in Canada is predicted to be economically dominant over the alternative of no donepezil, with lower 5-year costs and less time spent with severe AD

A RANDOMIZED TRIAL OF A DECISION AID FOR BREAST CANCER SURGERY

V Goel, C Sawka, E Thiel, E Gort, A O'Connor Institute for Clinical Evaluative Sciences, Sunnybrook Health Science Centre, University of Toronto, Toronto, Ontario, Canada

The choice of type of surgery for early breast cancer is a difficult decision that is made in an emotionally charged period between diagnosis and surgery Lumpectomy coupled with radiation therapy and mastectomy offer equivalent survival outcomes Variations in breast cancer surgery rates have suggested that patient choice factors may be critical Thus, decision aids are ideal in this setting

A randomized trial of a decision aid compared with an information pamphlet was conducted The decision aid consists of an audiotape and workbook, completed in thirty minutes The decision is based on clinical practice guidelines for breast cancer surgery and was developed after extensive focus group and pilot tests It includes explicit photographs of outcomes, graphical representation of probabilities, and a values clarification exercise The information content in the control pamphlet was identical Non-academic surgeons in office-based practice were cluster randomized to deliver one of the two interventions to women newly diagnosed with breast cancer Major outcome measures were the Breast Cancer Knowledge Questionnaire, Decisional Conflict, anxiety (STAI), and decisional regret Follow-up was before surgery and at six months

Thirty-eight surgeons enrolled 164 patients into the trial The two groups had comparable characteristics Complete follow-up is available on 136 patients, with no differences in the baseline characteristics of those who completed follow-up There are no differences between the decision aid and pamphlet groups in terms of knowledge and anxiety There were lower decisional conflict in the decision aid group, 1 98, compared to 2 08 in the pamphlet group (p=0 234) There were no differences in decisional regret and at the six month follow-up Qualitative comments from participants and surgeons suggest that they much preferred the decision aid

These negative results could be due to several reasons The study was designed to observe an effect size of 0 4 in decisional conflict The actual effect seems to be much smaller, about 0 2 and thus a larger power may be required to observe this smaller, less clinically significant effect Since the qualitative comments suggest that the decision aid was well received, perhaps other outcome measures are required Finally, we must consider whether a decision aid actually adds to decision-making beyond the information available to the control group in this context

A METHODOLOGY FOR EVALUATING CROSSING RECEIVER OPERATING CHARACTERISTIC (ROC) CURVES

RM Centor University of Alabama at Birmingham, Birmingham, AL

ROC analysis has become a primary tool for evaluating diagnostic tests Statistical techniques for comparing ROC curves assume that one ROC curve dominates another ROC curve Domination occurs when all true positive rates (TPRs) for a particular ROC curve curve are greater than all TPRs for a second curve at all false positive rates, i e , at all points along both curves However, at times domination does not occur In this case, the ROC curves cross The presence of significant crossing precludes the use of standard comparison methodologies

We can assess ROC curve crossing by visual inspection, however, we do not know the importance of such crossing Therefore, a mathematical standard for assessing ROC crossing, such as follows, will help investigators Using a parametric formulation of a ROC curve, one obtains a straight line in gaussian space Then one compares the two straight-line formulas and assess exactly where those two ROC curves cross They will always cross unless the slopes are identical Most pairs of ROC curves cross at false positive rates either less than 0 05 or greater than 0 95 If the crossing occurs at these extremes, we consider this insignificant crossing Here, one may perform standard statistical comparisons If the curves cross between 0 05 and 0 95, one then compares the slopes If the slopes are statistically indistinguishable, then one can not say that an important crossing has occurred, and, again, standard statistical comparisons can be made Only when the crossing occurs between 0 05 and 0 95 for the false positive rate and the slopes are different, should we refrain from statistical comparison of the ROC curves

THE EVALUATION OF AN INTERACTIVE COMPUTER PROGRAMME ON TREATMENT OPTIONS IN EARLY STAGE BREAST CANCER PRELIMINARY RESULTS

S Molenaar, MAG Sprangers, P Oosterveld, JC de Haes Academic Medical Center, Dept of Medical Psychology, Amsterdam, The Netherlands

Background An interactive computer programme (on CDROM) was developed for women with early stage breast cancer In order to facilitate patient participation in decision making, this programme provides written and audio-visual information on the treatment options available to patients

Study purpose To examine the acceptability of the interactive computer programme and its effect on patients' treatment preference and decision uncertainty

Method Patients were eligible if diagnosed with stage I or II disease All consecutive patients from three medical centers were invited to take part in the study A simple pretest-posttest design was used Data were collected through structured self-administered questionnaires Treatment preference was assessed with a measure developed for this study Decision uncertainty was assessed with the first subscale of the Decisional Conflict Scale (O'Connor, 1993) Patients' evaluations of the programme (e g , the level of clarity and usefulness, would you recommend it to others, evaluation of interactive method etc) were measured with a post-viewing evaluation form which was developed for this study

Results Ninety-six patients were included (mean age 55, range 33-85) At the pretest 9% 'definitely preferred mastectomy', 14% 'probably preferred mastectomy', 19% felt 'unsure', 21% 'probably preferred breast conservation therapy', and 38% said they would 'definitely prefer breast conserving therapy' At the posttest these proportions were 13%, 5%, 9%, 22%, and 51% respectively The shift in treatment preferences was statistically significant (p< 002) Treatment preferences were not related to age or education Posttest levels of decision uncertainty were significantly lower after using the interactive programme (p <001) Patients' evaluations of the interactive programme were highly positive

Conclusion The interactive programme was highly valued by patients It was found to be helpful in the selection of a treatment Treatment preferences shifted towards breast conservation therapy Uncertainty with regard to the decision decreased The effect of the programme on the decision finally made, decisional conflict, decision satisfaction and quality of life are yet to be established

A PILOT STUDY OF A DECISION AID FOR PATIENTS WITH LOCALLY-ADVANCED NON-SMALL CELL LUNG CANCER (LA-NSCLC)

M D Brundage, R H Cosby, D Feldman-Stewart, R. Gregg, P Dixon, Y Youssef, W J Mackillop The Radiation Oncology Research Unit, Queen's University, Kingston Regional Cancer Centre, Kingston, Ontario

The purpose of this study was to evaluate a decision aid designed to help eligible patients with LA-NSCLC choose between (a) radiotherapy (RT) alone, and (b) combined modality therapy (CMT) with Cisplatin-based chemotherapy and radiotherapy Each patient completed a one-hour interview designed to complement the medical consultation The aid included a structured description of the treatment options and trade-off exercises designed to help clarify the patient's values for the small median and three-year survival advantages conferred by CMT in light of its risks The trade-off exercise determined the survival advantage threshold, that is, the increase in survival conferred by CMT over RT that the patient deemed necessary for choosing CMT

Twenty patients met the eligibility criteria for study Of these, 7 declined the decision aid because they had a clear treatment preference (4 chose CMT and 3 chose RT) The remaining 13 participants completed the decision aid, 11 chose CMT and 2 chose RT Eleven of the 13 wished to be involved in the decision All patients reported that using the decision support was useful to them and that others should use the support No patient or physician reported that the aid interfered with the physician-patient relationship Patients' thresholds for 3-year survival ranged from < 0% to 45%, 11 of 13 patients had a threshold ≤ 5% Patients' thresholds for median survival ranged from zero months to > 24 months, 8 had a threshold ≤ 6 months For all patients, either the 3-year or median survival threshold was consistent with their final treatment choice Regarding the aid's impact, 5 patients showed a treatment preference shift, all towards CMT (in 3 of the 5 from no initial preference for either treatment) Patients' understanding of the treatment benefits increased modestly - none knew the chance of 3-year survival with RT and 25% knew the chance of 3-year survival with CMT prior to the aid In follow-up, only 63% knew the exact chances of survival with each treatment No patient reported an increase in decisional conflict, and 65% reported an improvement in decisional conflict or increased satisfaction

We conclude that the decision aid for LA-NSCLC patients is acceptable both to patients and to physicians Patients' preferences are split between the treatment options, the aid appears to help patients select a treatment that is most consistent with their values It may also decrease decisional conflict and increase satisfaction with the decision Further evaluation of the aid in a larger study is warranted

PREFERENCES FOR CARDIOPULMONARY RESUSCITATION DIFFERENCES IN PERSPECTIVES OF DOCTORS, NURSES, PATIENTS AND THEIR PARTNERS

R. de Vos, P M M. Bossuyt, R J de Haan, R W Koster MD PhD, J C J M de Haes, L J. Krol Academic Medical Center, University of Amsterdam The Netherlands

Background Differences in preference for resuscitation between participants in resuscitation decisions can be due to different expectations about the survival, but also to diverging value judgements about the prospects without cardiac arrest

Methods Doctors, nurses, patients and their partners estimated the general survival probability after resuscitation They also evaluated 16 paper case descriptions of patients with different prospects, and rated their preference for resuscitation in each case Factors contributing to the preference score were analysed through multivariate regression techniques

Results A total of 448 persons participated in this study (122 doctors, 173 nurses, 106 patients and 47 partners) 37 percent of the doctors, 20 percent of the nurses, 9 percent of the patients and 11 percent of the partners realistically expected the survival probability after resuscitation to be 10-20 percent Expectations about survival significantly influenced the preference for resuscitation In case of a short life expectancy without cardiac arrest, doctors were more inclined towards no resuscitation than patients, whereas depression and social isolation were more important for patients and their partners The preference score of the 4 groups was similarly influenced by aspects of pain and loss of self care

Conclusions Preference for resuscitation is influenced by expectations about survival after resuscitation and complex patterns of value judgements of the prospects without cardiac arrest Participants in decisions need to be informed about the realistic outcomes after resuscitation Even then, a misunderstanding of the patients' preference may occur Early discussions with patients are required

IS IT POSSIBLE TO INFER EFFECT DIFFERENCES OF DRUG A OVER DRUG B FROM SEPARATE PLACEBO CONTROLLED TRIALS OF DRUG A AND DRUG B?

E Trindade, C Coloma, D Menon Canadian Coordinating Office for Health Technology Assessment (CCOHTA), Ottawa, GIRAME, Montreal, Institute of Pharmaco-Economics (IPE), Edmonton, Canada

Purpose: To compare competing pharmacotherapies for major depression using meta-analyses

Introduction: Comparison of randomized placebo-controlled trials to show efficacy and safety of new drugs may give an early estimation of the quantity of benefit (or not) that health care systems and patients may (or may not) expect to receive from the introduction of a new technology

Methods: Various biomedical, bibliographic databases were searched to identify all randomized placebo-controlled and head-to-head trials of *serotonin uptake inhibitors (SSRIs)* or *tricyclic antidepressants (TCAs)*, published between 1990-May 1996 The effect size for each drug was estimated based on the difference in symptom scores measured using the Hamilton Depression Rating Scale before and after treatment Comparability of placebo and drug treated groups across the trials was tested by similarity of demographic and clinical inclusion and exclusion criteria, and statistical homogeneity of outcome measures The effect size difference between each drug and placebo, or the two types of drugs was then estimated These individual trial estimates were then used to calculate the pooled effect size difference, using Bayesian meta-analysis, for the following three groups (A) "new" (SSRI) vs placebo, (B) "old" (TCA) vs placebo and (C) "new" vs "old" The pooled value for the "old" drug was then subtracted from that for the "new" drug ((A) - (B)) and compared to the pooled value for the head-to-head trials (C)

Results: In 51 individual "trials" of SSRIs against placebo, the estimated pooled effect size differences over placebo was 0 55 (95% confidence interval 0 40 to 0 70) This was a statistically significant result indicating a moderate mean antidepressant effect In 32 trials of TCAs versus placebo, the mean pooled effect size difference was 0 56 (95% confidence interval 0 40 to 0 73) also a statistically significant result indicating a very similar, moderate mean antidepressant effect The meta-analysis of 96 randomized trials of SSRIs vs TCAs yielded a pooled head-to-head trials effect size difference of -0 013 between SSRIs and TCAs (95% confidence interval -0 10 to 0 07) The difference between ((A) - (B)) and (C) was -0 01

Conclusion: In early randomized, placebo-controlled trials, when comparability of placebo groups can be assumed, estimates of effect size differences obtained with a "new" drug over placebo may be compared with a similar set of studies of comparable existing technologies This may allow early inference of the incremental (or otherwise) benefit (or not) for the introduction of new technologies

INCOME EFFECTS OF REDUCED HEALTH AND HEALTH EFFECTS OF REDUCED INCOME IMPLICATIONS FOR QALY MEASUREMENT

Mark Sculpher, Bernie O'Brien University of York, UK, McMaster University, Canada

There is increasing use of multi-attribute health state utility systems such as the Health Utilities Index (HUI), the Quality of Well-Being (QWB) scale and the EuroQol (now EQ-5D) Proponents argue that data from such systems can be used in cost-effectiveness analysis (CEA) to calculate using quality-adjusted life years (QALYs) in a way that is unrelated to income However, income can influence QALY measurement at two stages raters' income when they value health states and patients' income when they are categorized to health states We assess the implications of these two income effects for the measurement of QALYs using health state valuation systems

