Methodological Issues in the development of an Australian algorithm for the EQ-5D

Rosalie Viney, Richard Norman, Madeleine King, Deborah Street, Paula Cronin, Stephanie Knox,
Julie Ratcliffe, John Brazier

This project is funded by a NHMRC Project Grant

Quality Adjusted Life Years (QALYS)

- QALYs used to value health outcomes for economic evaluation
- Combines length of life and quality of life in a single measure
- QALY weights represent an index of health states/quality of life
- Full health =1, Death = 0, Worse than death<0
- The total value of a health profile (in QALYs) is then

$$TotalValue = \int Q(t)dt$$

where Q(t) is the QALY weight at time t

- Methods for deriving QALY weights include Visual Analogue Scales, Time Trade-Offs (TTO), Standard Gamble (SG) (potential for DCEs)

Multi-attribute utility instruments (MAUIs)

Generic Quality of life Instruments
(SF-6D, EQ-5D, HUI, SF-36)
General population sample

↓ Preference based measurement (SG, TTO, DCE)

Health state scores

↓ Model and data analysis

Scoring ‘algorithm’ developed

QALY weights for each health state

The EQ-5D

Mobility
I have no problems in walking about
I have some problems in walking about
I am confined to bed

Self-care
I have no problems with self-care
I have some problems washing and dressing myself
I am unable to wash and dress myself

Usual activities (eg. work, study, housework, family or leisure activities)
I have no problems with performing my usual activities
I have some problems with performing my usual activities
I am unable to perform my usual activities

Pain/discomfort
I have no pain or discomfort
I have moderate pain or discomfort
I have extreme pain or discomfort

Anxiety/depression
I am not anxious or depressed
I am moderately anxious or depressed
I am extremely anxious or depressed

EQ-5D

- Developed by the EuroQoL group
- Standard version comprises:
  - descriptive system questionnaire
  - visual analogue scale
- 5 dimensions, 3 levels per dimension = 243 health states
- Easy to use but “lumpy”
- Algorithm based on use of TTO to directly value a selection of states and OLS regression for estimation

EQ-5D algorithm (UK)

	Full Health	1
	Constant	-0.081
Mobility	M2	-0.069
	M3	-0.314
Self-care	SC2	-0.104
	SC3	-0.214
Usual activities	UA2	-0.036
	UA3	-0.094
Pain/Discomfort	PD2	-0.123
	PD3	-0.386
Anxiety/Depression	AD2	-0.071
	AD3	-0.236
Any level 3	N3	-0.269

- Examples:

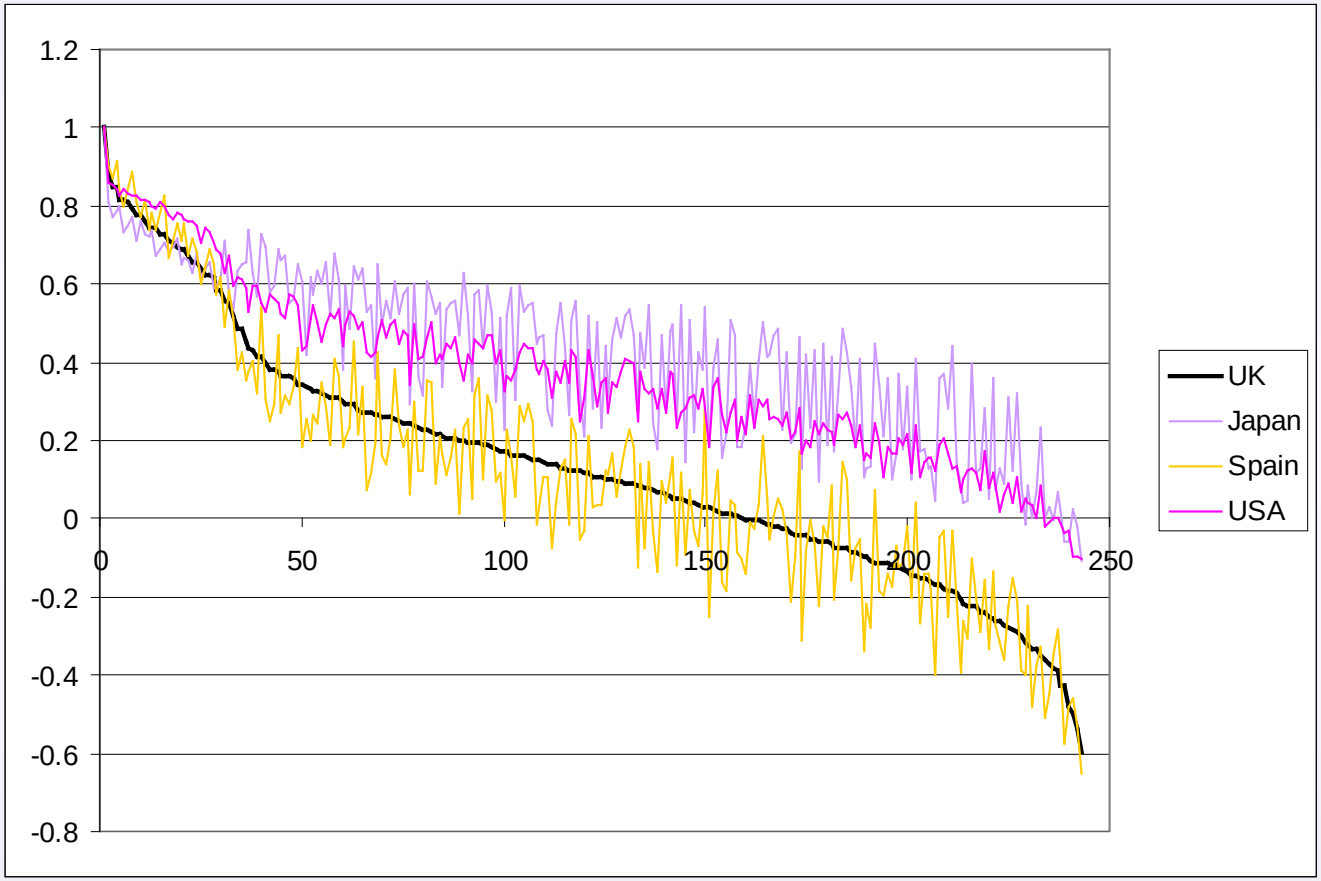
11111	1
22222	0.516
33333	-0.594
11211	0.883
21311	0.487
33121	-0.01

