Best **W**orst **S**caling

HEALTH ECONOMICS RESEARCH UNIT

Promoting Excellence in Health Economics

MSc Health Economics
University of Aberdeen

26/02/2015

HERU is supported by the Chief Scientist Office (CSO) of the Scottish Government Health Directorates

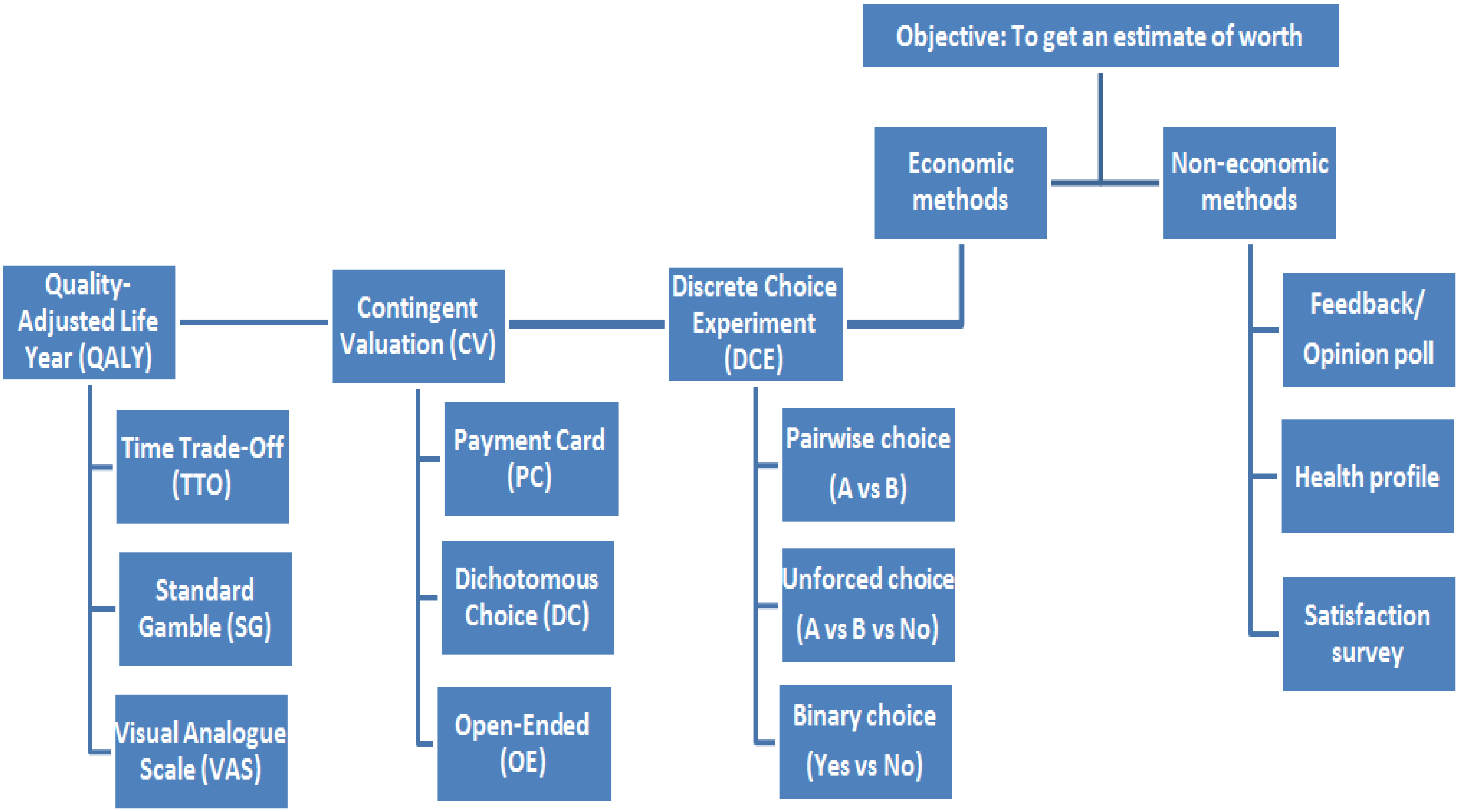
www.abdn.ac.uk/heru

Outline

BWS … Buy one get two free!

- BWS case 1 (or “object” case)
- BWS case 2 (or “profile” case)
- BWS case 3 (or “multi-profile” case)

Remember …



Objective: To get an estimate of worth
Economic methods
Non-economic methods
Quality-Adjusted Life Year (QALY)
Contingent Valuation (CV)
Discrete Choice Experiment (DCE)
Time Trade-Off (TTO)
Standard Gamble (SG)
Visual Analogue Scale (VAS)
Payment Card (PC)
Dichotomous Choice (DC)
Open-Ended (OE)
Pairwise choice (A vs B)
Unforced choice (A vs B vs No)
Binary choice (Yes vs No)
Feedback/ Opinion poll
Health profile
Satisfaction survey

BWS-1

- *An hospital manager wants to efficiently improve the quality of medical services from patients’ perspective.*