Categorization within a valuation system will primarily reflect the direct effects of a patient's health However, certain domains - in particular, mental health and social function - are likely to reflect the effects of income loss due to illness We argue that these health effects of reduced income are not separable from primary health effects At the valuation stage, descriptions focus on levels of health, but raters may make inferences about the effects of reduced health on income and hence on general wellbeing In principle, these inferences can be removed by asking raters to assume income would remain unchanged in a given health state We have contacted senior researchers associated with the most popular valuation systems currently used in CEA Only the HUI has actively ruled out income inferences at valuation, with the other instruments ambiguous in this respect The US Panel on Cost-Effectiveness has advocated that these income effects should be actively ruled in to reflect the productivity costs incurred by the individual, despite the absence of valuation instruments to achieve this If income inferences are ruled in, one source of double counting can be avoided by removing this part of productivity costs from the numerator However, another source of double counting exists if patients categorization of their health (e g social function) partly reflects limitations in their consumption due to income loss We conclude (1) health state valuation systems should actively rule in or out income inferences at valuation - only one instrument currently does this, (2) ruling in income inferences will make instruments less useful outside the location from which the raters were drawn, (3) the risk of double counting if income inferences are rules in can be avoided if greater detail is given to the rater about which health domains are affected by income loss

FUZZY CLUSTERING APPROACH TO HEALTH OUTCOMES PREDICTION

K Tabata, S Ikeda, N Ikegami, C Hamashima, K Yoshida Tokyo Institute of Technology, Tokyo Keio Univ Tokyo, St Marianna Univ Kawasaki

According to the standards, diagnosis should be relatively straight forward, based on clear cut-off points However, because there is a degree of uncertainty, this process is not necessary so simple, especially in subjects coming for screening We believe that the fuzzy set theory is an effective method in dealing with this situation

Object Applying fuzzy clustering to predict onset of hypertension from screening data

Methods In this study, the data are cross sectional and consist of 3,169 subjects, 1,721 men (mean age 57 6 ±8 9 yrs) and 1,448 women (mean age 54 6 ±8 7 yrs) Of this total, after a follow-up period of 10 years 10% were diagnosed with hypertension To increase the discriminatory function of each variable the data were converged by using fuzzy membership function Five independent variables [diastolic blood pressures, systolic blood pressures, uric acids, FBS, and high alcohol intake] were chosen in the order of their prevalence among the subjects Fuzzy clustering was compared with two other methods discrimination analysis and conventional cluster analysis for predicting the onset of hypertension

Result The prediction of the onset of hypertension was 91 3% with fuzzy clustering, 76 5% with discrimination analysis and 21 1% with conventional cluster analysis For the fuzzy clustering three clusters were obtained Cluster 1 those with high blood pressure Cluster 2 those with borderline hypertension with obesity tendencies high alcohol intake Cluster 3 those having abnormal values on all three aspects The onset of hypertension for each of the Cluster Groups was as follows 12% for Cluster 1 13% for Cluster 2, and 78% for Cluster 3

Conclusion Fuzzy clustering appears to best predict the onset of hypertension However, since the membership function was based on the observer s subjective view more analysis using other chronic diseases should be conducted and the methodology developed to obtain an objective membership function

COST-EFFECTIVENESS OF COLON CANCER FOLLOW_UP A MODELLING APPROACH

Kievit J, Eijkemans MJC, on behalf of the Leiden-Rotterdam TAO-group
Dept of Medical Decision Making, Leiden University Medical Center, and Center for Clinical Decision Analysis, Erasmus University Rotterdam, the Netherlands

Introduction: Patients who have been operated for colorectal cancer undergo follow-up with the aim to detect recurrences as early as possible, and thereby improve the resectability and post-treatment prognosis of these recurrences The rationale of this approach has not been substantiated by empirical research, as neither randomized nor non-randomized studies have demonstrated gains in quality adjusted life years The aim of the present study is 1) to identify (cost-)effective follow-up strategies and 2) to assess whether the cost-effectiveness of such strategies can be confirmed in RCT's of realistic sample size

Methods: We created a Monte Carlo Markov simulation model, using DATA 3 0, representing disease stage and progression, follow-up, and reoperative therapy as separate but interdependent determinants of outcome Input to the model was derived from 3 patient series treated for colon cancer (N's of 266, 1108, and 4819 respectively), and from a metanalysis of a 214 articles Eight main follow up strategies were tested for 12 different settings (defined on the basis of disease, patient, and care characteristics Simulations were performed so as to predict of a potentially feasible international multicenter RCT (5000-25000 patients per arm using a 10 year time horizon) Final outcomes calculated were costs, 2-, 5- and 10 year survival, cancer death and death from other causes, and (quality adjusted) life years Intermediate explanatory outcomes included suspicion of recurrence, recurrence detection, reresection with curative intent, and sample size needed to obtain a significant difference per outcome

Results: In the base case scenario 3 strategies (CEA, ultrasound, and a limited combined scenario for 3 years) provided better life years (>=+0 08 year), better 2- and 5-year survival (>=+0 01), and lower cancer mortality (<=-0 01) than no follow-up, provided sample sizes equal or exceed 20,000-45,000 per arm Cost differences are significant for most strategies at sample sizes of 60-100 per arm Cost-effectiveness ratios are lowest ($22,000/QALY) found for CT scan once after year, but may rise to $85,000/QALY or over $200,000 for limited or aggressive combined strategies respectively Although colon cancer follow-up is cost-effective under certain conditions, RCT's will almost certainly fail to confirm 1) gains in significant and clinically relevant outcomes, and 2) cost-effectiveness of follow-up

THE VALUE OF INFORMATION IN THE ECONOMIC EVALUATION OF NEW HEALTH CARE TECHNOLOGIES AN APPLICATION TO A POLICY MODEL OF ALZHEIMER'S DISEASE

K Claxton[a], PJ Neumann, SS Araki, MC Weinstein Department of Economics, University of York UK and Harvard School of Public Health Boston MA [a] Commonwealth Fund of New York [b] Supported by Pfizer

Objectives To estimate the value of additional information in the choice between alternative treatments of Alzheimer's disease and to establish a value of information for each of the uncertain model inputs

Methods A Bayesian approach to the value of information was applied to an established model of Alzheimer's disease[b] Prior distributions were assigned to transition probabilities based on longitudinal data from a national registry and to preference weights and costs based on data from a cross-sectional study Drug efficacy was modelled as a relative risk ratio (RRR) and was assigned a distribution based on the reported 95% CIs from a 24 week clinical trial The uncertainty surrounding the RRR when it was extrapolated beyond 24 weeks was characterised as either exchangeable (same as the trial period) or as partially exchangeable (an increasingly sceptical prior) The distributions were propagated through the model using Monte Carlo simulation A loss function represented the objective of maximising expected net benefits conditional on a monetary valuation of health outcome (1/g) The perspective was societal, and the future US population of Alzheimer's disease sufferers was estimated from published incidence rates and US census projections

Results The expected value of perfect information (EVPI) is the maximum societal decision makers should be willing to pay for additional information (clinical research) At 2 years and assuming partial exchangeability the EVPI for the treatment decision was $127 46 million (1/g=$50,000), which was higher than when exchangeability was assumed ($92 71m) The EVPI is dependent on the value of 1/g selected ($14 88m when 1/g=$80,000) and on the time horizon ($52 17m at 3 years) The analysis also included estimates of the value of information for each of the uncertain model inputs At 3 years the EVPI for the RRR after 24 weeks was $49 44m which was substantially greater than the RRR within the trial period ($2 72m) The EVPI for all other variables combined was estimated to be $18,819 at 3 years

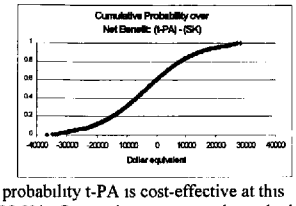
Conclusions The EVPI can be used to set research priorities In this case further research appears to be potentially cost-effective The value of additional information about costs and utilities is small compared to the value of effectiveness data and the value of longer term effectiveness data (extending a trial beyond 24 weeks) is substantially greater than additional information about effectiveness within the existing trial period

BAYESIAN COST-EFFECTIVENESS ANALYSIS ILLUSTRATION OF METHODS

DG Fryback, JO Chinnis Jr , J Uvila University of Wisconsin, Madison, WI, and Decision Science Associates, Inc , Manassas, VA

Cost-effectiveness analysis (CEA) as usually practiced computes a point estimate of the cost-effectiveness ratio (CER) and represents uncertainty about input parameters by sensitivity analyses Recent efforts have derived classical confidence intervals for CERs or net benefits (NB) to reflect uncertainty about some inputs A Bayesian analysis derives a full posterior probability distribution for the quantity of interest incorporating all uncertainties in the problem as well as prior information We illustrate a Bayesian CEA using published results of the Global Utilization of Streptokinase (SK) and Tissue Plasminogen Activator (t-PA) for Occluded Coronary Arteries (GUSTO) trial, a previous CEA (Marks, et al , 1995) concluded the incremental CER for t-PA vs SK is $32,678/QALY in their base case

We model 12 sources of uncertainty with full distributions Beta-binomial conjugate distributions model short-term mortality and stroke risk in the t-PA and SK arms, with and without prior information from earlier trials Short term and long term costs of the occlusive episode and of stroke are modeled using 4 gamma distributions Quality of life weights are modeled using both beta distributions and direct empiric distributions The entire analysis is implemented in Excel 97, the approximation method of bracket medians (Schlaifer, 1969) is used to convolute probability distributions Due to ambiguities in expressing uncertainty about CERs we used NB in dollar equivalents (Stinnett & Mullahy, 1998) calculated at $50,000/QALY as the summary quantity Without prior trial information, the probability t-PA is cost-effective at this threshold is 39 4%, with prior trial data included this is 38 9% Our analysis uses simple methods in reach of most analysts now More sophisticated Bayesian techniques are becoming available



PROSPECTIVE ECONOMIC EVALUATION OF A PHASE III TRIAL COMPARING IRINOTECAN WITH INFUSIONAL 5-FU IN PATIENTS WITH ADVANCED COLORECTAL CANCER

C Schmitt, B Jolain, E Van Cutsem ARCOS, Issy les Moulineaux, France, Rhône-Poulenc Rorer, Antony, France, UZ Gasthuisberg, Leuwen, Belgium

Objective To evaluate economic implications of treatment with irinotecan (CPT) based on the results from a phase III randomized trial which demonstrated significantly increased survival at one year with irinotecan (45%) compared to infusional 5-FU (32%) in patients with metastatic colorectal cancer

Methods Resource utilization data were collected prospectively alongside the phase III trial, with 256 patients followed for a median of 10 months Follow-up was prolonged beyond treatment failure and resource utilization was not protocol driven enabling a realistic evaluation of economic implications The problem with right-censoring of the economic data was handled by two recently proposed non-parametric methods one by Lin et al (Estimating medical costs from incomplete follow-up data Biometrics 53 419-34, 1997) and one by Zhao and Tsiatis (A consistent estimator for the distribution of quality adjusted survival time Biometrika 84(2) 339-48, 1997) The problem with right truncating of the data was explored using a Weibull survival model combined with an estimate of average resource utilization during the observational period The data from the trial were supplemented by a survey of the investigators to document resources used for chemotherapy administration according to their center's usual practice

Results Resource utilization due to chemotherapy administration was comparable for CPT and 5-FU, although wide differences were observed among centers The mean cumulative number of days in hospital due to complications which is the most important cost driver was 14 4 (95% CI 10 7 - 18 1) with CPT vs 17 5 (95% CI 11 7-23 3) with 5-FU, as estimated using the method by Lin et al The corresponding figures were 14 3 days for CPT vs 16 9 days for 5-FU, as estimated using the method by Zhao and Tsiatis Adjusting for right-truncation yielded an estimated 16 0 days for CPT vs 18 1 days for 5-FU

Conclusion Second-line irinotecan improves survival compared to infusional 5-FU in patients with advanced colorectal cancer without significantly increasing medical care consumption This result is not sensitive to the choice of the method to adjust for censored and missing economic data Cost-effectiveness was further analyzed combining these estimates with Belgian cost data (cf abstract submitted for this SMDM meeting by Neymark et al)

MANDATED COVERAGE FOR CLINICAL PREVENTIVE SERVICES: WHOSE GUIDELINES DO STATES FOLLOW?

SS Rathore, JD McGreevey III, R Snihurowych, S Lee, KA Schulman and D Atkins Clinical Economics Research Unit, Georgetown University Medical Center, Washington, DC and the Agency for the Health Care Policy and Research, Rockville, MD

States recognize the value of clinical preventive services (CPS) and mandate their provision for insured patients While recommendations for some CPS (i e Papanicolaou (Pap) smears) are based on extensive clinical evidence, others (i e prostate specific antigen (PSA) screening) are based on more speculative data We sought to evaluate state-mandated coverage for four oncologic CPS of differing preventive value

We contacted Insurance Departments in 50 states, DC and PR for copies of state code concerning provision of mammography, Pap smears fecal occult blood test/sigmoidoscopy (FOBT/S) and PSA by indemnity and HMO plans We evaluated code for mandated coverage of any of the four CPS and determined whether code specified general, aggressive (American Cancer Society, 'ACS') or conservative (United States Preventive Services Task Force, 'USPSTF') screening recommendations

CPS	GENERAL GUIDELINES	SPECIFIC GUIDELINES			TOTAL
		ACS	USPSTF	OTHER	
Mammography	2	36	2	3	43
Pap smears	-	9	3	11	23
FOBT/Sigmoidoscopy	-	-	-	2	2
PSA	-	4	-	-	4

45 states, DC and PR mandate some coverage of oncologic CPS Although codes rarely cited the source of guidelines, most states followed aggressive screening recommendations Legislators need to be aware of differing sources of CPS guidelines and should consider the clinical, economic and policy implications of mandating coverage based on aggressive CPS recommendations

IS POST-POLYPECTOMY SURVEILLANCE COST EFFECTIVE IN MALE PATIENTS?