- 84/243 HS are “worse than death”
- No health states with values between 0.883 and 1.0
- Shifts between states can have large impacts on utility

EQ-5D and economic evaluation

- EQ-5D is the most widely used multi-attribute utility instrument (MAUI)
  - Recommended by NICE
  - One of several recommended by PBAC in Australia
- Existing algorithms are based on Time Trade Off or Visual Analogue Scale (NZ)
- Algorithms exist for many countries
  - UK, Japan, Spain, Denmark, Germany, USA, Zimbabwe, Slovenia, Belgium etc – but not Australia
- Australian economic evaluation relies on international studies for EQ-5D weights, typically UK, but others available
- Choice of algorithm can lead to gaming

Scoring comparison



* This figure uses the UK preference ordering as a baseline for comparison

Aims

- Develop an Australian algorithm for the EQ-5D
- Explore methodological issues in algorithm development
- Investigate potential for DCE approach to algorithms
- Compare DCE and standard approach (TTO)

Issues with existing algorithms

- Valuation method
  - TTO accepted as standard for EQ-5D, but resource intensive
  - DCEs offer potential new approach
    - Alternatives described in terms of attributes
    - Experimental design principles used to choose a sample of choice sets to allow for efficient estimation of parameters of interest
    - Readily applicable to MAUI context
    - Use of DCE approach allows greater coverage of the response surface
- Survey administration
  - Potential for on-line administration
- Choice of health states
  - Most valuation surveys based on direct valuation of 42 or 17 health states + death
  - Limited investigation of interactions
- Duration of health states
  - Typically fixed at 10 years

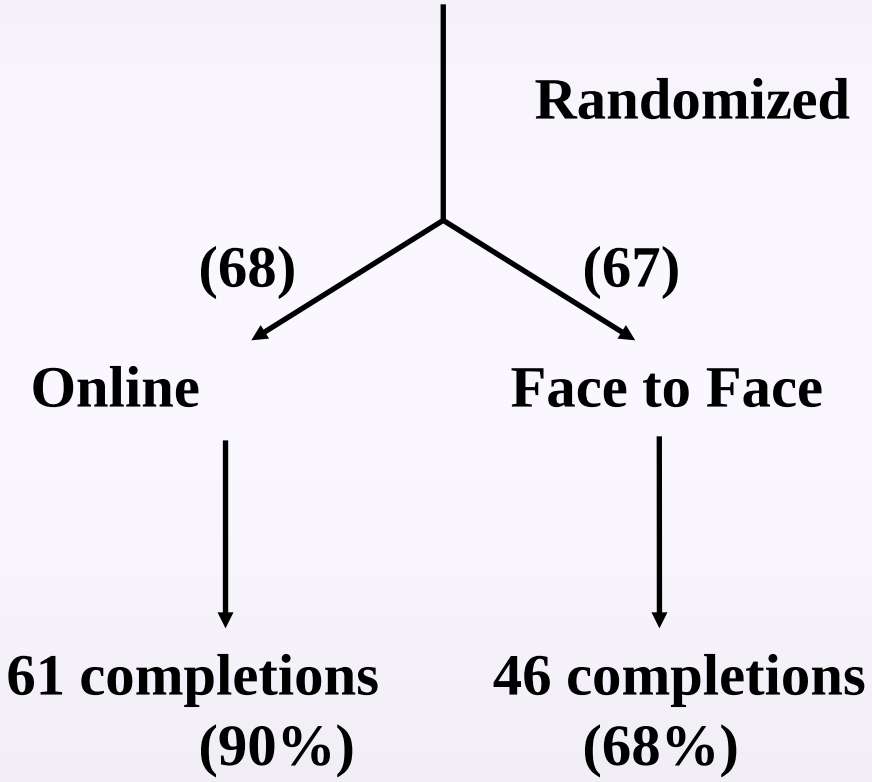
Overall approach

- Undertook both a conventional TTO and a DCE
- In the TTO
  - Varied from the conventional approach in terms of mode of admin and selection of health states
- In the DCE
  - Explored selection of health states and the effect of duration

Issue 1: Mode of Administration

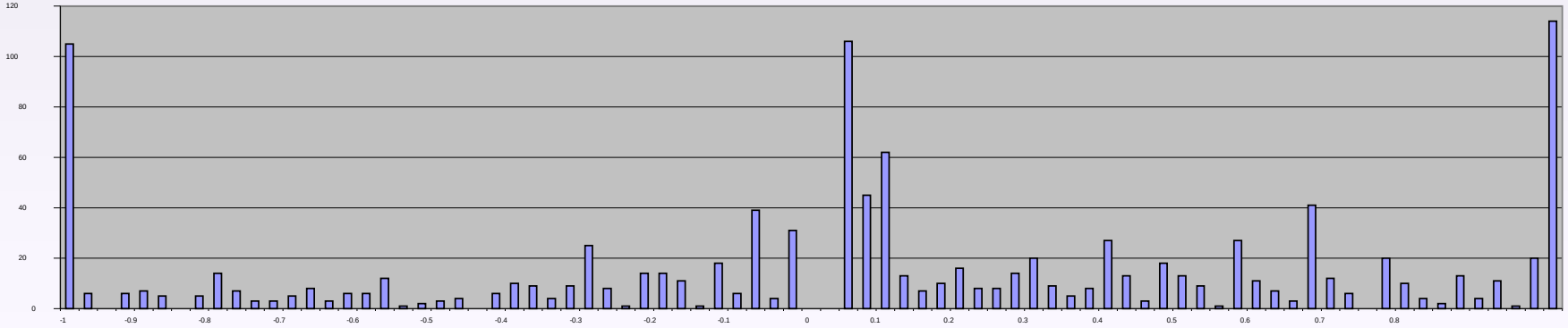
- TTO conventionally interviewer administered with physical props
- Potential for electronic data collection
  - minimise errors in data entry
  - considerably cheaper, but performance may be different
- Options for data collection
  - Fully online data collection via on-line panels
  - On-line administration with conventional recruitment
  - Assisted on-line administration with conventional recruitment
  - Conventional administration and recruitment
- Explored mode of administration in the TTO via a pilot study
  - Convenience sample
  - Randomised to on-line or conventional interviewer administration
  - Used a main-effects design with all respondents valuing all 18 states