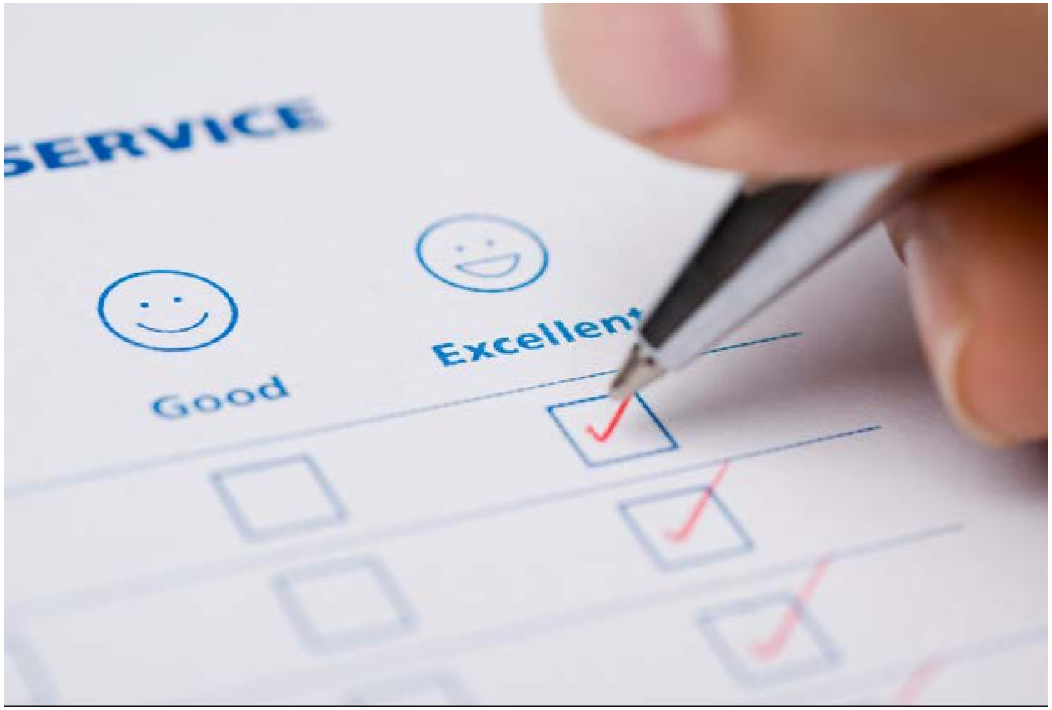
- What would you suggest to do?

“Classical” approach

- **Rating scale procedure** (e.g. satisfaction survey)
  - To identify most important (= priorities) aspects of medical services

Limitations of “classical” approach

- **L1**: Respondent **don’t need to discriminate** among items



Patients' priorities with respect to general practice care: an international comparison (Grol et al, 1999)

TABLE 4 *Description of patients' priority percentages 'very/most important' and rank numbers (n = 3540)*

What would make for a good GP?	Section[b]	Denmark	Germany	Israel	The Netherlands	Norway	Portugal	UK
During the consultation a GP should have enough time to listen, talk and explain to me.	AVAILIBIL	91 *1*	88 *2*	85 *5*	91 *2*	93 *1*	89 *1*	90 *2*
A GP should be able to provide quick service in case of emergencies.	AVAILIBIL	88 *2*	89 *1*	89 *1*	94 *1*	88 *4*	87 *2*	91 *1*
A GP should guarantee the confidentiality of information about all his patients.	RELATION	84 *5*	82 *5*	88 *3*	85 *3*	91 *2*	77 *8*	88 *3*
A GP should tell me all I want to know about my illness.	INFORMAT	85 *4*	84 *3*	89 *2*	82 *5*	76 *9*	69 *14*	84 *5*
A GP should make me feel free to tell him or her my problems.	RELATION	87 *3*	82 *4*	68 *16*	75 *9*	89 *3*	82 *6*	86 *4*
It should be possible to make an appointment with a GP at short notice.	AVAILIBIL	74 *11*	74 *9*	69 *14*	84 *4*	86 *5*	77 *10*	81 *6*

Everything seems to be important !?!

Limitations of "classical" approach

- **L2**: **Not clear how to interpret values** (and their differences)

Question	Mean score
Item 1	4.3
Item 2	4.3
Item 3	4.1
Item 4	3.7
Item 5	3.2
Item 6	3.1
Item 7	3
Item 8	2.9
Item 9	1.5
Item 10	1.4

Are these 4 items really different in terms of importance?

What "1.5" means in terms of importance?

5-points rating scale of importance: Unimportant [**1**]; Of Little Importance [**2**]; Moderately Important [**3**]; Important [**4**]; Very Important [**5**]

“Alternative” approach

- We are looking for a method that:

1) Forces respondents to discriminate among items (or at least prevent them to make “easy” decisions)

2) Provides results on a scale with known measurement properties (>= interval scale)

Any idea ???

BWS-1

- BWS-1 (or object case) is the initial approach (Louviere & Woodworth, 1990)
- **To ask people to select both the "best" (e.g. most important) and "worst" (e.g. least important) items in different subsets of all the items**
- It has been developed as an alternative to rating procedures
- It has been used in 9 health studies
  - Importance of adverse effects associated with tobacco use
  - Importance of factors affecting residency ranking
  - Preferences for preserving functional abilities with Alzheimer disease

BWS case 1

Marti et al (2012)

- 376 respondents
- 15 items
- 16 tasks
- 5 to 11 items per task

Categories	Adverse effects
Long-term consequences	Lung cancer
	Cardiovascular diseases
	Reduced life expectancy
	COPD
	Skin problems
Short-term consequences	Reduction of physical capacity
	Oral and dental problems
	Sexual and fertility problems
	High expenditures
Barriers to smoking cessation	Addiction
	Weight gain after cessation
Product focus	Tobacco industry manipulation
	Inhalation of chemicals
Impact on others	Disturbance of non-smokers
	Endangering relatives

List of items

MOST DETERRENT		LEAST DETERRENT
☐	**Lung cancer** (smokers have a much higher risk of getting lung cancer than non-smokers)	☐
☐	**Addiction** (smoking creates addiction and craving symptoms are uncomfortable (headache, nervousness, etc.))	☐
☐	**Oral/dental problems** (smoking is bad for teeth and gums and causes bad breath)	☐
☐	**Weight gain after cessation** (individuals who quit gain 2-3 kilos on average)	☐
☐	**Endangering relatives** (second-hand smoke is harmful to other's health, including relatives)	☐

Designing a BWS-1 experiment

- Requirement: All the items cannot fit in any one 'block' (i.e. task)
- We need to use a **Balanced Incomplete Block Design** (BIBD)
- BIBDs have 3 desirable properties:

  1) Each task includes same number of items

  2) [1st-order balance] **Each item occurs the same number of times** across tasks

$$r = \frac{Tasks \times Size}{Items}$$

Designing a BWS-1 experiment

3) [2nd-order balance] **Each item appears equally often with every other item**

$$\lambda = r\frac{(Size - 1)}{(Items - 1)}$$

- **BIBD doesn't exist for every combinations** of {Nbr Items; Nbr tasks; Size tasks}. There is a catalogue!
- Necessary but not sufficient conditions: **(r) and (λ) are integers**

Designing a BWS-1 experiment

- 3 methodological choices to do:
  - Define the **number of items**
  - Define the **number of items per task**
  - Define the **amount of information** needed (either Tasks or λ)

- More observations increase the reliability of results because decrease influence of random errors, but:
  - Too much tasks might be unmanageable (for the user and/or participant)
  - Diminishing marginal return

Designing a BWS-1 experiment

- Full (factorial) design:

$$Tasks = \frac{Items \times (Items - 1) \times \ \ldots \times (Items - Size + 1)}{Size!}$$

- Example: Items=8; Size=4 ***Manually defined by user***
  - Full design => Tasks=1680/24=70
  - **BIBD => 4 solutions**

Task	r	λ
14	7	3
28	14	6
42	21	9
56	28	12
70	35	15

Analysing a BWS-1 experiment

- BWS data can be analysed as:
  - **Count** data (i.e. nbr of times each item is selected as best/worst)
  - **Choice** data (i.e. which item is selected as best/worst)

1/ Count-based approach

- Several scores can be computed with different (measurement) properties
  - **Difference score** (= B-W) ➔ Location of items on **interval scale**
  - **Ratio score** $\left(= \sqrt{B/W}\right)$ ➔ Location of items on **(pseudo)-ratio scale**

- Empirical finding: LN(ratio score) => Values similar to the ones obtained with choices-based approaches (e.g. conditional logit model)

- If enough observations, can be computed for each respondent separately and then used in segmentation analyses (e.g. cluster analysis)

Analysing a BWS-1 experiment

2/ Choice-based approach

- Probability of choosing a particular item as best (or worst) in a task (subset of items)
- Different specifications are possible (example: choice among 3 items {A,B,C})
  - **Maximum Difference** (MaxDiff)

6 possible BW pairs

$$P(pair = AB) = \frac{\exp(V_{AB})}{\exp(V_{AB}) + \exp(V_{AC}) + \exp(V_{BC}) + \exp(V_{BA}) + \exp(V_{CA}) + \exp(V_{CB})}$$

  - **Simultaneous** (SIM)

$$\left[P(best = A) = \frac{\exp(V_A)}{\exp(V_A) + \exp(V_B) + \exp(V_C)}\right] \times \left[P(worst = B) = -\frac{\exp(V_B)}{\exp(V_A) + \exp(V_B) + \exp(V_C)}\right]$$

Analysing a BWS-1 experiment

– **Sequential** (SEQ)

$$\left[P(best = A) = \frac{\exp(V_A)}{\exp(V_A)+\exp(V_B)+\exp(V_C)}\right] \times \left[P(worst = B) = -\frac{\exp(V_B)}{\exp(V_B)+\exp(V_C)}\right]$$

- Empirical finding: All these specifications lead to similar results (***To be confirmed!***)
- All of these models can be accommodated to include individual covariates (e.g. age, gender)

Example of BWS-1 experiment

“Six Nations Championship”

We want to know how good are the different teams

- **Step 1**: Define the content of the experiment
  - 6 items {England; France; Ireland; Italy; Scotland; Wales} + 3 items per task
- **Step 2**: Identify the number of tasks
  - Grid search => 2 solutions (i.e. 10 and 20 tasks)

Task	Freq.	Co-freq.
1	0.5	0.2
2	1.0	0.4
3	1.5	0.6
4	2.0	0.8
5	2.5	1.0
6	3.0	1.2
7	3.5	1.4
8	4.0	1.6
9	4.5	1.8
10	5.0	2.0
11	5.5	2.2
12	6.0	2.4
13	6.5	2.6
14	7.0	2.8
15	7.5	3.0
16	8.0	3.2
17	8.5	3.4
18	9.0	3.6
19	9.5	3.8
20	10.0	4.0

Max number of tasks (cf full design)

Example of BWS-1 experiment

- **Step 3**: Generate the design + Check its properties

DESIGN *	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7	Task 8	Task 9	Task 10
England	No	Yes	Yes	Yes	No	No	No	Yes	No	Yes
France	Yes	No	No	Yes	Yes	No	Yes	Yes	No	No
Ireland	No	No	Yes	No	Yes	Yes	No	Yes	Yes	No
Italy	No	Yes	No	Yes	No	Yes	Yes	No	Yes	No
Scotland	Yes	No	Yes	No	No	Yes	Yes	No	No	Yes
Wales	Yes	Yes	No	No	Yes	No	No	No	Yes	Yes

* Generated with R package 'crossdes' ==> **find.BIB(item, task, size)**

DESIGN *	England	France	Ireland	Italy	Scotland	Wales
England	**5**					
France	2	**5**				
Ireland	2	2	**5**			
Italy	2	2	2	**5**		
Scotland	2	2	2	2	**5**	
Wales	2	2	2	2	2	**5**

* R package 'crossdes' ==> **isGYD(design)**

Example of BWS-1 experiment

- **Step 4**: Create the choice tasks + Collect data

Best	Task 1	Worst
X	France	
	Scotland	X
	Wales	

(… etc …)