RM Ness, RW Klein, AM Holmes, RS Dittus Indiana University, Indianapolis, IN and Vanderbilt University, Nashville, TN

Post-polypectomy surveillance for colorectal neoplasia is currently recommended for patients in whom large or multiple polyps are found. However, its cost effectiveness remains unproven We examined the cost-utility (C/U) of post-polypectomy surveillance strategies in male patients using decision modeling

We built a discrete-event simulation model of the natural history of colorectal neoplasia in male patients Estimates for the risk of colorectal cancer (CRC), the incidence of polyps, the mortality from CRC, the sensitivity and complication rates of colonoscopy, and the direct costs of CRC diagnosis and treatment were summarized from the literature Quality adjusted life years (QALYs) were computed as the measure of effectiveness, using utility estimates obtained from an instrument we designed to measure utilities for outcome states of CRC as the weights We created a hypothetical cohort consisting of male adults in whom either one large or more than one polyp had been discovered and removed during initial screening at age 50 We modeled alternative surveillance strategies to calculate their cost-utility Each surveillance strategy was designed to begin at initial screening and continue until death We discounted both costs and QALYs at a 3% annual rate No surveillance cost $1036 and resulted in a mean of 17 257 QALYs per person.

Strategy(yr)	Cost($)	QALYs	Marg. Cost($)	Marg. QALYs	Marg. C/U($/QALY)
q10	794	17 3373		-	-
3-10 conditional	991	17 3857	197	0 0484	4,070
3, then q5	1487	17 4041	496	0 0184	26,957

The mean per-person results of 3 alternatives are shown above The label "3-10 conditional" means survey q3 until one not large or no polyp is found, then switch to q10 Other recommended or commonly used strategies examined (q3, q5, 3-5 conditional) were dominated. When the sensitivity of colonoscopy for colon polyps was increased by 10%, q10 still dominated no surveillance, but marginal C/U ratios for more frequent surveillance exceeded $48,000/QALY In males screened for CRC at age 50 in whom either large or multiple polyps are found, post-polypectomy q10 surveillance is cost-effective Other strategies (3 then q5 or 3-10 conditional) become much less cost effective as the sensitivity of colonoscopy increases

THE COST-EFFECTIVENESS OF TELEPHONE VERSUS MAILOUT ADMINISTRATION OF HEALTH QUESTIONNAIRES RESULTS FROM THE VA HEALTH ASSESSMENT PROJECT USING THE SF-36V AND SF-12V

DL Jones, LE Kazis, AM Hendricks, L Cassar, N Wilson CHQOER, Bedford, MA

Purpose The Undersecretary of the Veterans Health Administration (VHA) has recommended that functional status be used to monitor patient outcomes and assess performance at VA hospitals Using the SF-36V and SF-12V developed in the Veterans Health Study, we sought to measure the costs and response rates for administering these questionnaires by telephone and mail to determine which mode was cost-effective Methods 9,000 ambulatory care patients were randomly selected from six VA facilities to receive either the SF-36V or SF-12V health questionnaire through mail or telephone interview The study employed a crossover design (a) after 12 unsuccessful attempts to contact patients by telephone, patients were administered the questionnaire by mail, and (b) after 2 unsuccessful mailouts, 3 attempts were made to interview patients by telephone Cost-effectiveness analysis provided the framework for describing the steps in SF-36V and SF-12V administration, assigning values (i e , costs and numbers of completed surveys) to possible outcomes, and comparing different options Cost-effectiveness models are presented in schematic form as decision trees Results Total fixed costs were roughly equivalent, but mail administration had lower total variable costs and higher response rates The average total cost per completed questionnaire by mail administration was $4 94 less for the SF-12V ($15 98 vs $20 92) and $9 34 less for the SF-36V ($15 87 vs $25 21) as compared to telephone administration The marginal cost of an additional completed SF-12V or SF-36V was also substantially lower for mail administration Results of a sensitivity analysis using a discounted telephone rate did not alter the conclusion that mail administration, both with and without crossover to telephone administration, remained the preferred mode for administering the SF-36V Conclusions Costs and response rates are both influenced by the mode of administration Mail administration of the SF-36V and SF-12V was clearly dominant over telephone administration as indicated by lower average total and variable costs and greater response rates Cost-effectiveness analysis can be used to determine the most cost-effective strategy for administering health questionnaires when health services researchers and managers want to maximize their response rates and minimize their costs

A DECISION ANALYSIS OF MANAGEMENT STRATEGIES FOR POLYPOID LESIONS OF THE GALLBLADDER

H Moriguchi, M Kobayashi, J Tazawa, C Sato Div of Health Sci , Fac of Med , Tokyo Med and Dent Univ , Res Dept , Inst for Health Economics and Policy Health Economics Res Group, Crecon Res and Consulting Inc Dept of Intern Med , Tsuchiura Kyodo General Hosp , Japan

Mass screening for abdominal diseases have recently been performed using ultrasonography, and polypoid lesions of the gallbladder are one of the most common findings However, management strategies for these lesions remain controversial and there is no agreement as to whether preventive operation or follow up observation offers the best strategy because of the lack of a strictly controlled trial

A decision analysis was conducted to compare the expected utility of preventive operation versus follow up observation for polypoid lesions of the gallbladder Probabilities were derived from clinical and prognostic data for 109 patients (58 females age median 54 range 25 89) who had been diagnosed as having polypoid lesions of the gallbladder by ultrasonography, from 4,343 patients who had presented to the outpatient clinic of Tsuchiura Kyodo General Hospital (Moriguchi H et al Gut 1996,39 860 862) and the published literature The utility for each health state was expressed on a scale of 0 (death) to 1 0 (perfect health) and obtained from 50 gastroenterologists in Japan using time trade off methods and the published literature

A baseline analysis showed that the expected utility for follow up observation was 0 9820 and that for preventive operation was 0 9787 Sensitivity analyses showed that there were four values having large potential effects on the expected utility the mortality, the rate of complications, the utility of complications after preventive operation and the malignant transformation rate of the polypoid lesions For follow up observation to be preferred, threshold values of these four factors were as follows the malignant transformation rate of the lesions had to be as low as 1 3% the mortality after preventive operation had to be as high as 0 41% the rate of complications after preventive operation had to be as high as 15%, and the utility of complications after preventive operation had to be as low as 0 946

The decision on management strategies for polypoid lesions of the gallbladder depends on the following four values the mortality the rate of complications, the utility of complications after preventive operation and the malignant transformation rate of the polypoid lesions

FUNCTIONAL MEASUREMENT A NEW PLATINUM STANDARD FOR MEASURING PATIENT PREFERENCES

Beth A. Jaworski, Gregory J. Zarow VASDHS, UCSD Dept of Psychology, San Diego, California bjaworski@ucsd edu, gzarow@ucsd edu

Our purpose is to introduce functional measurement (FM) as the platinum standard for measuring patient preferences and subjective value

Values play a critical role in decision making at the individual and policy levels There is wide disagreement about the validity of methods utilized to assess preferences, such as the standard gamble, time trade-off, willingness to pay, rating scales, and magnitude estimation Invalid values can lead to disordinal choice and suboptimal decisions Criteria are needed to test the validity of any method of value measurement In most cases, the criteria are not obvious Convergent validity and theoretical soundness of the methodology have have served as surrogate criteria, however the results of most studies of the convergent validity of common health preference measures have been disappointing Theoretically based models are infirmed because the axioms on which these methods are based appear to be flawed and lack empirical support Other methods do not produce clearly meaningful units of measurement

A platinum standard of preference measurement must meet three criteria 1) measurement of subjective stimulus values on an equal-interval scale, 2) measurement of subjective response values on an equal-interval scale, 3) generation of the mathematical integration rule which relates the subjective values of the stimulus and of the response FM simultaneously meets all three criteria Extensive empirical FM work in the behavioral sciences has demonstrated that information integration obeys simple rules of cognitive algebra people appear to add, average, subtract or multiply information to form an attitude Further, FM allows for seperation of scale weights from scale values, not possible with regression techniques

Methodology for disafirming hypotheses is prerequisite for any claim of validity FM employs the use of factorial designs and the data produced from these designs are analyzed with ANOVA With FM, a hypothesized integration rule of patient preference can be validated or invalidated with precision

FM provides an extremely powerful device for validating integration rules while at the same time validating scale values Unlike other methods, FM permits conclusions about the level of measurement (ordinal, ratio) of scaled health states FM provides meaningful and valid subjective values of health states necessary for making decisions about resource allocation at the individual and policy levels

INEQUALITY AVERSION IN PREFERENCES FOR MENTAL HEALTH STATES

Ann M Holmes, Indiana University-Purdue University Indianapolis, Indianapolis, IN

Traditionally, health state utilities (HSU) have been based on individual preferences It has been argued that such values should not be used to allocate health care resources across groups of people since they do not reflect people's preferences regarding the distribution of health in society In this analysis, I examine (1) whether individual HSU's are consistent with social HSU's and, if not, (2) whether social values can be expressed as a function of individual values

Data were collected from 54 graduate students in Health and Public Administration programs at Indiana University-Purdue University Indianapolis Students completed written questionnaires in which they were asked to evaluate two mental health states (depression and schizophrenia) A time trade-off (TTO) instrument was used to solicit respondents' individual preferences, while a person trade-off (PTO) instrument was used to solicit their social preferences

Mean TTO values for depression and schizophrenia were 0 67 (SE 0 030) and 0 41 (SE 030), respectively Mean PTO values for depression and schizophrenia were 0 84 (SE 0 026) and 0 65 (SE 0 038), respectively Paired t-tests indicate mean PTO and TTO values were significantly different ($p = 0\ 00$) for both health states considered, and three-quarters of respondents assigned higher values using the PTO instrument than the TTO instrument PTO values were then expressed as a "mean of order r" function of individual TTO values ("r" captures the degree of inequality aversion $r=1$ indicates inequality neutrality/utilitarian ethics, while smaller values indicate greater inequality aversion) Median inequality aversion parameters were 0 38 (95% CI [0 19, 0 44]) based on depression scores, and 0 39 (95% CI [0 23, 0 58]) based on schizophrenia scores (confidence intervals are based on bootstrapped samples)

Results suggest these respondents were moderately averse to inequality in the distribution of health As a consequence, the use of their individual-based HSU scores to allocate health care resources in society would assign too low a priority to the treatment of severe illnesses (e g , schizophrenia) and too high a priority to the treatment of milder disorders (e g , depression)

PATIENT REPORTED UTILITY FOR RADIOFREQUENCY ABLATION FOR TREATMENT OF ATRIOVENTRICULAR NODAL RE-ENTRANT TACHYCARDIA

KM McDonald, M Larson, D Boothroyd, MA Hlatky Stanford University, Stanford, CA.

Patients with highly symptomatic atrioventricular nodal re-entrant tachycardia (AVNRT) are commonly treated with radiofrequency ablation (RFA) Small studies report improvements in quality of life for patients with severe symptoms We evaluated patient reported utility of RFA for treatment of AVNRT in a patient group with a range of symptoms and clinical utilization patterns

We performed a retrospective telephone survey of 161 AVNRT patients recruited from a consecutive series of patients who had undergone RFA within Northern California Kaiser Patients were asked about symptom frequency and severity, emergency room utilization and time-tradeoff elicited utilities for their current post-RFA health and for the period of one year preceding RFA

The median age of patients was 55 years, and 71% were female We found that symptoms decreased from a mean of 5 8 to 3 1 ($p<0\ 001$) per patient and moderate or severe symptoms decreased from 4 6 to 1 1 ($p<0\ 0001$) Emergency and urgent care visits decreased from 4 6 to 0 4 per patient per year after RFA Utility scores improved from a mean of 0 71 to 0 88 after RFA ($p<0\ 0001$) The change in utility from before to after RFA increased with increasing frequency of severe or moderate symptoms and with increasing number of urgent care visits

	TTO Utility Measured After RFA			
Urgent Visits Pre-RFA	Pre-RFA	Post-RFA	Improvement	Number of Patients
0	0 81	0 90	0 09	26
1	0 73	0 88	0 14	26
2 or 3	0 65	0 83	0 18	40
4 or more	0 67	0 90	0 23	59

Although the impressive changes in utility, symptoms and clinical utilization may include some recall bias, the measurements reported in this study accurately reflect the patients' perspective about their health state improvements at one point in time — postablation procedure

EXPERTISE VS CONTEXT IN THE EVALUATION OF ACUTE CHEST PAIN PATIENTS IN THE EMERGENCY DEPARTMENT

A Fontaine, D Soyez, P Estagnasie, H Puy, JC Deybach, Hopital Louis Mourier Faculte Xavier Bichat

Following the introduction of new sensitive but expensive tests for the evaluation of acute chest pain patients we wanted to assess how physicians determined the pre test likelihood of acute coronary disease

Residents and senior staff prescribing Troponine Ic assays in the emergency department recorded clinical and electrocardiograph findings on a standardized prescription form Four diagnostic hypotheses ranging from acute myocardial ischemia to non coronary disease were ranked independently without knowledge of the test results by three different physicians 1) initially by the physician prescribing the test 2) later by one other resident and/or senior staff of equivalent expertise on the basis of the information recorded on the prescription form and 3) by one certified cardiologist using the same information