Completion rates

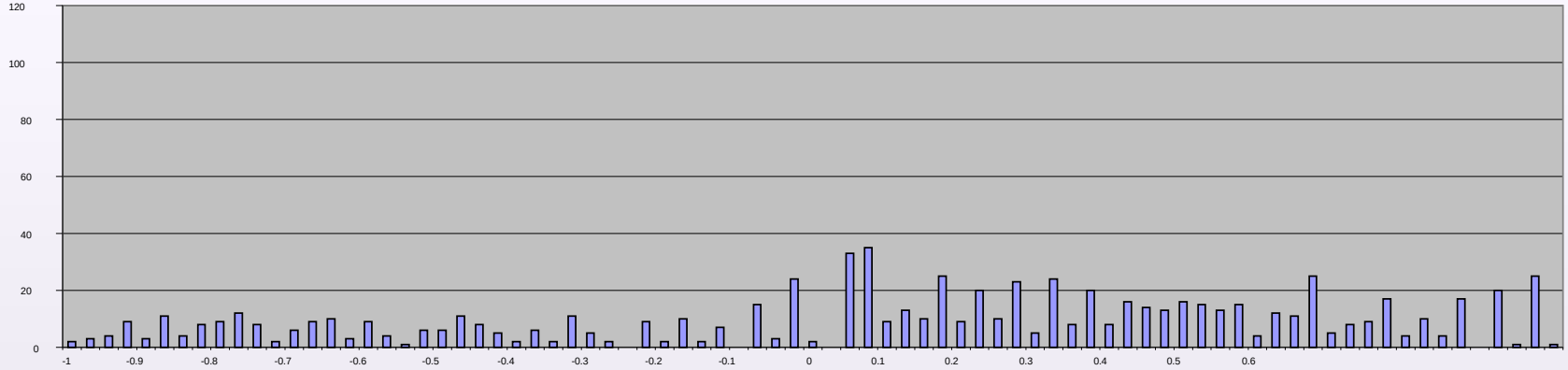


Frequencies of valuations

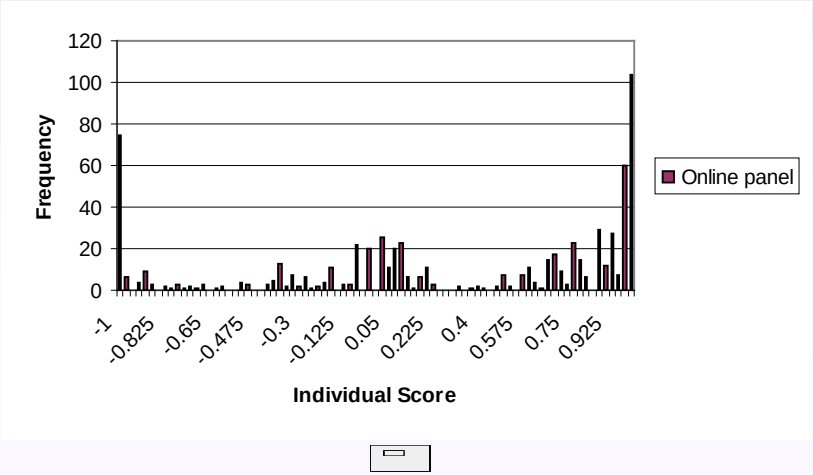
Student TTO Online

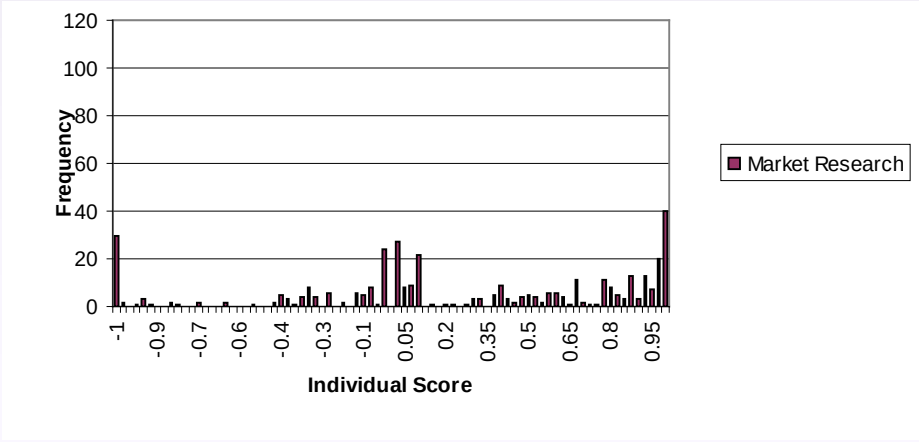


Student TTO Face to Face

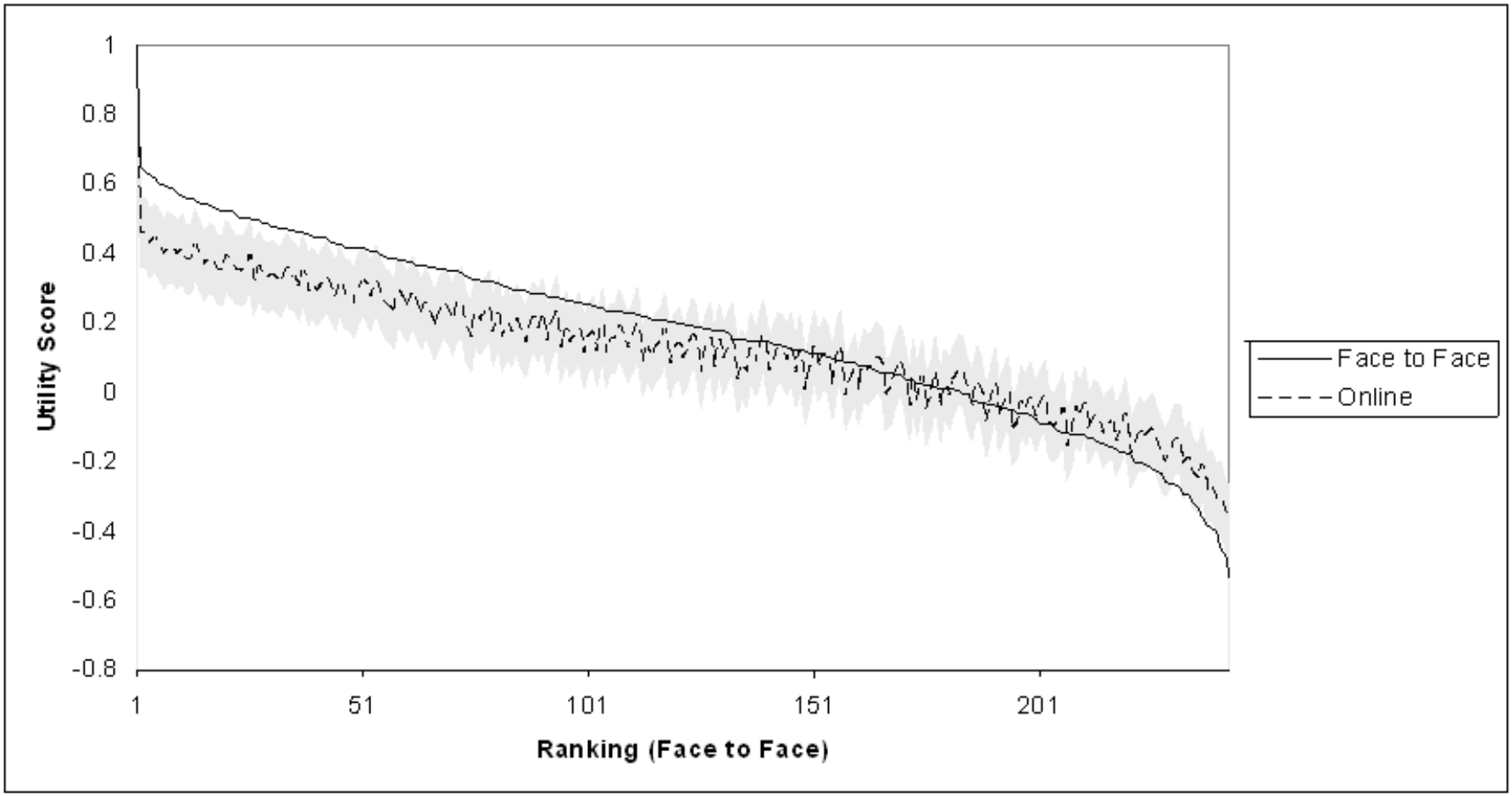


Further comparison of modes of administration from different samples





Assessing mode of administration



Issue 2: Selection of health states

- Two main approaches to selecting which health states are directly valued
  - Dolan approach (original UK valuation study)
    - 43 health states (42 + “pits” state) defined as either “very mild”, “mild”, “moderate” or “severe”
    - Each respondent values 13 of 43
    - Excluded implausible states (UA1 with Mob3 or SC3)
  - Tsuchiya approach
    - Each respondent values 17
    - Subset of the original Dolan set described as “minimum required to estimate the value set”
  - Both approaches have led to similar algorithms, which include an N3 interaction term

Directly Valued States

Category	Tsuchiya	Dolan (the number valued by each respondent is in parentheses)
Full health	**11111**	**11111**
Very Mild	**11112, 11121, 11211, 12111, 21111**	(2 from 5) **11112, 11121, 11211, 12111, 21111**
Mild	**11113, 11131, 11133, 11312**	(3 from 12) 11122, **11131, 11113,** 21133, 21222, 21312, 12211, **11133,** 22121, 12121, 22112, **11312**
Moderate	**13311, 32211, 32313, 22222**	(3 from 12) 13212, 32331, **13311,** 22122, 12222, 21323, **32211,** 12223, 22331, 21232, **32313, 22222**
Severe	**23232, 32223, 33323**	(3 from 12) 33232, **23232,** 23321, 13332, 22233, 22323, **32223**, 32232, 33321, **33323**, 23313, 33212