Best	Task 10	Worst
	England	
	Scotland	X
X	Wales	

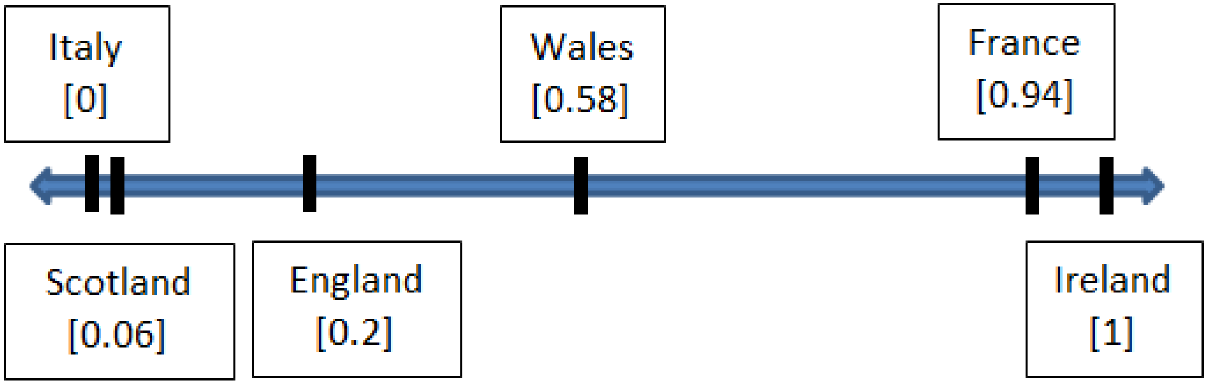
Task	Best	Worst
1	FRA	SCO
2	WAL	ITA
3	IRE	SCO
4	FRA	ITA
5	IRE	WAL
6	IRE	ITA
7	FRA	ITA
8	IRE	ENG
9	IRE	ITA
10	WAL	SCO

Example of BWS-1 experiment

Step 5: Analyse the data (e.g. count approach)

CHOICES	Best (B)	Worst (W)	B-W	Rank	ln(√(B/W)) *	Rescaled
England	0	1	-1	4	-1.20	0.20
France	3	0	3	2	1.72	0.94
Ireland	5	0	5	1	1.97	1.00
Italy	0	5	-5	6	-1.97	0.00
Scotland	0	3	-3	5	-1.72	0.06
Wales	2	1	1	3	0.32	0.58

* Based on (B+0.1) and (W+0.1) columns to avoid computational issues



BWS case 1

Limitations

- **Relative anchoring/scaling**: We know how important the items are relative to each other, but we still do not know if their absolute level of importance.

- **Not flexible**: Impossible to add new items a posteriori.

- **No ties**: How can I make my choice(s) if I consider 2 items as being equally important? Random choice(s)? Which consequences for the items ranking if few observations?

- **Unrealistic**: Under some circumstances, it might be difficult to identify the Best or the Worst items. For example, what would be the Best disease among {Leukemia; Lung cancer; Pancreatic cancer}?

Other "advantages" of BWS-1

- **Less prone to cultural biases**

Response Style and Cross-Cultural Comparisons of Rating Scales Among East Asian and North American Students (Chen et al, 1995)

Abstract

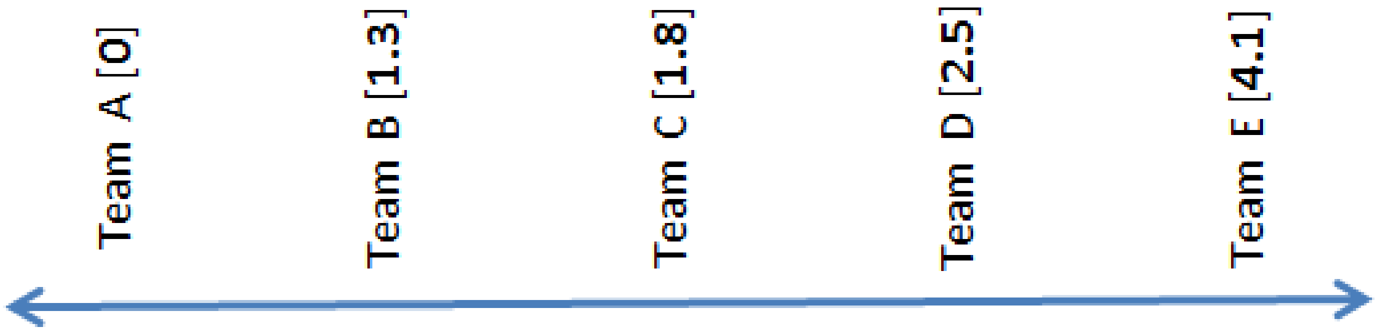
This report examines cross-cultural differences in response style regarding the use of rating scales Subjects were high school students 944 from Sendai (Japan), 1,357 from Taipei (Taiwan), 687 from Edmonton and Calgary (Canada), and 2,174 from the Minneapolis metropolitan area and Fairfax County, Virginia Responses to fifty-seven 7-point Likert-type scales were analysed. **The Japanese and Chinese students were more likely than the two North American groups to use the midpoint on the scales, the US subjects were more likely than the other three groups to use the extreme values**. Within each cultural group, endorsement of individualism was positively related to the use of extreme values and negatively related to the use of the midpoint. These small, albeit statistically significant, differences in response styles generally did not alter cross-cultural comparisons of item means

Limitations of the BWS-1

How to interpret results?

- There is **no ‘benchmark’ value**. Only relative measurements.

Limitations of the BWS-1



- We know that "Team E > Team D > Team C > Team B > Team A" …. But it could be the case that all of these teams don't perform well !

Think about BWS study on dictators. Does a positive measure means that you like them?

Solution

- **Anchored BWS**
  - Add a new question (e.g. Lagerkvist et al, 2014)

Considering only these five features, which is the Most Important and which is the Least Important?		
Least Important	Feature	**Most** Important
	I can use Peepoo as fertiliser when I grow vegetables	
	The use of Peepoo reduces the smells in my house/garden	
	Peepoo makes going to the toilet cleaner	
	It is easy for me to dispose of the bag to a collector	
	Peepoo is easy to use	

Considering just these five features…

- ☐ **None of these five are important**
- ☐ **Some are important, some are not**
- ☐ **All five are important**

Figure 1 | Example of ABWS scaling choice set.