A total of 126 consecutive patients were included over a 6 month period 15 (12%) had elevated Troponin Ic levels indicating myocardial ischemia (MI) All three sets of diagnostic hypotheses were available for 91 patients, including 12 patients (13%) with MI Physicians prescribing the test ranked either acute or sub acute myocardial infarction among the first 2 likely diagnoses for 23 patients (25%), 4 of whom had MI and thought that some degree of coronary disease was likely for 61 additional patients (67%) of whom 7 had MI Residents or senior staff reviewing the prescription forms ruled out coronary disease for 4 of the 23 patients initially classified at high risk and for 19 of the 61 patients considered at intermediate risk The cardiologist ruled out coronary disease for 13 additional patients None of the patients considered non coronary by secondary reviewers or by the cardiologist had elevated Troponine Ic Differences in ranking of diagnostic hypotheses by initial prescribers secondary reviewers and the cardiologist were not significantly associated with the seniority of the prescribing physician nor with differences in the interpretation of EKG findings or with the timing of the tests (i e at night or during weekends)

Systematic consultation with an expert cardiologist or additional training of residents might help avoid unnecessary tests However, the finding that both residents and senior emergency physicians reached more conservative (and accurate) conclusions when they reviewed "paper" cases than when they had to deal with real acute chest pain patients suggests that in an emergency context physicians regardless of their expertise may give more weight to the risks associated with overlooking acute myocardial infarction

LOGISTIC REGRESSION IN DIAGNOSTIC DATA ANALYSIS
COMBINING ADEQUATE MODELING WITH IMPROVED LAYOUT

EW Steyerberg, MJC Eijkemans, J Hilden Department of Public Health, Erasmus University Rotterdam, The Netherlands and Department of Biostatistics, University of Copenhagen, Denmark

Background: The path from data-collection to a practical diagnostic work-up scheme contains several logically independent steps statistical modeling and estimation, choosing a rule of interpretation, choosing a rule of execution (a 'flow chart'), and presenting the combined rule of interpretation and excecution in a nice layout The posterior logodds of disease can often be expressed as a linear combination of test results, e g with logistic regression (LR) The Bayes-with-independence model is a special case (no interaction), once test results are coded as 0/1 in the LR model However, sequential diagnostic work-up is not satisfactorily presented, since the LR coefficients presuppose that all test have already been done

Purpose: We aimed to improve the representation of LR models for sequential diagnostic problems, extending a proposal by Knottnerus [1]

Results: We found that is was possible to edit the standard LR formula such that it gradually unfolds as test data become available, while giving correct logodds at each step from no data to having all test results available Moreover, advantage can be taken of regression properties whenever it is assumed that a test does not interact with those already included (similar effect in all 2x2 tables), which leads to a more parsimonious model with a main effect of the test only It was readily possible to apply this approach to the examples provided before, where 2 or 3 dependent dichotomous tests were considered with or without interaction [1]

Conclusion: While logically separate, the several steps of diagnostic data analysis can be combined by applying logistic regression formalisms and software This refinement of the previously proposed method makes it more widely applicable, e g for small data sets and for larger numbers of tests

1 Knottnerus JA Application of logistic regression to the analysis of diagnostic data exact modeling of a probability tree of multiple binary variables Med Decis Making 1992 12 93-108

SELECTION OF PATIENTS WITH ATRIAL FIBRILLATION FOR A TRIAL OF THE IMPLANTABLE ATRIAL DEFIBRILLATOR VS AMIODARONE

BF Gage, K Woods, C Weimer, D Reuter, DK Owens Washington University, St Louis, MO, InControl Corporation, Seattle, WA, VA Health Care System and Stanford University, Palo Alto, CA

Although antiarrhythmic agents can help maintain normal sinus rhythm in patients who have atrial fibrillation (AF), their use is limited by their side effects One potential alternative to these agents is the implantation of an atrial defibrillator The purpose of this study was to define the patient population who might benefit from the implantation of an atrial defibrillator

We developed a decision-analytic model to compare quality-adjusted life expectancy in patients with AF who received either oral amiodarone or implantation of an atrial defibrillator Based on current practice patterns, we assumed that 75% of patients who had the atrial defibrillator implanted also would be prescribed amiodarone Patients receiving either treatment continued their warfarin therapy until they had maintained normal sinus rhythm for 1-month The utility of AF versus normal sinus rhythm was estimated by administering the standard-gamble technique using a paper-based questionnaire Utilities for other states were obtained from previously published interviews of patients who had atrial fibrillation In the base case, all patients were 65 years of age and had recently been cardioverted back into normal sinus rhythm

The two key variables in the analysis were the utility of AF and the utility of receiving the (4-Joule) cardioversion from the atrial defibrillator The threshold utility of AF was 0 96 3 of 10 patients rated their utility of AF lower than this figure indicating that their quality-adjusted life expectancy could increase with implantation of the atrial defibrillator Patients would not benefit from the defibrillator if their utility of being cardioverted intermittently was lower than their utility of AF or if implantation of the defibrillator increased their mortality significantly

Our findings suggest that implantation of an atrial defibrillator offers potential to improve quality-adjusted life expectancy in those patients who are most bothered by their AF

COMPARING ANALYTIC APPROACHES FOR THE ECONOMIC EVALUATION OF A CLINICAL TRIAL. THE BALLOON VS OPTIMAL ATHERECTOMY TRIAL (BOAT)

DA Taira, TB Seto, DJ Cohen Beth Israel Deaconess Medical Center, Boston, MA

The BOAT multicenter clinical trial compared directional coronary atherectomy to conventional balloon angioplasty The clinical findings and economic costs of BOAT have been reported previously The purpose of this study was to compare the cost estimates obtained using 4 analytic approaches 1) Total hospital charges, 2) Hospital costs generated using cost center-specific cost-to-charge ratios, 3) Catheterization laboratory costs based on measured resource utilization (e g, number of balloons, procedure duration) and unit costs, with ancillary hospital costs generated using cost center-specific cost-to-charge ratios, 4) Approach 3 with linear regression models to impute the costs for patients with missing bills For all approaches, physician fees were based on the Medicare RBRVS fee schedule

Results Mean charges exceeded mean costs by >$10,000, resulting in the largest cost difference between the two arms When catheterization laboratory costs were itemized, the difference between the 2 arms was twice as high as when procedure costs were not itemized

Approach	n	angioplasty	n	atherectomy	difference	p
1 Hospital charges	329	$20,897	335	$24,664	$3,630	030
2 Hospital costs	329	$ 6,664	335	$ 7,894	$1,183	004
3 Itemized procedure costs	329	$10,058	335	$12,310	$2,251	< 001
4 Imputed hospital costs	466	$10,114	471	$12,001	$1,887	< 001

Conclusion Although the cost of atherectomy is significantly higher than angioplasty under each approach, the magnitude of the cost difference varies considerably, with potential implications for cost-effectiveness analyses

COST-EFFECTIVENESS OF POPULATION INTERVENTIONS TO REDUCE THE INTAKE OF SALT (SODIUM CHLORIDE)

R Selmer, IS Kristiansen, DS Thelle, A Haglerød, K Bønaa National Health Screening, Oslo, Odense University, Odense, University of Oslo, Oslo, University of Tromsø, Tromsø

Aim The aim of the study was to estimate economic and health consequences of reducing the daily intake of salt (sodium chloride) by 6 grams in the Norwegian population Health promotion (information campaigns), food regulation and tax on salty food were used as interventions

Methods Based on a literature review we assumed that systolic blood pressure (BP) would be reduced by 2 mm Hg (range 1-4) through the interventions Health and cost consequences were estimated in a Markov model The risk of myocardial infarction (MI) and stroke was based on the observed mortality The impact of the blood pressure reductions on the risks was based on 35 years of follow-up of the Bergen Hypertension Study We included the following cost components Cost of interventions *per se*, welfare losses from price interventions (based on elasticity of demand and supply), cost of avoided MIs and strokes, cost of avoided antihypertensive treatment, health care costs in additional life (based on age and sex specific hospital utilization rates) and productivity gains from reduced morbidity/mortality,

Results The estimated increase in life expectancy was 1 8 months in males and 1 4 in females The net discounted (5%) cost of the interventions was minus $120 millions (*i e* cost saving) in the base case Sensitivity analyses indicate that the interventions would be cost saving unless the systolic BP reduction were only 1 mmHg, future costs were doubled, productivity gains were disregarded or the welfare losses from price interventions were greater than $70 millions per year If these losses were as great as $142 millions per year, the cost per life year gained would be $29,000

Conclusion Population interventions to reduce the intake of salt are cost saving under a range of assumptions

MANAGEMENT OF SINUS NODE DYSFUNCTION IN THE ELDERLY- A DECISION ANALYSIS

Glenn A. Pacio and Thomas McGinn, Montefiore Medical Center, Bronx, NY

Background: The natural history of untreated sinus node dysfunction has not been completely elucidated and current definitive indications for permanent cardiac pacemakers require that the arrhythmia be clearly associated with symptoms or that it is a result of a necessary medication Theophylline has been tried as treatment to prevent symptoms primarily arising from bradycardia We compared two management strategies for a 70-year old man who has had complaints of near-syncope with note of asymptomatic bradycardic and tachycardic episodes on telemetry and with a non-diagnostic electrophysiologic study

Methods: The Markov decision analysis model was used to compare immediate pacemaker insertion versus trial of medical therapy with theophylline Health states consisted of absence or presence of recurrent syncope or near-syncope, complications of treatment including pacemaker syndrome, atrial fibrillation, heart failure, gastrointestinal intolerance and death over a time period of five years Outcome measured was quality-adjusted life years based on a recently published quality of life and clinical outcomes in elderly patients treated with pacemakers (New England Journal of Medicine 338 1097-1104, April, 1998) Sensitivity analysis were performed over a range of probabilities for pacemaker complications, mortality from sinus node dysfunction, recurrence of symptoms after pacemaker therapy, side effects and effectiveness of medical therapy as gathered from available literature

Results: Medical therapy had a higher quality adjusted life years compared with the immediate pacemaker insertion strategy (2 79 versus 2 59) Medical therapy remained the preferred strategy across all probability values in the sensitivity analysis

Conclusions: Elderly patients with no definitive indications for cardiac pacemaker insertion for the management of sinus node dysfunction may benefit from trial of medical therapy with theophylline to alleviate symptoms

PHYSICIANS' JUDGEMENTS OF OUTCOMES OF TREATMENT FOR HEART FAILURE

Roy M. Poses, Brown University School of Medicine, Providence, RI, Maria Woloshynowych, D. Mark Chaput de Saintonge, St Bartholomew's and the Royal London School of Medicine and Dentistry, London, UK

Physicians frequently fail to use angiotensin converting enzyme inhibitors (ACEI's) for patients with congestive heart failure (CHF) and systolic dysfunction despite strong evidence that they prolong survival, decrease morbidity, and cause few adverse drug reactions (ADR's) Physicians may base decisions to prescribe ACEI's on their judgments of mortality risk reduction due to ACEI's and of the probability of adverse drug reactions (ADR's) Further, physicians may underestimate mortality risk reduction by ignoring factors that positively predict increased risk reduction for specific patients or may overestimate the probability of ADR's by taking account of factors that might plausibly appear to be predictive of ADR's, but are not

Our physician subjects were 20 general practitioners, 21 hospital based internists other than cardiologists, and 19 hospital-based cardiologists practicing in London, England, selected essentially randomly We asked the physicians to consider 27 case vignettes constructed such that selected clinical characteristics systematically varied across them These included characteristics that predict either risk reduction or ADR's for CHF patients (e g , NYHA class IV CHF predicts greater risk reduction than III and hypotension predicts a higher likelihood of ADR's), and characteristics which do not predict such outcomes (e g a history of a kidney stone or gastroesophageal reflux, or low normal blood pressure) For the majority of physicians, we found relationships between judgments of risk reduction and of ADR probability due to ACEI's and ACEI treatment decisions How physicians weighted clinical variables when making judgments of outcome rates varied widely 12/60 (20%) physicians used at least one positively predictive or non-predictive variable as a significant negative predictor of risk reduction, and 12/60 (20%) also used at least one non-predictive variable as a significant positive predictor of ADR risk

When judging outcome rates physicians may fail to take into account predictive variables and/or take into account non-predictive variables Appreciating how doctors make judgments and decisions may lead to better interventions to decrease unnecessary practice variation and improve quality

A STATISTICAL METHOD FOR MONITORING HEALTH CARE OUTCOMES

MA Rosenberg, University of Wisconsin Business and Medical Schools

Purpose To develop a statistical method that inexpensively monitors health care outcomes

Background Monitoring health care outcomes helps ensure the delivery of quality health care Expenditures on unnecessary services can increase the cost of providing care and potentially reduce quality The Non-Acceptable in-patient Claims (NAC) rate is an outcome measure that identifies unnecessary services, which if lowered, would allow for costs to be allocated more effectively One approach to monitor the NAC rate involves an expensive audit of patient medical records

Method Expensive audited claim records (model data) consist of claim data submitted for reimbursement, plus an indicator of NAC These data were used to develop a hierarchical Bayesian logistic regression model (NAC model) to screen for NACs (Rosenberg 1998) The model data establish the in-control state at time 0 A Bayesian approach determines whether an in-control state has changed since time 0 using future claims (monitor data) The state has not changed if the NAC rate posterior distribution (NPD) for the monitor data is the same as the model data One decision-rule (DR1), a variation of classical statistical quality control theory, finds the control limits of the model data NPD using +/- 2sd from the posterior mean The area of the monitor data NPD outside those limits, is an indicator of an out-of-control process A second decision-rule (DR2) finds the control limits of the model data NPD using specified percentiles and the amount of the monitor data NPD outside those limits

Results Abdominal pain claims had a 34 3% NAC rate in the model data and the NAC model predicted 33 1% The expensive audit for the next year's data show a NAC rate of 12 9%, indicating a change in the process Using DR1, 4 4% of the NPD for the model data is beyond the 2sd limits, while 20 7% for the monitor data is beyond the 2sd limits Using DR2 with 2 5% in each tail, 23 8% of the monitor data NPD is beyond the control limits Both rules conclude that the process may have changed and an investigation warranted