Pits State	**33333**	**33333**

Issues

- In the 43 or 17 health states approach used in TTO
  - Focus on obtaining precise estimates of a selection of health states and interpolating between these
  - Focus on relatively better health states
  - Greater precision at one end of the scale
  - Pairs of attributes missing
  - Not all interactions can be estimated
  - Underlying design strategy opaque
- In the EQ-5D all 243 health states could be valued
- Little discussion in literature of selection of impact of selection of health states

Simulation study

- Five approaches to selection of health states
  - Dolan (n=43), Tsuchiya (n=17)
  - Orthogonal Main Effects Plan (OMEP) (n=18)
  - Full factorial (FF) (n=243)
  - Exclude implausible states (FFP) (UA1 or SC1 with Mob3) (n=198)
- Two underlying utility functions assumed
  - Main effects model with an N3 term
  - Main effects + two factor interactions between dimensions
- Monte carlo simulation of 100 independent samples for each design
  - N=300 respondents each valuing 15 states
  - Simulated data used to estimate the assumed utility functions
  - GLS estimation

Findings from simulation study

- For the “N3” model (no 2 factor interactions)
  - All design approaches perform similarly
- If 2 factor interactions are assumed
  - Only FF and FFP allow estimation of all these parameters
  - Dolan allows estimation of some
  - FF has the least bias and best precision
  - Potential trade-off between statistical and respondent efficiency in FF
- FFP considered to be the most appropriate design approach

