predict within-sample subgroup mean EQ-5D scores, the mean errors (mean absolute errors) range from 0.021 to 0.077 (0.045 to 0.083). When predicting baseline cohort EQ-5D scores using published mean dimension scores the models produce mean errors ranging from 0.048 to 0.099 with 76% of values correct to within the minimal important difference. When predicting out-of sample incremental differences between study arms and incremental changes over time, over 71% of values are within the minimal important difference. **CONCLUSIONS:** The models provide researchers with a mechanism to estimate EQ-5D utility data from published mean dimension scores. This research is unique in that it uses mean statistics from published studies to validate the results. While further research is required to validate the results in additional health conditions, the algorithms can be used to derive additional preference-based measures for use in economic analyses.

PMC34

INTERNATIONAL SURVEY ON WTP FOR ONE ADDITIONAL QALY GAIN—HOW MUCH IS THE THRESHOLD OF COST-EFFECTIVENESS ANALYSIS

Shiroiwa T[1], Sung YK[2], Fukuda T[1], Bae SC[2], Tsutani K[1]
[1]The University of Tokyo, Tokyo, Japan, [2]Hanyang University College of Medicine, Seoul, South Korea

OBJECTIVE: Threshold of cost-effective analysis is thought to be £20,000–£30,000 in UK and $50,000–$100,000 in the US, however it is known that these values are not based on explicit scientific evidence. We measured WTP for one additional QALY gain to determine the threshold of cost-effectiveness analysis. **METHODS:** We measured willingness to pay (WTP) for additional one year survival in perfect health status to determine threshold of incremental cost-effectiveness ratio (ICER) by the internet. The number of subjects is 1000 (500 male and 500 female in their 20 s to 50 s) in Japan, Republic of Korea (ROK), Australia and UK. We asked them four kinds of WTP, i.e., 1) WTP for their own additional QALY: (WTPsel); 2) WTP for their own additional QALY 5 years later: (WTP5sel); 3) WTP for their family's additional one QALY: (WTPfam); and 4) the amount they think society should pay for someone's additional one QALY: (WTPsoc). The double bound dichotomous choice was applied to this research analyzed by nonparametric Turnbull method. **RESULTS:** The value of WTPsel is ¥5 million (Japan), ROK70 million Won, AUS$64,000 (Australia) and GBP23,000 (UK). Discount rate of outcome was calculated by comparing WTPsel with WTP5sel, and it is estimated to be 6.8% (Japan), 3.7% (ROK), 1.9% (Australia) and 2.8% (UK). The order of four kinds of WTP is WTP5sel < WTPsel < WTPsoc < WTPfam in Japan and ROK, and WTP5sel < WTPsel < WTPfam < WTPsoc in Australia and UK. **CONCLUSION:** Considering the value of WTPfam (¥6.4 million), we think the threshold of ICER should be determined to be ¥5 million to ¥6 million per QALY in Japan. This result also shows the threshold and discount rate adapted by NICE is reasonable. The difference of four WTP's order may represent cultural gap between Asia and Western world.

PMC35

A SYSTEMATIC REVIEW OF APPLICATIONS OF CONJOINT ANALYSIS IN MEDICINE

Kinter ET[1], Bridges JF[1], McCormick C[2], Kidane L[3]
[1]Johns Hopkins University, Bloomberg School of Public Health, Baltimore, MD, USA, [2]Johns Hopkins Medical University, Baltimore, MD, USA, [3]Johns Hopkins University, Baltimore, MD, USA

OBJECTIVE: To conduct a systematic review of studies that employ conjoint analysis methodology in outcomes research in medicine published between 1985 and 2006 in order to document: i. clinical areas of focus; ii. sample size; iii. method of design; iv. method of analysis and other quality parameters. **METHODS:** Papers published between 1985 and 2006 were identified on Medline, using the search terms conjoint analysis/analyses, stated preference(s), discrete choice analysis/analyses, and discrete choice modeling/experiments(s). All papers were then reviewed for content by three reviewers, with papers not actually related to conjoint analysis being deleted. Remaining papers were then classified as: i. a clinical application; ii. an application focusing on heath systems; or iii) papers focusing on methods. We classified all clinical applications by ICD-9 codes and identified key methodological characteristics such as sample size, design methodology and type of analysis when available. **RESULTS:** We began our review in 1985 due to the limited number of publications between 1970 and 1985 (n = 4). Post 1985, 27% (n = 48) discussed the methodology of conjoint analysis with no application, 25% (n = 45) focused on health systems in medicine and 48% (n = 86) were clinical and therapeutic applications of conjoint analysis. There is a steady increase in the use of conjoint in medicine between 1985 and 1999, most common clinical areas being HIV, cancer and STI. The average sample size is 267. Use of orthogonal arrays was the most common design, 74% (n = 50) followed by adaptive conjoint analysis, 16% (n = 11). Primary analysis techniques were probit and logistic regression, 39% (n = 27) and 30% (n = 21) respectively. **CONCLUSIONS:** We find insufficient information on the methods used in a significant proportion of manuscripts reviewed. Given the importance of preference elicitation in medicine, we need to focus on developing standardized research practices for the application of conjoint analysis in outcomes research.

PMC36

THE USE OF A MOBILE PHONE FOR ASSESSING MOOD AND PERFORMANCE IN EVERYDAY LIFE

Tiplady B, Oshinowo B, Thomson J, Drummond GB
University of Edinburgh, Edinburgh, UK

OBJECTIVE: Handheld computers and mobile phones allow assessments of a variety of aspects of behaviour, experience and performance in an everyday life setting. We have evaluated the use of a mobile (cell) phone to collect data on alcohol consumption, subjective state, and psychomotor performance and to assess the relationships between these variables. **METHODS:** Thirty eight healthy volunteers (20 male) aged 18–54 years (mean 22.8) took part. The study program ran on any phone supporting downloadable Java® applications. Text (SMS) messages were sent twice a day for 14 days according to a randomised schedule, and volunteers completed their entries as soon as possible after receiving the text. They recorded alcohol consumption and mood, and completed performance tasks. **RESULTS:** Compliance was good, with responses being made to over 80% of text messages. Subjective drowsiness (% of visual analogue scale length) ranged from 35.7–42.9 between 09:00 and 19:00 and from 51.5–57.7 between 21:00 and 01:00. Mean ratings of drunkenness assessed between 21:00 and 01:00 were 5.5% (S.D. 12.2) where no alcohol consumption was recorded and 53.1% (S.D. 21.0) where five or more drinks had been consumed in the last six hours (ANOVA $p < 0.001$). Mean errors on a memory scanning task were 6.6 (S.D. 4.7) with no alcohol and 10.5 (S.D. 6.8) after 6 or more drinks (ANOVA $p < 0.01$). **CONCLUSION:** Both objective and subjective measures showed the expected relationship with reported alcohol consumption. The widespread use of mobile phones means that this type of study can be carried out economically and with reasonable train-