Conclusions The NAC model uses claim data as a proxy for expensive information gathered from medical record audits Inexpensive information is used to continuously monitor a process (by claim type or by hospital) for changes While this study shows that the NAC model and monitor can control the admissions process effectively, the methodology is transferable to other outcome measurements

THE TREATMENT OF INSOMNIA IN THE ELDERLY A COST-UTILITY ANALYSIS

L Bell, P Tousignant Clinical Epidemiology, Royal Victoria Hospital,/McGill University, USAGE Research Group

Insomnia afflicts an estimated 50% of people over age 65 in Canada and the USA For many, the condition is chronic - lasting months to years To compare therapeutic options available to physicians for the treatment of chronic insomnia in the elderly, cost and patient utilities for the treatment outcomes were analysed using decision-tree analysis Treatment consequences, probabilities, costs, and utility estimates were obtained from the literature and the analysis was enhanced by expert opinion and a seniors focus group interview

The main treatment options identified were to do nothing, to suggest non-pharmacological therapies (NPT), to prescribe benzodizepines or to prescribe another non-benzodiazepine medication i e zopiclone Cost-utility analysis indicates that the « do nothing » treatment option is neither the most effective nor the least expensive treatment option due to hidden medical costs and non-alleviation of symptoms in the majority of cases Because of the risks, costs and negative patient preferences associated with adverse events such as withdrawal syndrome, falls/accidents and or cognitive impairment which can result from benzodizepine and zopiclone therapies, treatment with these commonly used medications also does not optimize treatment risks and benefits NPT, on the other hand, maximize quality of life while minimizing costs Unfortunately, general practitioners continue to rely heavily on benzodiazepine therapies Over-prescribing may be due to lack of consultation time, patient resistance to changing regimes, inadequate knowledge of and belief in NPT and poor access to non-pharmacological counselling programs

When compared with benzodiazepines for the average patient, NPT produce only 0 37 QALYs of gain and saves only $2781 over 10 years In view of the high prevalence of insomnia, however, the savings to the health system of choosing this type of therapy option could be significant - over $200 million annually in Canada alone Thus, if no underlying health problem can be identified, non-pharmacological therapies should be considered the therapy option of first recourse Qualitative data suggests that the key may rest with preventing the initiation of benzodiazepine use for insomnia in the « young » elderly - to avoid dependence and chornic drug use, as well as the other negative outcomes associated with benzodiazepine use

A COST-ANALYSIS OF ALTERNATIVE TREATMENT STRATEGIES FOR ACHALASIA

TF Imperiale, JB O Connor, MF Vaezi, JE Richter Dep t of Medicine Indiana Univ School of Medicine Indianapolis IN and Department of Gastroenterology Cleveland Clinic Foundation Cleveland OH

Background The optimal treatment (Rx) for otherwise healthy patients with achalasia is controversial

Objective To compare the costs per cure of alternative strategies for Rx of achalasia

Methods Using decision analysis we compared the cost per cure for achalasia among otherwise healthy adults who would be candidates for any of the following Rx strategies 1) laparoscopic Heller myotomy with fundoplication (LHM) 2) pneumatic dilation (PD) with LHM reserved for Rx failures and PD perforations 3) botulinum toxin (BT) injection with PD reserved for BT failures LHM failures and PD perforations were managed with open Heller myotomy with fundoplication with an assumption of long-term cure Probabilities were derived from literature-based weighted averages and included 5% for PD perforation 80% and 70% immediate success rates for PD and BT respectively 2 5% for PPI-requiring reflux following BT and PD and 5% for post-LHM reflux Only direct costs were considered, and were derived from hospital accounting departments and Medicare-allowable payments During the 5-year time horizon BT responders were assumed to require a mean of 7 Rx sessions, and PD responders who developed a recurrence were assumed to require a mean of 2 additional PD sessions

Results Under base-case conditions the respective costs per cure of PD BT and LHM were $3111 $3723 and $10 792 respectively Despite rates of 96% and 94% for short- and long-term efficacy, respectively LHM remained the most costly strategy In one-way sensitivity analysis LHM was less costly than BT when LHM cost < $3244 a highly unlikely scenario PD remained less costly than BT as long as rates of PD efficacy and perforation were ≥ 70% and < 9 5% respectively and cost of BT (including endoscopy) was ≥ $450 Results were not sensitive to the probabilities of short- and long-term BT response recurrence following PD LHM efficacy post-LHM reflux or to costs of LHM or PD

Conclusion For otherwise healthy patients with achalasia PD is the least costly initial strategy provided that the PD perforation rate remains less than 10% BT becomes less costly only when its costs decrease by 25% LHM is the most costly strategy under all clinically plausible scenarios Subsequent analysis should include a longer time horizon and an assessment of patient preference for each Rx strategy

DO PATIENTS PREFER ERYTHROPOIETIN OR TRANSFUSION FOR THE TREATMENT OF CANCER CHEMOTHERAPY-RELATED ANEMIA?

MES Sorbero, JI Low, JJ Griggs University of Rochester Rochester New York

OBJECTIVE The purpose of this study was to elicit patient utilities associated with treatment with transfusions (tnfs) and erythropoietin (epo) for cancer chemotherapy-induced anemia Preferences for the two treatment options have not been described

HYPOTHESIS We hypothesized that there is no difference in utility between epo (Uepo) and tnfs (Utnfs) in the treatment of chemotherapy-induced anemia (null hypothesis)

METHODS A convenience sample of 38 chemotherapy patients treated at the University of Rochester Cancer Center were invited to participate We conducted annual time-tradeoff interviews using a multi-attribute scenario for both of the states Patients were invited to comment at the conclusion of the interview The difference in utility associated with the two treatment strategies (ΔU) was calculated for each participant

RESULTS Two of the patients were excluded from the analyses because of inconsistencies between their utilities and concluding comments Among the 36 included in the analysis 3 (8 3%) were not willing to trade off any time to achieve perfect health for either scenario 12 (33 3%) had no preference between the two treatment options 18 (50%) preferred thrice-weekly epo injections and 3 (8 3%) preferred tnfs The ΔU s and the equivalent numbers of days are described in the Table below

	Uepo	Utnfs	ΔU	# Days	p value	
All patients (n=36)	854	781	073	26	01	
Previously transfused (n=11)	777	754	023	8	05	001
No h/o transfusion (n=25)	888	793	095	34	001	

CONCLUSIONS Although our patients prefer epo to tnfs for the treatment of chemotherapy-induced anemia this preference is significantly less pronounced in those experienced with tnfs These data on utility should be incorporated into cost-utility analyses of erythropoietin

MODELING BREAST CANCER RECURRENCE TIMES TO AID TREATMENT DECISIONS

Jarrett Rosenberg Clement and Associates, Redwood City, CA

Purpose. Recurrences of breast cancer have different causes leading to different prognoses However, even though it is known that different mechanisms with different prognoses are involved, physicians have little but actuarial statistics on which to base treatment decisions This paper describes the use of statistical and simulation techniques to derive a growth-rate model of breast cancer recurrences that can aid physicians in distinguishing between recurrences due to regrowth of primary lesions and recurrences due to new primaries

Method. Summary data from (1) ipsilateral breast recurrences after lumpectomy alone and lumpectomy plus radiation from the NSABP B-6 study, and (2) recurrences after mastectomy alone from the Milan Cancer Center were analyzed for peak recurrence times The results were then used with published data about breast cancer growth rates to create simulations of recurrence patterns

Results. Analysis of the two data sets revealed them both to be multimodal These multiple modes in recurrence times represent the times to detection of regrowths and new primaries Regrowths appear to have a slightly faster mean doubling time than new primaries, with extensive overlap between the two distributions Earlier and later recurrences are associated with different histological markers, as well as different survival outcomes

Conclusion. A simple model of multiple recurrence mechanisms is consistent with reported recurrence times Since the model predicts that different mechanisms can lead to recurrences with the same time delay after treatment but different prognoses, physicians can use the model in conjunction with estimates of the tumor's growth rate and other information to aid diagnosis of which mechanism is responsible for the recurrence, and thus which form of treatment is most appropriate

SURGICAL DECISIONS IN PATIENTS WITH SOLITARY LUNG NODULES AND CHRONIC OBSTRUCTIVE LUNG DISEASE

S Cykert and J. Jones from the Internal Medicine Program of the Moses H Cone Health System, Greensboro, NC and the Division of General Internal Medicine and Clinical Epidemiology of the University of North Carolina School of Medicine, Chapel Hill, NC

Patients who are considered for lung resection surgery because of possible cancer are most often chronic smokers who suffer pulmonary function impairment Past reports describe predictors of adverse post-operative outcomes in lung surgery patients, but it is unclear that these traditional outcomes studied (transient atelectasis, mechanical ventilation, or pneumonia) represent quality of life reductions large enough to alter surgical decisions Therefore, we sought to assess patient preferences concerning outcomes reported in the lung resection literature as well as other outcomes of potential importance and use the results in a model designed to clarify the surgical decision- making process A structured questionnaire employing the standard gamble technique was administered to 64 patients aged 50 to 75 years at 3 practice sites Results were converted to a utility scale with perfect health valued at 1 0 and death at 0. The utility scores (0 8 for the traditional outcomes, 0 5 for myocardial infarction, 0 3 for oxygen dependence, and 0 2 for states of severe physical debility) were then inserted into a decision model The model assumes that the patient is 60 years old with a 2 cm rounded solitary pulmonary nodule on chest x-ray and has a 1 5 pack per day smoking habit Based on life expectancy, surgical risk, certainty of diagnosis, and probability of cancer and surgical cure, expected utilities expressed in quality adjusted life years (QALYS) were calculated for surgical intervention compared to no surgery In patients for whom the diagnosis of cancer was known and surgical risk was low, the surgical group could expect 9 QALYs compared to 0.3 without surgery For the group with uncertain diagnosis and low surgical risk these data were 9 QALYs with surgery and 5 QALY's without; for cancer diagnosis known and high surgical risk, 1 9 QALY's with surgery and 0 3 QALY's without, and for those without pre-operative diagnosis and high surgical risk, 1 9 QALY's with surgery and 5 QALY's without Probability of surgical cure and peri-operative death were the factors in the model most sensitive to variation Outcomes traditionally used in the lung surgery literature had no effect on the decision model but uncertainty of pre-operative diagnosis did. In conclusion, this decision analysis shows that patients with reasonable probability of surgical cure should undergo lung resection unless surgery presents a high risk of death or debility Predictors of death and physical limitation need to be identified since atelectasis, pneumonia, and moderate duration of mechanical ventilation do not affect surgical decisions

MEDICAL DIAGNOSIS OF CEPHALALGIA USING INDUCTIVE MACHINE LEARNING

GD Dounias, S Drivalou, VS Moustakis, EP Nikolakaki Technical University of Crete, Greece, Institute of Computer Science, FORTH, Crete, Greece, and Psychiatric Diseases Therapeutic Center of Chania, Greece

The purpose of our study was to test the efficacy of inductive machine learning algorithms, when dealing with a medical diagnostic domain such as cephalalgia Special characteristics which exist in the case of cephalalgia, like the fact that it does not really constitute an illness or abnormality itself, but appears as a common symptom of various other health problems, makes the domain complicated as well as attractive in applying modern approaches of computer based diagnosis

The method we used, belongs in the area of inductive machine learning "Learning from examples" was performed using C4 5 algorithm, which utilizes entropy information theory in order to form a decision tree, translated to "if/then" rules in the basis of discriminant descriptions among the diagnostic characteristics Examples were taken from reviewed medical records, were preclassified and represented past medical diagnostic experience The application domain consisted of 300 patients (training set), suffering from 18 different types of cephalalgia (cluster headache, migraine with aura, episodic tension-type headache, etc), described by 84 different nominal or numerical diagnostic characteristics (questions on symptoms like location and frequency of pain, photophobia, nausea, biochemical results, etc)

Our results presented classification accuracy of the training set (100%) and of unseen cases (over 80%, exceeding 90% for the most representative classes in the training set), while special attention was focused on determining the complexity and the comprehensibility of the acquired decision rules, in collaboration with medical experts Complexity of the 50 diagnostic rules produced by C4 5, was expressed by the number of premise parts (conjunctions) constituting each rule, while comprehensibility of rules was expressed in qualitative characters of completeness and correctness There were no rules consisting of more than 4 premise parts (low rule complexity) while 7 of them were characterized as completely meaningful covering almost half of the training examples According to the experts opinion, 17 rules were not important and 2 rules had the wrong meaning

Inductive machine learning proved to be an efficient and promising tool in the approach of computer based diagnosis of cephalalgia New medical records are now added to our training set, in our attempt to test the system's performance in rare types of cephalalgia In addition, other, competitive inductive learning algorithms (CN2 and AQ) are applied on our data and are compared with the results of C4 5

PROPHYLACTIC MASTECTOMY OR BREAST CANCER SCREENING IN BRCA1/2 MUTATION CARRIERS A DECISION ANALYSIS USING THE TRADEOFFS OF (SUSPECTED) CARRIERS

LCG Verhoef, IJ Unic, PFM Stalmeier, WAJ van Daal Institute for Radiotherapy, AZN, Nijmegen, The Netherlands

Female carriers of a BRCA1/2 mutation have a high risk to develop breast cancer (50-85%) They may consider bilateral prophylactic mastectomy as a preventive measure The purpose of our study was to provide guidance for counseling individual women by estimating the benefit of prophylactic mastectomy in terms of (quality adjusted) life expectancy