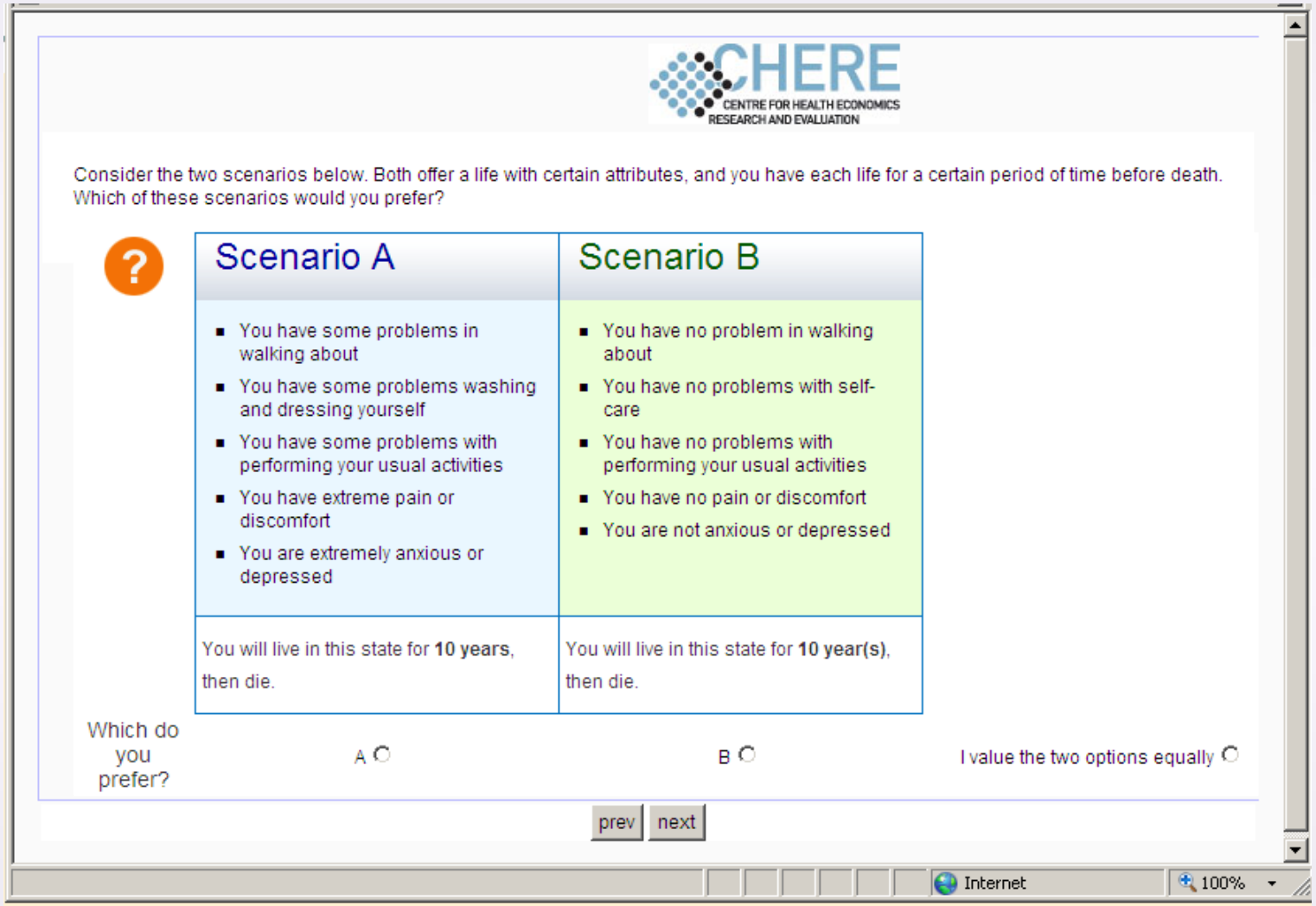
TTO main study

- Respondents recruited from an existing panel
  - Recruited in four locations
  - Sydney Metro, Sydney Suburbs, Melbourne, Orange
  - Sample comparable in age, gender, country of birth and household income to Aust'n population
- 417 completed the task
- Interviewer assisted on-line administration
- Each respondent valued 11 randomly selected health states + pits state

TTO Task

- Rank the health state as worse than death or better than death
- Ten year duration
- States better than death
  - x years in FH vs 10 years in HS*i* ($v_{HSi} = x/10$)
- States worse than death
  - Immediate death vs x years in FH, then (10-x) yrs in HS*i* ($v_{HSi} = -(10-x)/x$),
  - “transformed” to [(x/10)-1] (bounded at -1)
- Ping-pong method

TTO Task


CHERE
CENTRE FOR HEALTH ECONOMICS RESEARCH AND EVALUATION
Consider the two scenarios below. Both offer a life with certain attributes, and you have each life for a certain period of time before death. Which of these scenarios would you prefer?
Scenario A
You have some problems in walking about
You have some problems washing and dressing yourself
You have some problems with performing your usual activities
You have extreme pain or discomfort
You are extremely anxious or depressed
You will live in this state for 10 years, then die.
Scenario B
You have no problem in walking about
You have no problems with self-care
You have no problems with performing your usual activities
You have no pain or discomfort
You are not anxious or depressed
You will live in this state for 10 year(s), then die.
Which do you prefer?
A
B
I value the two options equally
prev
next
Internet
100%

Variables

Variable	Definition	Used in
MO2	1 if mobility is level 2; 0 otherwise	All models
MO3	1 if mobility is level 3; 0 otherwise	All models
SC2	1 if mobility is level 2; 0 otherwise	All models
SC3	1 if mobility is level 3; 0 otherwise	All models
UA2	1 if mobility is level 2; 0 otherwise	All models
UA3	1 if mobility is level 3; 0 otherwise	All models
PD2	1 if mobility is level 2; 0 otherwise	All models
PD3	1 if mobility is level 3; 0 otherwise	All models
AD2	1 if mobility is level 2; 0 otherwise	All models
AD3	1 if mobility is level 3; 0 otherwise	All models
N3	1 if any dimension is level 3; 0 otherwise	Models 2/2b
XXa x YYb	1 if dimension XX is level a (where a≠1) and dimension YY is level b (where b≠1) and XX ≠ YY; 0 otherwise	Models 3/3b

Estimation: GLS with random effects

Model	Constant constrained to unity	N3 term	2 factor inter-actions	No of parameters	LL	AIC	BIC
1	No	No	No	11	-3070.7	6167.32	6252.06
1b	Yes	No	No	10	-3092.8	6209.68	6287.90
2	No	Yes	No	12	-3029.5	6086.96	6178.21
2b	Yes	Yes	No	11	-3037.5	6101.05	6185.79
3	No	No	Only for level 3	21	-2987.8	6021.51	6171.43
3b	Yes	No	Only for level 3	20	-2999.1	6042.21	6185.60
4	No	No	All	49	-2975.8	6053.57	6385.99
4b	Yes	No	All	48	-2983.6	6067.13	6393.03

Results cont.