Limitations of the BWS-1

What if one wants to measure a new item?

- There is no correction procedure so far. You will have to run a new BWS study.

Limitations of the BWS-1

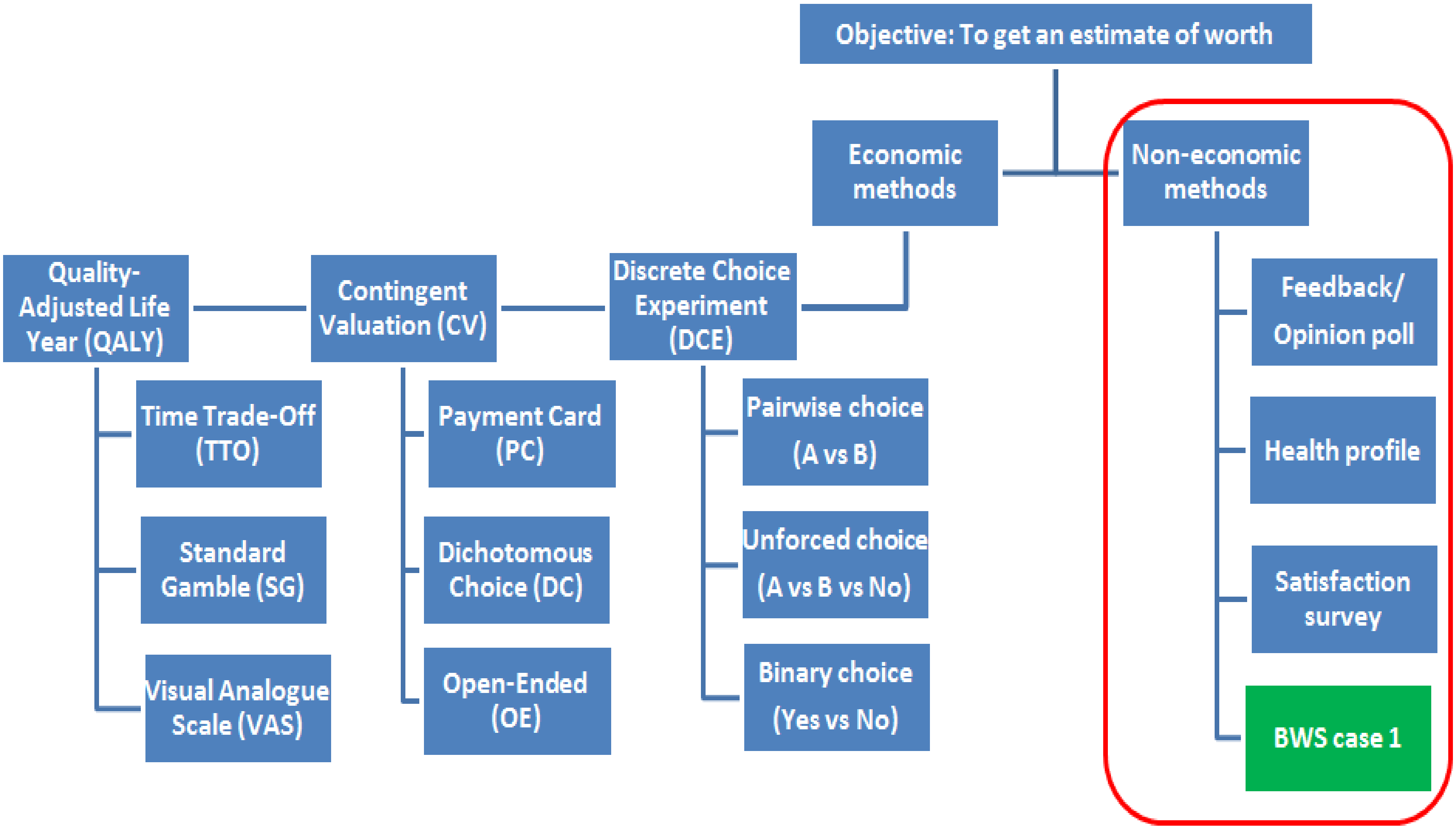
How can I make my choice(s) if I consider 2 items as being equally important?

- By definition BWS does not allow for “indifferent” choices
- Which consequences? Random choice(s)? What if ranking of items is based on few observations?

Sometimes BWS can be “behaviourally awkward”

- What would be the <u>MOST DESIRABLE</u> disease among {Leukemia; Lung cancer; Pancreatic cancer}?

Conclusion for BWS-1



BWS-2

- BWS case 2 (or profile case) has been introduced to health economics few years ago (Flynn et al, 2007)

- **To ask people to select both the "best" (e.g. most important) and "worst" (e.g. least important) characteristics of a product/service**

- Developed mainly as an alternative to DCE/SG/TTO procedures

- Used in 12 health studies
  - Preferences for contraceptive methods
  - Preferences for genetic tests
  - Health utility index (e.g. CHU-9D, ICECAP, ICECAP-A)

BWS-2

Al-Janabi et al (2012)

- 397 respondents
- 6 attributes (6x3L)
- 18 tasks

Imagine being in situation 1 below
What would be the best thing and the worst thing about it?