Employing decision analysis, prophylactic mastecomy and screening were compared with respect to their effect on (quality adjusted) life expectancy Estimates of (quality adjusted) life expectancy were produced for three risk levels, for different ages and for different utility values Utilities were assessed in healthy women with a family history of breast cancer who were referred for genetic counseling and breast cancer screening, using the time-tradeoff test at two occasions and at three occasions if large differences between two replications existed

Forty-eight women completed the time tradeoff test The test-retest reliability for the two last replications of the time tradeoff test was 0 96 The mean utility was 0 70 (SD 0 30) The gain in life-expectancy after prophylactic mastectomy was higher for younger women at higher risk levels, and may amount to 6 9 years Using the utility value of 0 70, no gain in quality adjusted life expectancy is found at any age nor at any risk level Only women whose utility value is higher than average may benefit from prophylactic mastectomy The benefit is greater if the utility is higher, in younger women and at a higher risk level

Most carriers will not benefit from prophylactic mastectomy in terms of quality adjusted life expectancy Decision analysis may help to select those who will benefit by providing a suitable starting point to discuss the options with the patient

HOW DOCTORS DIAGNOSE URINARY TRACT INFECTIONS THE POTENTIAL INFLUENCE OF LABORATORY REGULATIONS ON TEST PREVALENCE AND USE

SD Flach, RS Wigton, TG Tape, C Longenecker, C Parenti, TJ Bryan Univ of Nebraska Med Ctr, Omaha NE, Johns Hopkins Univ, Baltimore MD, Univ of Minnesota, Minneapolis MN, Univ of Alabama, Birmingham AL

Critics of laboratory regulations claim that the high cost of compliance impacts negatively on patient care We examined the effects of physician office lab (POL) regulations on testing for urinary tract infection (UTI) by re-analyzing a 1994-95 survey of primary care physicians in four states (AL, MN, NE, PA) concerning their diagnostic approach in cases of possible UTI One of the states, PA, has had statewide POL regulations since the 1970s, POLs in the other states were not regulated until the implementation of the Clinical and Laboratory Improvement Amendment of 1988 (CLIA-88) in late 1994 Thus, the survey includes one regulated and three unregulated states, allowing for an examination of the possible effects of POL regulations We analyzed 2139 survey responses using multivariate logistic regression to test whether POL testing capabilities (availability) differs by state, and if use of the test is related to its availability, controlling for organizational and physician characteristics, including the reported value of the test in diagnosing UTI We found that many tests were available more commonly in unregulated states than in PA, this includes the dipstick, microscopic UA, wet prep, and urine culture (OR from 2 6 to 5 2) Subsequently, doctors were more likely to order micro UAs and cultures when those tests were available (OR of 3 4 and 1 7, respectively), being able to do a micro UA in the office was associated with greater use of the leukocyte esterase test (OR = 1 7) and a decrease in urine cultures (OR = 0 5) Increased belief in the value of the test results was significantly related to both the presence and use of the micro UA and dipstick (OR from 1 8 to 10 5) We found decreases in test availability in the state with POL regulations and a subsequent decrease in testing when availability is reduced, suggesting that POL regulations may influence clinical behavior Further study will be required to determine the level and pattern of testing that maximizes patient welfare

COST-EFFECTIVENESS ANALYSIS OF ADJUVANT TREATMENTS IN THE MANAGEMENT OF BREAST CANCER

Selke B [1], Lenne X [1], Bercez C [2], Bonneterre M E [2], Lebrun T [1], Bonneterre J [2] [1] Centre de Recherches Economiques Sociologiques et de Gestion, Lille, France , [2] Centre Oscar Lambret, Lille, France

Objective The aim of the study was to assess the incremental cost-effectiveness (ICE) ratios of four therapeutic strategies in the management of breast cancer chemotherapy (CT), hormone therapy (HT), both treatments (CT+HT) and no adjuvant therapy (NO)

Method An ICE analysis was performed using a Markov process model with a Monte-Carlo simulation Relevant clinical data introduced into the model were obtained from published data or from experts The ICE analysis assessed the cost of one additional life year saved The inpatient costs of the strategies were estimated according to the type of adjuvant treatment and to the patient monitoring carried out during the period without recurrence A prospective survey was performed on a cohort of local patients to quantify the costs external to the hospital (including the cost of transportation) The costs of recurrence were evaluated from the medical records of 146 patients with a minimal follow-up of 10 years Costs were expressed in 1995 French Francs (FF), with an accounting rate of 5 5% per year A sensitivity analysis using a probabilistic sensitivity analysis with a Monte-Carlo simulation (Doubilet et al , Med Decis Making 5 157-177, 1985) was also performed

Results According to the trials, NO is strictly dominated by average by HT and CT+HT The frequency according which NO is strictly dominated by HT varies between 81% and 100% For CT+HT, this frequency equals to 76% For CT, it varies between 0 07% and 60% and the cost per life year saved when using CT instead of NO between 61,286 FF ($US 12,282) and 89,382 FF ($US 17912) CT is strictly dominated by CT+HT (100%) Compared to HT, the cost per life year saved by using CT+HT is equal to 20,312 FF ($US 4,071)

Conclusions In terms of frequency of dominance, these results show variability for the ICE ratios, depending on the trial and the prognostic factors taken into account Whereas new drugs in CT are expensive and have not yet shown a major survival improvement, new specific HT could increase the patient benefit in the future

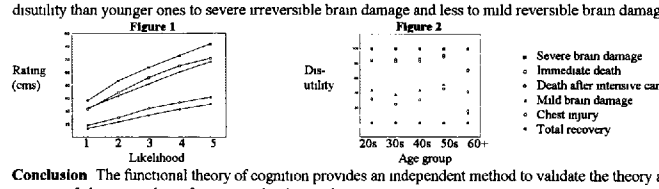
THE CONTRIBUTIONS OF ANDERSON'S FUNCTIONAL THEORY OF COGNITION TO EXPECTED UTILITY ANALYSIS THE CASE OF CARDIOPULMONARY RESUSCITATION

Muñoz Sastre MT, Sorum PC Université François Rabelais, Tours, France, Albany Medical College, Albany, NY

Aim To investigate the theoretical and practical contributions of NH Anderson's functional theory of cognition to the elicitation of utilities for expected utility analysis

Methods 58 French adults rated 27 scenarios of possible outcomes of CPR on a linear scale with anchors of "what would be the best [or worst] for me" In 25, the scenario was that the result would be either total recovery or 1 of 5 adverse outcomes (chest injury, mild reversible brain damage, severe irreversible brain damage, death after intensive care, immediate death) with 1 of 5 likelihoods (1,2,3,4, or 5 out of 10 chances) The 26th case was CPR with 10 out of 10 chances of total recovery, the 27th no CPR with 10 out of 10 chances of immediate death The data were analyzed according to Anderson's methods

Results The shape of the graph (Figure 1) of the mean ratings of each adverse outcome in relation to its frequency is a fan opening to the right, showing that subjects integrated information about the event and its probability in a multiplicative fashion This was true regardless of age group, sex, or educational level The mean disutility attributed to each outcome, relative to 0 for total recovery and 100 for severe brain damage (judged as the worst outcome), is shown in Figure 2 Subjects older than 60 attributed greater relative disutility than younger ones to severe irreversible brain damage and less to mild reversible brain damage



Conclusion The functional theory of cognition provides an independent method to validate the theory and practice of eliciting utilities for expected utility analysis

EXTENDED DOMINANCE IN COST-EFFECTIVENESS DECISION RULES FINE IN PRINCIPLE BUT

Bernie O'Brien, Mark Sculpher,McMaster University, Hamilton, Ontario, Canada, Centre for Health Economics, University of York, Heslington, York, UK

Two dominance relations are typically used when constructing an efficient frontier of mutually exclusive options in cost-effectiveness (CE) analysis (1) simple dominance excludes programs when a single alternative program has lower cost with higher outcome, (2) extended dominance excludes programs when a weighted combination ("mixture") of two other mutually exclusive programs can produce higher (same) outcomes at lower (same) costs We argue that, although the concept of ED is logical at a level of principle, it may be of limited value in practice These limitations are presented using two CE studies as examples alternative drug treatments for gastroesophageal reflux disease and antenatal HIV testing We have identified a series of limitations with the concept of ED relating to the practical difficulty of providing mixed programs The first problem is that there is likely to be resource costs associated with providing mixed programs - for example, relating to allocation of (homogeneous) patients between different programs This limitation could be overcome if these costs were explicitly factored into an analysis The second problem is that some programs exhibit public good characteristics (e g clean air programs, vaccination programs), making it difficult to exclude individuals from their benefits The third limitation is that programs may have non-constant returns to scale, so combining it with other programs for a given population would result in different costs and effects compared to if it were used for the entire population The fourth limitation concerns the equity implications of mixing programs - providing different sections of a clinically homogeneous population with different treatments would run counter to the principles of equal access or equal utilization for equal need A fifth limitation concerns the operational difficulties in providing mixed programs how should a clinician decide which patient has which program? Although allocation by the tossing of a dice would be logical, it may be perceived as politically unacceptable

We conclude that, if ED is to be used as part of the decision rules of a given CE analysis, prior thought should be given to the implications for the design of the analysis (e g factoring in the additional costs of mixed programs) and to how mixed programs could be provided in practice

COLORECTAL CANCER SCREENING A MULTICRITERIA DECISION ANALYSIS

JG Dolan, JE Kerr Primary Care Institute of Highland Hospital and the University of Rochester, Rochester, NY

Current colorectal cancer prevention guidelines endorse several screening procedures and recommend that patients and clinicians choose among them using a shared decision making process This decision is characterized by several alternatives that each have advantages and disadvantages The purpose of this study was to create a multicriteria decision making model to help practitioners and patients make better decisions regarding colorectal cancer screening We performed the analysis using the ideal mode of the Analytic Hierarchy Process (AHP) We defined the goal of the decision as "make the best choice about colorectal cancer screening" Our model contained two major criteria "avoid illness & death due to colorectal cancer" and "avoid illness & death due to screening" The latter criterion was further broken down into 3 subcriteria avoid physical side effects (screening-related deaths, intestinal perforations, major & minor bleeding), minimize inconveniences (number of tests, test preparation and procedure characteristics), and minimize false positive screening tests Decision alternatives were "wait & see" and the 5 screening options endorsed by a multidisciplinary expert panel for average risk patients aged 50 to 85 (Winawer SJ et al Gastroenterology 1997,112 594) annual fecal occult blood tests (fobt), flexible sigmoidoscopy every 5 years, fobt + sigmoidoscopy, barium enema every 5 years, and colonoscopy every 10 years Data for the analysis were obtained from the simulation study published in conjunction with the panel's recommendations Monetary costs were not included Two analyses were run, one for each author's judgments In the first, colonoscopy was the best option (0 199), followed by barium enema (0 184), fobt and sigmoidoscopy (0 163), sigmoidoscopy (0 161), wait and see (0 158), and fobt (0 135) In the second sigmoidoscopy was preferred (0 201) followed by wait & see (0 183), colonoscopy (0 179), barium enema (0 173), fobt and sigmoidoscopy (0 151) and fobt (0 112) Sensitivity analysis revealed that the results were affected by the weights assigned to the 2 major criteria and several subcriteria including test inconveniences, physical side effects, and false positive results

Decisions regarding colorectal cancer screening are complex and likely to vary from person to person Realization of the recommended shared decision making process in practice will probably require the development of clinically appropriate decision aiding methodologies

ASSESSING THE PROBABILITY OF ECONOMIC DOMINANCE WITHOUT DATA FROM A HEAD TO HEAD TRIAL

B van Hout, D Gagnon, P McNulty Dutch Centre for Health Economics and System Science IJssel, The Netherlands ICOM Health Economics, Raritan, NJ

Purpose

To compare the cost effectiveness of two new drugs that have been assessed in randomized clinical trials with similar comparators (for example placebo) but not in a head to head trial

Methods

Under the condition that the costs of treatment (excluding the costs of the new drugs) decrease in a linear manner with additional effectiveness, we can show that the relative cost effectiveness comparing two new drugs depends only on (1) the prices of these drugs, and (2) the net effectiveness of these drugs (in comparison to a common comparator) Subsequently, Fieller's Theorem can be used to estimate - for each price combination - the probability of economic dominance (i e , being both more effective and having a more favorable cost effectiveness ratio)

Results

The methodology is applied comparing two new anti-epileptic drugs, topiramate and gabapentin Net effectiveness, expressed as the additional percentage of patients with a >50% reduction in seizures, is estimated at 28% (14%-43%) for topiramate ($6 03 per patient per day) and at 18% (5%-31%) for gabapentin ($5 78 per patient per day) At prices up to $9 16 per day, topiramate exhibits economic dominance over gabapentin When considering the uncertainties, the probability that topiramate is more effective than gabapentin is estimated at 0 854, and the probability that topiramate is economically dominant over gabapentin is estimated at 0 838 Conversely, the probability that gabapentin is economically dominant over topiramate is estimated at 0 146

Conclusion

When it can be assumed that there is a linear relationship between costs (excluding the costs of the new drug) and effects, data about relative prices and relative effectiveness may be sufficient to assess the comparative cost effectiveness of two drugs, including an assessment of the uncertainties

PREDICTING QUALITY OF WELL-BEING SCORES FROM THE SF-12 HEALTH SURVEY RESULTS FROM THE BEAVER DAM HEALTH OUTCOMES STUDY

WF Lawrence, DG Fryback, KF Gold, BEK Klein, R Klein, Georgetown University Medical Center, Washington, DC, University of Wisconsin, Madison, WI