- Constant term positive and significant and significantly different from one in all models
- N3 term significant in model 2
- Some 2 factor interactions significant in Models 3 and 4
  - combinations of level 3 in Mobility, Pain and Anxiety
  - interaction terms tended to be positive

Comparing Models 2 and 3

	Model 2	Model 3		Model 3 (cont)
Constant	0.910 (0.022)**	0.895 (0.022)**	MO3_SC3	0.064 (0.034)
MO2	-0.071 (0.014)**	-0.068 (0.014)**	MO3_UA3	-0.025 (0.034)
MO3	-0.264 (0.019)**	-0.374 (0.033)**	MO3_PD3	0.092 (0.033)**
SC2	-0.104 (0.016)**	-0.087 (0.016)**	MO3_AD3	0.013 (0.035)
SC3	-0.169 (0.017)**	-0.267 (0.025)**	SC3_UA3	-0.055 (0.030)
UA2	-0.048 (0.017)**	-0.053 (0.017)**	SC3_PD3	0.090 (0.030)**
UA3	-0.085 (0.018)**	-0.139 (0.024)**	SC3_AD3	0.105 (0.031)**
PD2	-0.082 (0.015)**	-0.068 (0.015)**	UA3_PD3	0.025 (0.030)
PD3	-0.268 (0.016)**	-0.449 (0.022)**	UA3_AD3	0.043 (0.030)
AD2	-0.086 (0.015)**	-0.097 (0.015)**	PD3_AD3	0.185 (0.029)**
AD3	-0.214 (0.016)**	-0.397 (0.023)**		
N3	-0.180 (0.020)**			

** Significant at 1% level

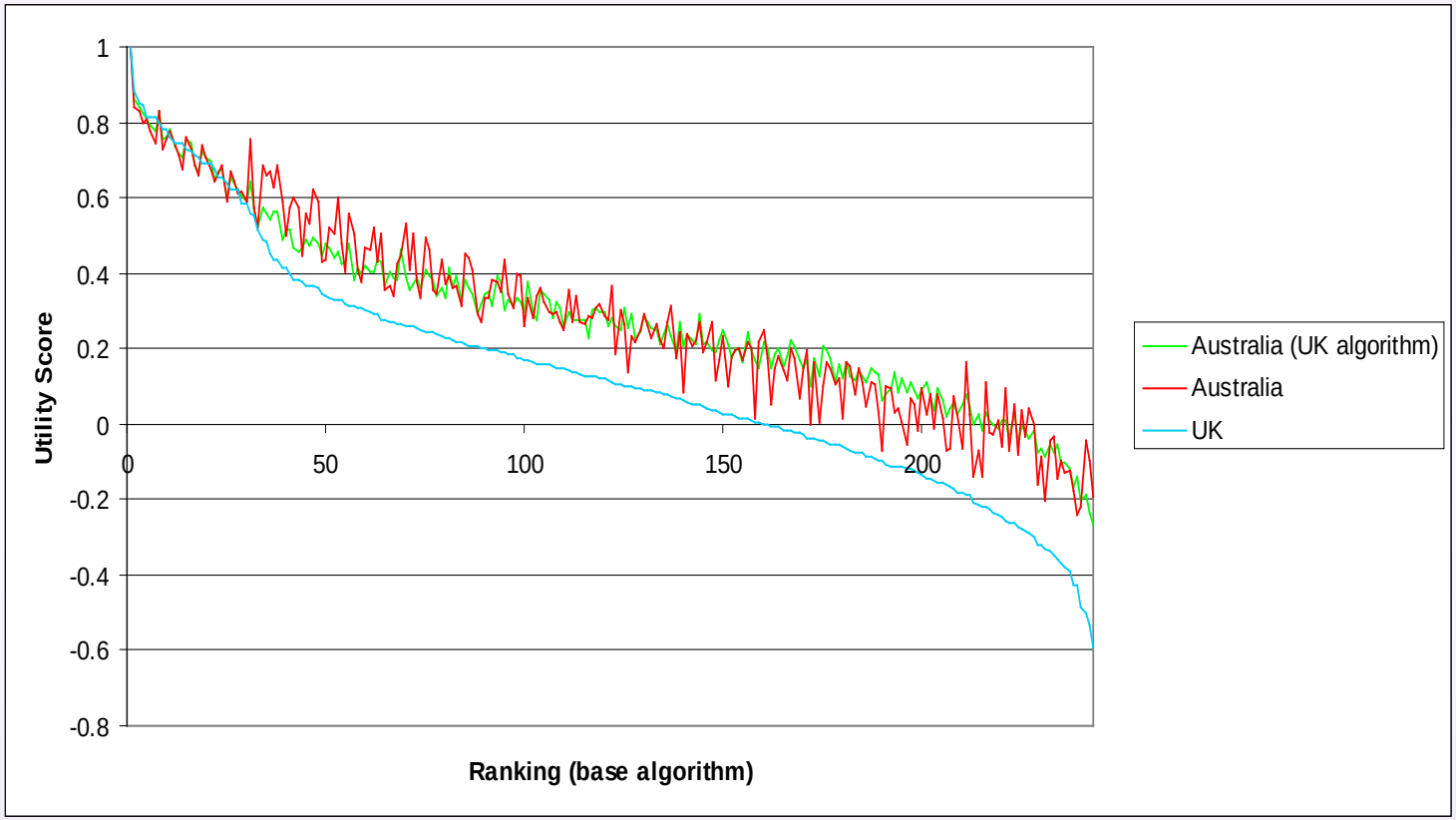
Issues

- Australian Model 2 similar to UK algorithm
- Model 3 preferred to Model 2
- Pattern of interactions more complex than the N3 term suggests
- Effect of an improvement in one dimension depends on how many other dimensions remain at level 3
- Some of these effects are very small
- Some are non-monotonic