Best thing		*Worst thing*
	You can do some of the other things you want to do outside caring	
	You get little support from family and friends	
	You get a lot of assistance from organizations and the government	
	You mostly find caring fulfilling	
	You are in control of some aspects of the caring	
	You sometimes get on with the person you care for	

Table 1 Attributes and Levels of the Carer Experience Scale

Attribute

Activities outside caring

Levels

- You can do *most* of the other things you want to do outside caring
- You can do *some* of the other things you want to do outside caring
- You can do *few* of the other things you want to do outside caring

Support from family and friends
- You get *a lot of* support from family and friends
- You get *some* support from family and friends
- You get *little* support from family and friends

Assistance from organizations and the government
- You get *a lot of* assistance from organizations and the government
- You get *some* assistance from organizations and the government
- You get *little* assistance from organizations and the government

Fulfillment from caring
- You *mostly* find caring fulfilling
- You *sometimes* find caring fulfilling
- You *rarely* find caring fulfilling

Control over the caring
- You are in control of *most* aspects of the caring
- You are in control of *some* aspects of the caring
- You are in control of *few* aspects of the caring

Getting on with the person you care for
- You *mostly* get on with the person you care for
- You *sometimes* get on with the person you care for
- You *rarely* get on with the person you care for

BWS-2

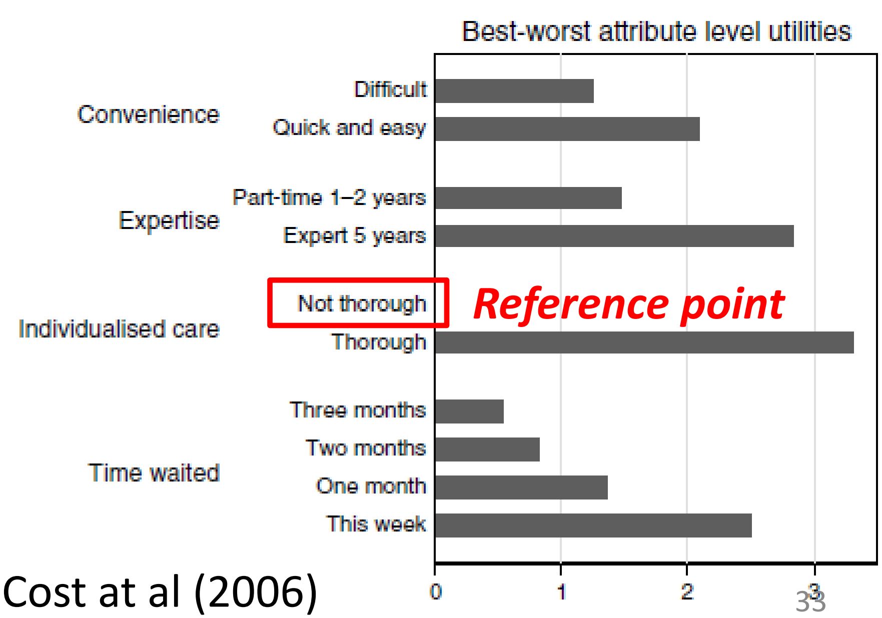
- ***Does it look familiar?***
- ***Why using a BWS-2?***

Table 3 Utility values obtained for the different levels of the different attributes compared with the lowest value (not thorough) which has been rescaled to zero

Attribute	Utility	95% confidence interval	P-value
Time waited			
3 months	0·55	0·28–0·81	< 0·001
2 months	0·83	0·54–1·12	< 0·001
1 month	1·38	1·06–1·70	< 0·001
This week	2·52	2·21–2·82	< 0·001
Expertise			
Part-time specialist	1·48	1·18–1·79	< 0·001
Team led by expert	2·84	2·52–3·16	< 0·001
Convenience			
Difficult	1·26	0·94–1·58	< 0·001
Easy	2·10	1·76–2·44	< 0·001
Thorough care			
Thorough	3·33	2·95–3·70	< 0·001
Not thorough	0		

Table 2 Importance weights for the attributes as a whole

Attribute	Weight
Time waited	1·32
Expertise	2·16
Convenience	1·68
Thorough care	1·66



Cost at al (2006)

BWS-2

- Some comparative advantages of BWS-2:
  - BWS tasks would be **easier to complete** than DCE tasks
  - Provide **more (statistical) information** on respondents' choices
  - Estimate **influence of more attributes' levels**
  - Locate all the attributes' levels on a **common scale**
  - Make a distinction between influence of attributes and levels (i.e. **separate weights and scale values**)

Designing a BWS-2 experiment

- BWS-2 shows the alternatives one at time
- We can use an **Orthogonal Main Effects Plan** (OMEP)
- The OMEP has 3 desirable properties:
  - **Level balance**: For a given attribute, each level appears same number of times
  - **Orthogonality**: The attribute's levels occur independently of each others
  - **Constraint**: Each attribute is represented once per task

- OMEP:
  - Does not exist for any number of tasks
  - Several solutions can be possible (with different number of tasks)

- Necessary (but not sufficient) conditions: Nbr of tasks is a multiple of the number of attributes' levels (e.g. 15 tasks is not a multiple of 4 levels)

Designing a BWS-2 experiment

- The user has 3 methodological choices to do:
  - Define the **number of attributes**
  - Define the **number of levels per attribute**
  - Define the **amount of information** needed (Lowest/Highest number of tasks)

- More observations:
  - Increase the accuracy of estimated preferences (i.e. lower SEs)
  - Might lead to 'artificial' preferences because of some psychological effects (e.g. fatigue, simplifying heuristics)

Always a good to find a compromise between

'not enough' and 'too much'

Designing a BWS-2 experiment

- Full (factorial) design, with (k) attributes:

$$Tasks = \prod_{k} L_k$$

- <u>Example</u>: 4 attributes, each having 3 levels
  - Full design => 3x3x3x3 = 81 tasks

- OMEP can be find with Software or online catalogue (Sloan, 2005)

- Catalogue → 9 solutions {9; 18; …; 81}

Analysing a BWS-2 experiment

- Best using econometric models considering Best and Worst decisions as choices among subsets of attributes' levels

- Again, 3 conceptualisations can be used:
  - MaxDiff, SEQ, SIM

- MaxDiff is based on the assumption that respondents try to select the items furthest apart on the latent scale (i.e. maximise 'utility' distance)

- MaxDiff is compatible with Random Utility Theory (RUT): Utility gained from alternatives cannot be fully recovered => We model probability of choices

$$P(pair = AB) = \frac{exp(\beta_A Best_A - \beta_B Worst_B)}{\sum_{i,j} exp\left(\beta_i Best_i - \beta_j Worst_j\right)}$$