Purpose To predict Quality of Well-Being (QWB) Index, a utility index, scores from the SF-12 Health Survey, a non-preference based health profile

Methods Data collected on a community sample of 1,430 adults aged 45-87 yo included the QWB and the SF-36 Health Survey The questions in the SF-12 are a subset of those contained in the SF-36 We used two approaches to deriving prediction equations 1) we used stepwise linear regression to identify significant predictors of the QWB scores from the health profile, including squares and first-order cross products, and then performed a best subsets regression using the significant predictors, using the adjusted R-squared and Cp statistics for parameter selection, and 2) we used the regression parameters from the SF-36 to QWB regression from Fryback, et al (Med Decis Making, 1997,17 1-9) using SF-12 questions as surrogates for SF-36 scale scores All prediction equations were derived from 1,356 participants, with the remaining 74 serving as a validation set A second validation set included 57 people with renal insufficiency interviewed at the University of Wisconsin Renal Clinic

Results Correlation between the 8 scales of the SF-36 and the corresponding subset of component questions from the SF-12 ranged from 0 79 (General Health) to 0 97 (Role Emotional) Using the first method of predicting the QWB, we derived a 4 variable equation with an R-squared of 46 4% and 49 0% for community and renal validation sets respectively The second method yielded an R-squared of 45 0% and 53 6% for the 2 validation groups Since the amount of variance explained between the two methods is quite similar, we chose to use the second method for consistency with the SF-36 equation The resulting equation is QWB = 5 92E-1 + 1 26E-3 x PF - 1 17E-3 x MH -1 43E-3 x BP +7 05E-6 x GH x RP + 1 14E-5 x PF x BP + 1 93E-5 x MH x BP, where scales represent raw component question scores linearly rescaled to range from 0 to 100, and PF = Physical Function, MH = Mental Health, BP = Bodily Pain GH = General Health, and RP = Role Physical scales

Conclusions Questions from the SF-12 Health Survey, a health profile, can be used to predict QWB, a utility index The amount of variance explained by the prediction equation is only approximately 5% less than that explained by the full SF-36 equation

AUTOMATING THE CODING OF MEDICAL EVENTS IN MULTIPLE DICTIONARIES

WJ Long, HSF Fraser , JH Ostroff, DC Trost MIT, Cambridge, MA, New England Medical Center, Boston, MA, and Pfizer, Groton, CT

The use of patient data for decision support or analysis depends on having the data coded in a standard medical dictionary The process of turning the words and phrases entered by physicians into standard terms includes abbreviation and acronym expansion, spelling correction, alternative term and spelling adjustment, and handling modifiers Coding data into different dictionaries may cause loss of information and sometimes overcommitment, but automating coding can assist the process

We developed an algorithm for coding terms into an arbitrary medical dictionary The minimum requirement is that we have a list mapping the known terms into the preferred terms From this we generate a mapping of all of the words in terms into the terms that contain them In addition we have a list of expansions of common abbreviations, a list of suffixes, a translation table between spellings, and a mapping of words to their importance From these the algorithm can code a term if the term is known or provide a list of terms with maximum commonality The most important word in the phrase is used to access terms containing the word This list is filtered to find the shortest terms with the most matching or nearly matching words

We tested this algorithm using adverse event data from drug studies In this sample there are 2,776 different physician entered terms The dictionaries used include five in the UMLS plus READ and MEDDRA The table gives the number determined by the algorithm of exact matches, terms that are probable matches, ones for which a list of possibilities was generated, and those with no matches

	MEDDRA	SNMI9	CST93	WHO93	ICD91	MSH96	READ
match	645	638	540	506	628	519	586
close	181	165	86	85	91	78	355
possibles	1258	1796	1637	1462	1705	1876	1704
no term	692	177	513	723	352	303	131

The large number of close terms in READ is a reflection of the British alternative terms Looking at both the matches or near matches and the small numbers of complete non-matches, READ and SNOMED performed best, while MEDDRA and ICD9 had a large number of exact matches

PHYSICIANS' DECISIONS AND AGENCY FOR HEALTH CARE POLICY AND RESEARCH GUIDELINES PREDICT MEDICAL OUTCOMES FOR PATIENTS WITH CHEST PAIN

Lloyd FJ, Reyna VF, Woodard RL University of Arizona, Tucson, AZ
University of Arizona College of Medicine, Department of Medicine, Tucson, AZ

The purpose of this study is to compare the predictive accuracy of physicians' triage decisions for patients with chest pain to those recommended by the Agency for Health Care Policy and Research's (AHCPR) clinical practice guideline for unstable angina A prospective cohort study was conducted of 550 patients (enrolled from March 1997 to May 1998) presenting to the emergency room with symptoms suggesting unstable angina, excluding patients with acute myocardial infarction or shock 61 physicians completed questionnaires eliciting their judgments of the probability of coronary artery disease (CAD) and risk of myocardial infarction (MI) for each patient, as well as triage decisions (to discharge or admit either to telemetry or cardiac intensive care unit) Medical outcomes included readmission rate, length of stay, inpatient and outpatient visits, cardiac procedures, functional status as assessed by the SF-36, and development of subsequent myocardial infarction Significant differences in outcomes were obtained for patients at high risk (according to the AHCPR guideline), compared to low- and intermediate-risk patients who did not differ 40% of patients were discharged, 47 % were admitted to telemetry and 12 % were admitted to cardiac intensive care Physicians' triage decision also predicted outcomes However, there were significant disparities between decisions and guideline recommendations The guideline suggests that low-risk patients be discharged and intermediate- and high-risk patients be admitted Both physicians' decisions and guideline recommendations contributed unique variance to the prediction of outcomes These results suggest that both low and intermediate risk patients may be safely discharged from emergency rooms with appropriate follow-up, contrary to guideline recommendations In addition, patients who were classified differently by physicians and guidelines had outcomes intermediate between low-risk discharged patients and high-risk admitted patients suggesting that guidelines would enhance clinical decision-making for patients with chest pain

ASSOCIATIONS BETWEEN HEALTH STATUS AND UTILITIES IN DEPRESSION

LA Lenert, JR Treadwell, CE Schwartz Stanford University School of Medicine, Stanford CA and Frontier Science, Chestnut Hill, MA

Objectives 1) To determine if patients' utilities for hypothetical states are associated with their health status, and, 2) to determine if patients' utilities for current health differ from their utilities for matched hypothetical states

Methods Patients completed a computerized survey that measured their health status using the Short–Form (SF) 12 questionnaire and their preferences for their current health and three levels of hypothetical health described using the language of the SF-12 Preferences were measured using the standard gamble and a visual analog scale

Setting Patients from three primary care practices in managed–care organizations with various medical illnesses complicated by symptoms of depression

Results 153 patients enrolled in the study and 148 patients completed the survey Utilities for the best and worst states were similar across different levels of health status, however, standard gamble utilities for intermediate health states were *higher* for patients in poorer health than patients in better health (p=0 03) This result was not seen in visual analog scale ratings We also observed an interaction difference between patients' preferences for their current health and the corresponding hypothetical state Patients in better health status tended to rate their current health as worse than the most similar hypothetical state, whereas, patients in poorer health showed the opposite tendency (p=0 035, standard gamble, p=0 030 visual analog scale)

Conclusions In depression, there appear to be significant interactions between health status and values Measurement of preferences in patients that are healthier than those who would benefit from the therapy may result in under-valuation of the treatment's effects There are systematic differences between a patient's values for his or her current health and the corresponding hypothetical states that are associated with health status Use of utilities for current health in cost-effectiveness analyses could under-estimate the significance of partial treatment of more severe health impairments

MEASURING THE APPROPRIATENESS OF CORONARY SURGERY DELPHI PANEL RATINGS VS MORTALITY GAINS BASED ON POOLED TRIAL DATA

JE Hux, CD Naylor, P Slaughter, K Sykora for the Coronary Surgery Appropriateness Group
Institute for Clinical Evaluative Sciences, Clinical Epidemiology and Health Care Research Program and Department of Medicine, University of Toronto, Toronto, ON

Both Delphi panel ratings (DPR) and trials-based scoring algorithms have been used to evaluate the appropriateness of coronary artery surgery (CABG) but have led to differing conclusions We compared these two constructs in a cohort of patients and assessed sources of disagreement

We evaluated in 1000 patients undergoing CABG at 3 Ontario hospitals in 1995-96 using data from a prospective CABG registry supplemented with retrospective chart review Ratings generated by a Delphi panel process (RAND methodology) were applied to classify cases as "appropriate", "equivocal" or "inappropriate" Regression parameters from the pooled trials analysis were applied to estimate survival benefit (SB) as low, moderate or high (average gain of 0, 4 and 9 months respectively) Ratings were compared with Kappa statistics

Results are shown in the table The majority of CABG cases in Ontario are appropriate under either construct Kappa was 0 177, p<0 001 The DPR gave a higher proportion of mid-range scores Among cases rated equivocal, those with high SB - when compared to those with moderate SB - were more likely to have poor left ventricular function, a key factor in SB (83% vs 31%), but also more likely to have elevated Parsonnet scores, reflecting poor operative risk (43% vs 16%) Cases rated appropriate despite low SB had low operative risk, high-risk non-invasive tests and used maximal medication, however also had less severe angina

Survival Benefit	Delphi Panel Ratings (DPR)			
	Inapprop	Equiv	Approp	
Low	2	14	4	20
Mod	3	45	62	110
High	1	171	698	870

Despite similarities, discordance between these two constructs reflects an emphasis on different factors - net survival benefit in a trial setting vs symptoms and perioperative risk It may suggest a range of benefit across appropriate cases and a range of perioperative risk across high SB cases

WOULD MANICS RATHER LAND? FUNCTION AND QUALITY OF LIFE IN PATIENTS WITH BIPOLAR DISORDER

CL Vojta, H Glick, MS Bauer, B Kinosian University of Pennsylvania, Philadelphia, PA, and Providence VAMC, Providence Rhode Island

PURPOSE To compare patient values, mental and physical function in patients with bipolar disorder followed in VA Mental Health clinics

METHODS We evaluated 94 bipolar patients enrolled in a VA multicenter randomized controlled clinical trial Mood states were determined by psychiatrist-applied DSM IV criteria, and the Internal State Scale (ISS) The Short Form-12 (SF-12) for mental and physical function and the EuroQol global health rating thermometer for overall value were used to evaluate if patients experiencing manic symptoms feel better than those in euthymic and depressed states

RESULTS Mood states at time of enrollment included 7 manic, 4 mixed manic/depressive, 25 major depressive, 8 hypomanic, 11 mixed hypomanic/major depressive and 33 euthymic by DSM IV criteria Euthymics scored highest on SF-12 and EuroQol quality of life scales, followed by manic and finally mixed and depressed mood states (see Table) The mental SF-12 subscale and EuroQol were able to reflect differences among mood orders ($p< 05$), but the physcal SF-12 subscale did not Linear regressions to test the predictive ability similarly revealed the mental SF-12 and EuroQol scales' ability to predict ISS categorization ($p< 05$) The ROC area for logistic regression of ISS mood state categorization, EuroQol and SF-12 mental subscale using DSM IV as a gold standard had a values of 83, 72 and 85 Correct classifications of these three scales were 81%, 73% and 77% respectively

CONCLUSION We conclude 1)SF-12 mental subscale and EuroQol are sufficient to accurately reflect mood states in clinical trials, and 2) mania is not a preferred state, with bipolar patients in euthymic mood states reporting the highest quality of life

Table Mean Scores by Mood State

	Euthymic	Manic	Hypomanic	Mixed	Depressed
EuroQol	78 1	71 3	72 2	51 8	54 3
SF-12 Mental	51.1	41 3	43 2	31.7	27 3
SF-12 Physical	49 3	44 1	50 8	42 8	45 3

SCALES FOR ASSESSING PERCEPTIONS OF EVENT PROBABILITY A VALIDATION STUDY

S Woloshin, LM Schwartz, SJ Byram, B Fischhoff, HG Welch VA Outcomes Group, White River Junction, VT, Dartmouth Medical School, Hanover, NH & Carnegie-Mellon University, Pittsburgh, PA

Background Whether patients' medical decisions are based on an accurate perception of the chance of relevant health outcomes is of fundamental interest to clinicians and researchers Surprisingly little work has been done to develop and validate scales to facilitate communication about event probabilities, particularly low probability events (i e , < 1%)

Objective To assess the validity and reliability of 5 scales measuring perceptions of event probability

Design Survey with retest after approximately 2 weeks The perceived probability of 6 events was assessed with 5 scales a crude frequency question and 4 visual analog scales (2 linear and 2 magnifier scales- one of each type labeled with words, one with numbers) The magnifier scales featured a "magnifying glass" to represent probabilities between 0 and 1% on a logarithmic scale. Participants were randomly given surveys differing only in question order

Participants 100 veterans and family members and 107 university faculty and students 178 (88%) completed the retest

Measures For each scale, we assessed (1) Validity (correlation between participants' ranking of the relative probability of 6 events (i e , numbering them from 1 to 6) and their inferred ranking based on responses using the 5 scales) , (2) Test-Retest Reliability (correlation of perceived risks over 2 weeks)

Results The crude frequency question performed poorly on all measures The 4 visual analog scales all had similar validity (correlation between direct and inferred rankings ranged from 72 to 78) and usability ($\leq$ 2% of responses were missing or unusable, 61 to 77% rated the scales as good indicators of their feelings about chance) Scales labeled with words were slightly more reliable than those using numbers (r = .69, 61 vs 57, 54) The presence of the magnifier facilitated expression of perceptions of low probability For example, perceptions of the chance of parenting sextuplets were orders of magnitude lower on the magnifier scales (median perceived chance 10^{7}) than on their linear counterparts (10^{-3}) No such shift was noted in participants' assessment of high probability events such as the chance of catching a cold in the next year