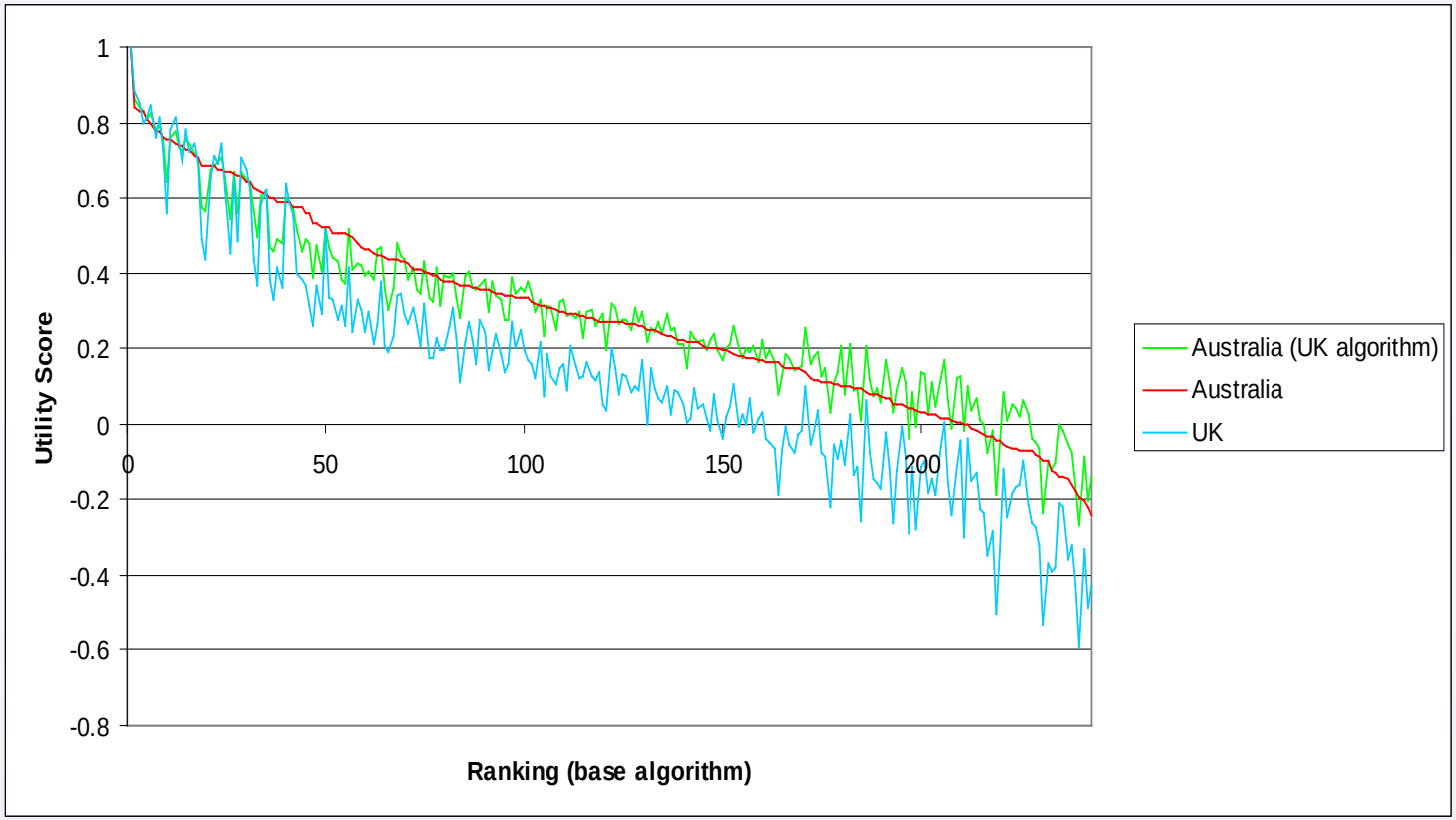
Examples of how algorithms differ

- Gain in utility in moving from UA3 to UA2
  - In Model 2 = 0.273 for all health states with only UA3, and = 0.037 for all other health states (ie with 2, 3, or 4 dimensions at level 3)
  - In Model 3 = 0.086 for all health states with only UA3
    - 0.043 when AD remains at level 3
    - 0.141 when SC remains at level 3
    - 0.111 when Mob remains at level 3
    - 0.018 when AD and PD remain at level 3
    - 0.098 when SC and AD remain at level 3
    - 0.068 when Mob and AD remain at level 3
    - 0.166 when Mob and SC remain at level 3
    - 0.073 when SC, PD and AD remain at level 3
    - 0.043 when Mob, PD and AD remain at level 3
    - 0.141 when Mob, SC and PD remain at level 3
    - 0.098 when all everything else remains at level 3

UK algorithm as base



Our algorithm as base



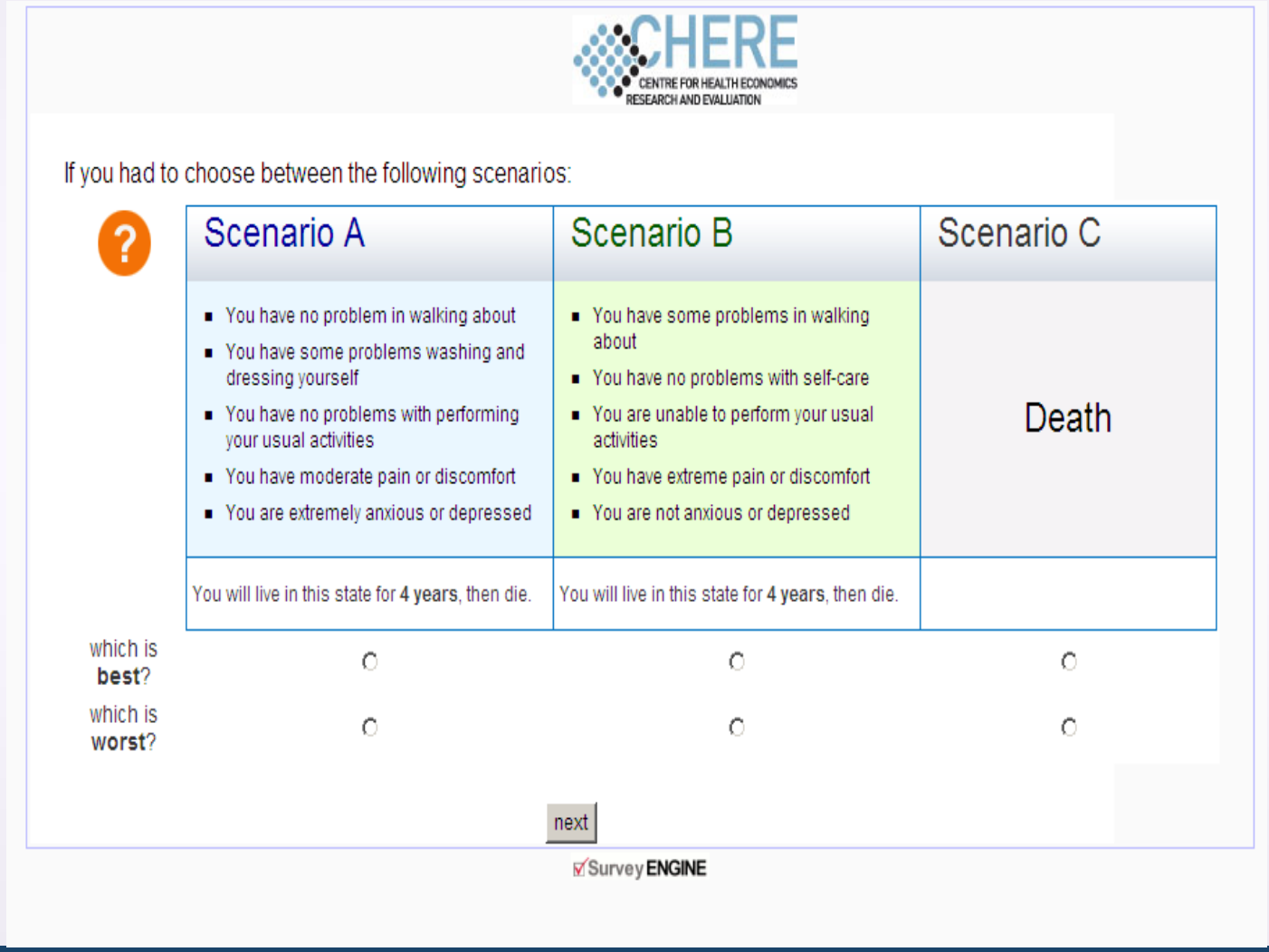
Issue 3: DCE as an approach to health state valuation