Example of BWS-2 experiment

- **Step 1**: Define the content of the experiment
  - 3 attributes {Quantity; Quality; Cost} with 2 levels {Low; High}

- **Step 2**: Identify the number of tasks
  - Full design = 2x2x2 = 8 alternatives
  - OMEP (1 solution) = 4 alternatives

DESIGN *	Quality	Quantity	Cost
Alt 1	High	Low	High
Alt 2	High	High	Low
Alt 3	Low	High	High
Alt 4	Low	Low	Low

* R package 'DoE.base' => **oa.design(2,2,2)**

- **Step 3**: Check the design properties

FREQUENCY	Low	High
Quality	2	2
Quantity	2	2
Cost	2	2

CORRELATION	Quality	Quantity	Cost
Quality	1	0	0
Quantity	0	1	0
Cost	0	0	1

Example of BWS-2 experiment

- **Step 4**: Create the tasks + Collect the data

Task 1	Best	Worst
High quality	X	
Low quantity		
High cost		X

Task 2	Best	Worst
High quality		
High quantity		X
Low cost	X	

Task 3	Best	Worst
Low quality		X
High quantity	X	
High cost		

Task 4	Best	Worst
Low quality		X
Low quantity		
Low cost	X	

MaxDiff structure

ID	TASK	BW PAIR	CHX
1	1	Qual-Quant	0
1	1	Qual-Cost	1
1	1	Quant-Cost	0
1	1	Quant-Qual	0
1	1	Cost-Qual	0
1	1	Cost-Quant	0
(etc.)			

Example of BWS-2 experiment

- **Step 5**: Analyse the data (e.g. conditional logit model)

Table 3 Utility values obtained for the different levels of the different attributes compared with the lowest value (not thorough) which has been rescaled to zero

Attribute	Utility	95% confidence interval	P-value
Time waited			
3 months	0·55	0·28–0·81	< 0·001
2 months	0·83	0·54–1·12	< 0·001
1 month	1·38	1·06–1·70	< 0·001
This week	2·52	2·21–2·82	< 0·001
Expertise			
Part-time specialist	1·48	1·18–1·79	< 0·001
Team led by expert	2·84	2·52–3·16	< 0·001
Convenience			
Difficult	1·26	0·94–1·58	< 0·001
Easy	2·10	1·76–2·44	< 0·001
Thorough care			
Thorough	3·33	2·95–3·70	< 0·001
Not thorough	0		

Coast et al (2006)

Limitations of BWS-2

- **Conditional demand**: BW decisions are analysed as economic choices but we don't know if respondents are 'demanders' of the attributes' levels (➔ no MRS, no CV, no forecast)
- **Cognitive difficulties** … Not easier than DCE tasks
- **Misleading**: The 'Worst' decisions provide additional information to estimate preferences, but decision making processes can differ between Worst and Best decisions
- **Restrictive**: How taking into SQ effect? Label/Naming effect?
- **Unrealistic behavioural assumptions** of the choice model (e.g. MaxDiff)

BWS-2 results seem to be significantly different from DCE results.

Which method should be used? For what purpose?

Solution

- To combine binary-choice DCE + BWS-2

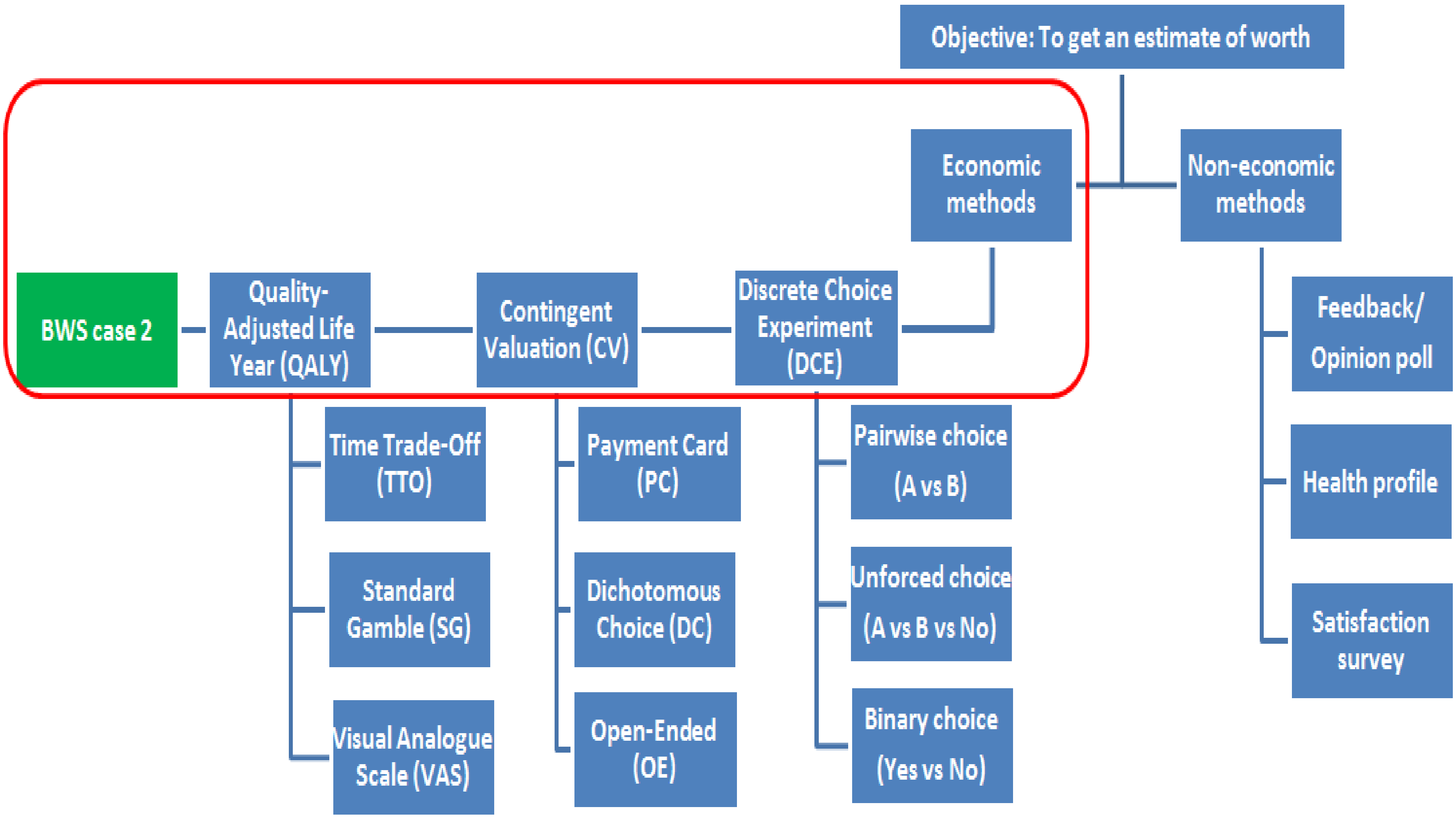
Best thing	The appointment with the specialist	Worst thing
	You will have to wait *one* month for your appointment	
	Getting to your appointment will be difficult and time-consuming	
	The consultation will be as thorough as you would like	
	The specialist is in a team led by an expert who has been treating skin complaints full-time for at least 5 years	