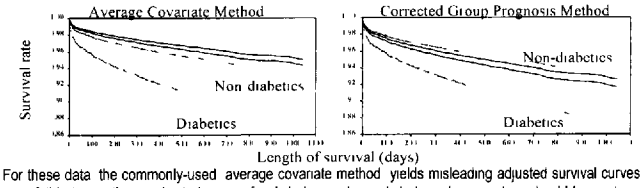
Conclusions The 4 visual analog scales are valid, reliable and usable methods which allow individuals to express their perceptions of event probabilities The magnifier scales make it possible to elicit perceptions of low probability events in a range practically inaccessible on standard linear scales

PLOTTING ADJUSTED SURVIVAL CURVES FROM PROPORTIONAL HAZARDS MODELS A COMPARISON OF TWO METHODS

WA Ghali, H Quan, CM Norris, R Brant, ML Knudtson for the APPROACH (Alberta Provincial Program for Outcome Assessment in Coronary Heart Disease) Investigators University of Calgary Calgary AB

APPROACH is a large inception cohort study which captures all patients undergoing cardiac catheterization in Alberta Canada We used data on 11 804 patients from this study to examine the two year survival experience of diabetics vs non-diabetics while controlling for covariates such as left ventricular ejection fraction coronary anatomy sociodemographic variables and comorbidities Here we present a comparison of two methods for calculating covariate-adjusted survival curves

The most commonly-used method for generating such curves is the average covariate method in which the average values of covariates of interest are entered into a proportional hazards regression equation to generate adjusted survival estimates We used data from the APPROACH study to compare the survival curves generated by the average covariate method with those generated by a newer method - the corrected group prognosis method - in which a survival curve for each level of covariates is calculated after which an average survival curve is calculated as a weighted average of the individual survival curves The resulting adjusted survival curves are shown below along with the corresponding unadjusted survival curves (dotted lines) for diabetics and non-diabetics



For these data the commonly-used average covariate method yields misleading adjusted survival curves which do not fall between the unadjusted curves for diabetics and non-diabetics Investigators should be cautious when using the average covariate method and should consider the 'corrected group prognosis method if resulting adjusted survival curves fail to fall between the unadjusted curves

A MULTILINEAR MULTI-ATTRIBUTE UTILITY FUNCTION FOR THE HEALTH UTILITIES INDEX MARK 3 (HUI3)

GW Torrance, D Feeny, C Goldsmith, W Furlong, Z Zhu, S DePauw, McMaster University, Hamilton, ON, Canada

Purpose To estimate and test a multilinear multi-attribute utility function for the HUI3

Methods Utilities were obtained from a random sample of the general population age 16 and over (Modeling Survey, n=256) The survey design allowed estimation of both a multiplicative and a multilinear multi-attribute utility function The design for the multilinear was a ½ fraction 2^8 factorial, with each health state consisting of a combination of the best and worst levels on the 8 attributes of the HUI3 The design allows the estimation of all 8 main effects, 26 of the 28 possible two-factor interactions, and 4 three-factor interactions The multilinear model was estimated in unrestricted unscaled form, and in a form restricted by Lagrangian multipliers to pass through 0 and 1 The functions were tested regarding their agreement with directly measured utility scores for health states from the same survey not used in the estimation, and for health states from a separate survey (Direct Survey, n=248)

Results The simple main effects additive model was not supported The unrestricted unscaled model contained sixteen two-factor interaction terms and one three-factor term, all statistically significant The restricted scaled model contains all 26 estimable two-factor and 4 estimable three-factor terms, all statistically significant In all models, the coefficients displayed a consistent structure regarding the interactions among attributes all main effects were negative, all two-factor interaction terms were positive, and all three-factor terms were negative This appears to mimic the type of interactions implied by the multiplicative model The multilinear model with all estimable terms, in both unrestricted unscaled (A) and restricted scaled (B) form, and the multiplicative model (C), predicted the measured utility scores of the various sets of validation states as follows

	Mean Difference(P-M)			Standard Deviation		
	A	B	C	A	B	C
Modeling Survey marker states (n=3)	0 020	0 189	0 011	0 068	0 235	0 084
Direct Survey marker states (n=3)	0 031	0 200	0 022	0 064	0 246	0 095
Direct Survey all states (n=73)	0 019	0 166	-0 008	0 089	0 183	0 103

Conclusions The multilinear formulation was useful in understanding the utility interactions among attributes, but for practical applications the more parsimonious multiplicative model is recommended

TIME PREFERENCES, FUNGIBILITY, AND DECISION PERSPECTIVE

GB Chapman Rutgers University

According to normative theory, the same temporal discount rate should be applied to delayed health and monetary outcomes because, if money and health are fungible, applying different discount rates to the two will result in inconsistencies and money pumps (Keeler & Cretin, 1983) Arguably, health is always tradeable for money at the societal level, since money saved on one health intervention can be used to fund another At the individual level, however, health and money are often not fungible Thus, one reason for previous demonstrations that individuals do not use the same discount rate for health and money may be that respondents did not view health and money as tradeable In two experiments, participants read a scenario in which health and money were either tradeable (money could be used to purchase a health-improving medication) or were not tradeable (medication was not available for sale) Participants specified the amount of money in 3 years that was just as attractive as $1000 now and the amount of medication after a 3-year delay that was just as attractive as a 12-month supply now In Experiment 1, 106 college students and 100 train station commuters responded to the scenarios from an individual perspective As predicted, the Spearman correlation between health and money discount rates was higher in the tradeable condition (r= 0 44, N=105, p < 0 0001) than in the not tradeable condition (r= 0 02, N=101, p > 0 8) In Experiment 2, 144 college students responded to the scenarios from an individual or policy maker perspective For the individual perspective, the correlation between health and money discount rates was higher in the tradeable than the not tradeable condition In the policy maker perspective, fungibility had a smaller effect on the correlation (see table) This result can be explained by the fact that at the policy level, health and money are always tradeable regardless of trade-offs available to individual patients Thus, policy decisions may be more in line with normative theory

Spearman correlations between health and money discount rates, *p<0 01

	perspective	
	individual	policy
tradeable	r = 0 50* N = 36	r = 0 41* N = 37
not tradeable	r = 0 01 N = 35	r = 0 27 N = 36

BIAS IN MEDICAL STUDENTS EFFECT OF PATIENT RACE AND GENDER ON PERCEIVED HEALTH VALUES

SS Rathore, LA Lenert, KP Weinfurt, A Tinoco, CK Taleghani, W Harless, KA Schulman Clinical Economics Research Unit, Georgetown University Medical Center, Washington DC, Department of Medicine, Stanford University, Palo Alto, CA and Interactive Drama, Bethesda, MD

Diminished health care by women and minorities in the United States has been documented in several settings While prior studies have attributed differences to several factors, few have investigated the possible role of physician bias We sought to determine if physician bias exists early in medical training by evaluating the influence of race and gender on students' perceptions of patients

164 first and second year medical students at a private medical school were randomized to view a video of a 55 year old African-American female (AAF) or 55 year old Caucasian male (CM) patient actor presenting identical angina symptoms Patient actors read from identical scripts and were standardized in presentation, tone and appearance A computer assisted survey instrument elicited students' preferences for the patient actor's health state by standard gamble (SG) and visual analog (VAS) utility metrics Students also rated type of chest pain (definite/possible/no angina), probability of coronary artery disease and subjective perceptions of the patient actor on 15 Likert scale items

Students imagining their own quality of life with coronary artery disease assigned a lower utility value to symptoms when presented by the African-American female than the Caucasian male patient (VAS 80 3 vs 87 1, p<0 02, SG 67 0 vs 71 8, p<0 01) Further analysis of SG measures found the observed effect was primarily due to lower health values assigned to the AAF by Caucasian students (SG 74 1 AAF vs 89 2 CM, F(1,157)=3 98, p=0 03) and male students (SG 76 4 AAF vs 84 4 CM, F(1,157)=3 73, p=0 055) Despite assigning lower health values for the AAF, students reported similar probabilities of coronary disease (x=73 8 AAF vs x=77 5 CM, p<0 60) Further, students viewed the AAF more favorably than the CM from a psychosocial perspective, rating her as more friendly (p<0 001), a better communicator (p<0 005) and having a more positive affect (p<0 003) than the CM

Our findings indicate race and gender influence students' perception of a patient's symptoms Differences in health values at an early stage in training may be explained by latent prejudices which may impair a student's ability to allocate health resources in an unbiased manner We recommend the inclusion of diversity training in medical curriculum nationwide to address this issue

QUALITY OF LIFE AFTER COMBINED PANCREAS-KIDNEY TRANSPLANT The case of response shift and measurement by proxy

EMM Adang, D Postulart, Maastricht University, Faculty of Health Sciences, Dept BEOZ, Maastricht, The Netherlands

The aim of this study is to investigate the way quality of life (QL) outcomes can differ depending on the way they are measured i e , retrospectively or prospectively, and if the use of patients or proxies in QL research lead to the same results A cohort analytic study was performed to assess QL in 22 patients with type I diabetes and end-stage renal disease (IDDM-I with ESRD) in whom a combined pancreas-kidney transplant was performed sucessfully The patients were interviewed at home before transplantation, and at 6,12 and 18 months following transplantation At these follow-up assessments, patients were also asked to re-evaluate their pretransplant QL (i e , a so called "thentest") QL was assessed with a 10cm Visual Analogue Scale (VAS) ranging from (0) worst imaginable QL to (10) best imaginable QL At a later stage 55 students in "Health Sciences" at our University were asked to evaluate the "IDDM-I with ESRD"health state with the same VAS as in the patient study Analyses revealed that before transplantation the mean QL score was 5 23cm for the patients and 2 78cm for the proxies (Mann whitney U, p<0 000) At the follow-up assessments, patients retrospectively assessed their pretransplant QL at 3 27,3 14 and 3 05, which did not differ significantly from the proxies QL rating of 2 78 (Mann Whitney U, p=0 266, p=0 794, p=0 838) These findings show that assessing QL in a patient study in a retrospective way lead to the same results as using proxies to assess a specific health state Prospective QL measurement leads to different outcomes depending on the use of patients or proxies Apparently the value a patient assigns to a health state does not only depend on that health state but also on how that health state compares to the health state the patient might have had if he or she had made a different choice (i e no transplant and having an IDDM-I with ESRD health state) The findings in the patient study can be explained by response-shift effects that result from adaptation to changing health status (i e , from relatively poor to relatively well) The fact that proxies (having no reference about a health state other than the description given and therefore cannot be adapted to that particular health state) rate the IDDM-I health state the same as the patients who retrospectively assessed their pretranplant QL suggests that pre intervention QL, measured prospectively, is hightened by adaptation to the imperfect health state and for that reason leads to different QL outcomes than retrospectively or by proxy assessed QL

MODELING OF MULTIOUTCOME LIFE EXPECTANCY LOTTERY CHOICES, CLASSICAL EXPECTED UTILITY (EU) VERSUS PROSPECT-LIKE (PT) THEORIES

LM Bernstein, GB Chapman, AS Elstein, University of Illinois at Chicago

Choices between multioutcome life expectancy lotteries presented in alternative editings were modeled with PT and EU Scenarios described a chronic disease which, if wholly untreated would result in living about two years Treatment with a baseline drug would prolong life to about 20 years if accompanied by 1 of 2 supplementary drugs which themselves have additional effects on longevity of increasing, decreasing or either increasing or decreasing survival Condition 1 (50 problems, 116 subjects) used "segregated editing", baseline drug effects common to both treatment options were described in the text, but lottery displays presented only the gains or losses attributable to two supplementary drugs Condition 2 (100 subjects) used "integrated editing", the combined effects of both baseline and supplementary drugs were presented in the displayed lotteries All compared lotteries had equal value life expectancies Lottery pairs differed in numbers of outcomes or probability distribution The fit of each theory to the data was measured by agreement between observed majority preference and the lottery with the higher calculated expected utility (or Value) Modeling was limited to problems with moderate or high majority preference rates (56% to 79%), 44 in Condition 1 and 36 in Condition 2 RESULTS For the segregated versions , parameter fitting showed that the utility function was s-shaped, concave in the gain and convex in the loss domains, and probability weighting did not improve the fit The overall concordance rate was 82% (for gains, 80%, for losses, 75%, for mixed lotteries, 100%) The overall agreement rate for EU was 48%, close to chance With segregated presentations, subjects established the 20 year life expectancy effect of the baseline drug as their reference point, and their choices largely conformed with PT models For the integrated versions of these formally equivalent problems, results are the same for both EU and PT since all problems fall into the gain domain Overall agreement was only 58%, precluding reliable inferences regarding the shape of the utility function or underlying decision processes Despite full disclosure of effects of both baseline and supplementary drugs and visual presentation of the lotteries based on net effects of both drugs (the classical EU theory form), the subjects did not use the net years of life expectancy from current time to guide their preferences Behavior was not in accordance with either EU or the PT-like theories We speculate that the decision processes used resulted from conflict between the textual description of years of life lost due to the supplementary drug effects coupled with the visual presentation of the overall net effects as pure gain lotteries CONCLUSIONS On balance, these data provide little support for EU theory as a descriptive theory of choice, and suggest that medical decision making might benefit from explorations of circumstances in which s-shaped utility functions and prospect-like theories operate