- Potential advantages in terms of
  - Ease of administration
  - Readily adaptable to on-line administration
  - Cognitive and interviewer burden
  - Coverage of response surface
  - Potential to explore other attributes (eg impact of duration)
- Potential issues
  - Appropriate design of the DCE task
  - Anchoring at 0 and 1
  - Consistency with the QALY model (or not?)

Design of the DCE

- On-line administration using existing panel + TTO sample for validation
- Experimental design
  - Included all plausible states
  - Allowed 5 levels of duration (1,2,4,8,16)
  - Allowed for 2 fis between EQ-5D levels and all 2 and 3 fis with duration
  - Street and Burgess design
  - 1620 choice sets blocked into 108 blocks of 15 choice sets

Example of a choice set: DCE



Summary results: DCE

		Worst				
		Option A	**Option B**	**Death**	**Total**	
Best	**Option A**		3,065	4,899	**7,968**	**43.7%**
	Option B	2,915		4,653	**7,571**	**41.6%**
	Death	954	989		**1,944**	**10.7%**
	Total	**3,871**	**4,057**	**9,554**	**18,213**	
		21.3%	**22.3%**	**52.5%**		

Deriving a QALY model from DCE results

- DCEs yield ordinal ranking, QALY model is a cardinal ranking
  - Impose assumption that best health state=1 and worst health state=0
  - Use external anchors (eg from a TTO)
  - Use death as a natural anchor
  - Impose the QALY model by estimating duration-health state interactions
    - Linear version of this specification is consistent with QALY restrictions

Preliminary results: Clogit for a QALY model

Variable	Coefficient	SE	Z	P>|z|
Duration (linear)	0.269	0.007	39.660	0.000
mo2_durlin	-0.031	0.004	-7.890	0.000
mo3_durlin	-0.140	0.004	-34.960	0.000
sc2_durlin	-0.032	0.004	-7.320	0.000
sc3_durlin	-0.078	0.004	-17.960	0.000
ua2_durlin	-0.027	0.004	-6.140	0.000
ua3_durlin	-0.051	0.004	-11.850	0.000
pd2_durlin	-0.029	0.004	-7.400	0.000
pd3_durlin	-0.129	0.004	-32.600	0.000
ad2_durlin	-0.037	0.004	-9.610	0.000
ad3_durlin	-0.100	0.004	-25.460	0.000

To value an individual EQ-5D state, this can then be rescaled, so health state 21232 is valued at (0.269-0.031-0.027-0.129-0.037)/0.269=0.167

Heterogeneity - Mixed Logit for QALY model

	Variable	Coefficient	SE	Z	P>|z|
Mean	Duration (linear)	0.553	0.017	33.350	0.000
	mo2_durlin	-0.060	0.007	-8.790	0.000
	mo3_durlin	-0.308	0.010	-30.270	0.000
	sc2_durlin	-0.055	0.008	-6.520	0.000
	sc3_durlin	-0.161	0.009	-17.560	0.000
	ua2_durlin	-0.053	0.009	-6.230	0.000
	ua3_durlin	-0.109	0.009	-12.570	0.000
	pd2_durlin	-0.054	0.007	-7.680	0.000
	pd3_durlin	-0.283	0.010	-29.310	0.000
	ad2_durlin	-0.066	0.007	-9.590	0.000
	ad3_durlin	-0.215	0.009	-23.770	0.000
Standard deviation	Duration (linear)	0.260	0.011	24.230	0.000
	mo2_durlin	0.005	0.015	0.310	0.754
	mo3_durlin	0.186	0.011	16.320	0.000
	sc2_durlin	-0.101	0.012	8.310	0.000
	sc3_durlin	0.135	0.012	11.000	0.000
	ua2_durlin	0.091	0.013	6.760	0.000
	ua3_durlin	0.095	0.015	6.170	0.000
	pd2_durlin	0.004	0.014	0.270	0.789
	pd3_durlin	0.157	0.011	14.430	0.000
	ad2_durlin	0.030	0.012	2.570	0.010
	ad3_durlin	0.165	0.011	15.660	0.000

Issue 4: Duration linear?

- DCE allows for estimation of the effect of different durations
- In preliminary results have estimated models with linear and quadratic terms, and models with dummy for each duration
- Results suggest that duration is non-linear

Conclusions and next steps

- Conventional algorithm for the TTO may not be appropriate
  - More complex pattern of interactions
- QALY assumptions may not hold
  - Duration linear
- DCE approach an option for developing preference index for EQ-5D
  - Best approach to developing algorithm still to be explored
- Heterogeneity in preferences an issue for determining algorithm

Next steps

- Refine the TTO algorithm
  - Deal with non-monotonic results
  - Determine minimum “real” utility change
- Compare alternative approaches to deriving DCE algorithm
- What are the implications of the findings for use of algorithms?