BWS part

Would you attend this appointment? Yes ☐ 1 No ☐ 2

DCE part

Conclusion for BWS-2



BWS-3

- BWS-3 (or multi-profile case) is the most recent approach (Marley & Philens, 2012)

- **To ask people to select both the “best” (e.g. most important) and “worst” (e.g. least important) product/service among subsets of at least 3 products**

- It has been developed as:
  - Extended-version of DCE
  - Alternative to classical ranking procedures

- It has been used in 4 health studies
  - Preferences for nursing jobs
  - Preferences for EQ5D health states
  - Preferences for management of cardiac arrest

BWS-3

Doiron et al (2011): 526 respondents; 12 attributes; 8 tasks

Features of Job	Job A	Job B	Job C
Location	Private hospital	Private hospital	Public hospital
Clinical rotations	None	None	Three
Work hours	Fulltime only	Part-time or fulltime	Fulltime only
Rostering	Inflexible, does not allow requests	Flexible, usually accommodating requests	Inflexible, does not allow requests
Staffing levels	Usually well-staffed	Frequently short of staff	Usually well-staffed
Workplace culture	Unsupportive management and staff	Unsupportive management and staff	Supportive management and staff
Physical environment	Poorly equipped and maintained facility	Poorly equipped and maintained facility	Well equipped and maintained facility
Professional development and progression	No encouragement for nurses	Nurses encouraged	No encouragement for nurses
Parking	Abundant and safe	Limited	Abundant and safe
Responsibility	Appropriate responsibility	Appropriate responsibility	Too much responsibility
Quality of care	Poor	Excellent	Excellent
Weekly salary	$1,250	$800	$1,100
Considering these three jobs:			
Which would you MOST like to get?	○ Job A	○ Job B	○ Job C
Which would you LEAST like to get?	○ Job A	○ Job B	○ Job C

BWS-3

- ***Why using BWS-3 instead of classical (pairwise choice) DCE?***

- Comparative advantages of BWS-3:
  - Provide **richer information** than methods based on ‘pick one’ approach (Less than 30% of initial sample size is needed)
  - Take advantage of human propensity to **best identify ‘extremes’ objects** (Best-Worst easier than ranking)
  - Can be used for **individual-level analyses** (i.e. estimate preferences of each respondent separately)

Designing a BWS-3 experiment

- Same approaches than for DCE with 2 main possibilities:
  - **Orthogonal** design (OD) [cf BWS case 2]
  - **Efficient** design (ED)
- Better to use ED but more difficult to design (need specialised software)
  - Minimise variance of parameter estimates
  - Variance measured by determinant of covariance matrix, which is related to the (Fisher) information matrix
  - Finally, the design optimality is measured by the D-Error (we want the lowest value)

$$D\text{-error} = \det(I_{BW}(X, \beta))^{-\frac{1}{q}}$$

- **Issue**: How to select the prior information for the parameters? (Ongoing research)

Designing a BWS-3 experiment

- Many methodological choices to do:
  - Define the **number of attributes and levels**
  - Define the **number of alternatives per task**
  - [If more than 3 alt] Define the **number of BW rounds**
  - Define the **format of BW decisions** (e.g. 1st best – 1^{st} worst – 2^{nd} best – 2^{nd} worst, etc)

- Different choice processes can lead to different results. One should consider the cognitive facilitation. Best is probably 'free choice process' but requires computer-based survey to record order of decisions.

Designing a BWS-3 experiment

- Full (factorial) design, with (k) attributes and (m) alternatives per task:

$$alt = \prod_k L_k$$ [*cf BWS case 2*]

$$task = \frac{alt \times (alt-1) \times \cdots \times (alt-m+1)}{m!}$$ [*cf BWS case 1*]

- <u>Example</u>: 4 attributes, each having 3 levels
  - Full design => 9x8x7/3x2 = 84 tasks
  - ED (many solutions) => 9 tasks (D-Error = 0.0167)

Analysinga BWS-3 experiment

- Slightly different than BWS case 1 and 2
- Key notion: Order/Sequence of choices (**A** $\succ$ B $\succ$ **C**)
  - **Rank-Ordered (RO) model**

$$P(AB) = P_{\{ABC\}}(Best = A) \times P_{\{BC\}}(\boldsymbol{Best = B})$$

  - **SIM**

$$P(AB) = P_{\{ABC\}}(Best = A) \times P_{\{\boldsymbol{ABC}\}}(Worst = C)$$

  - **SEQ**

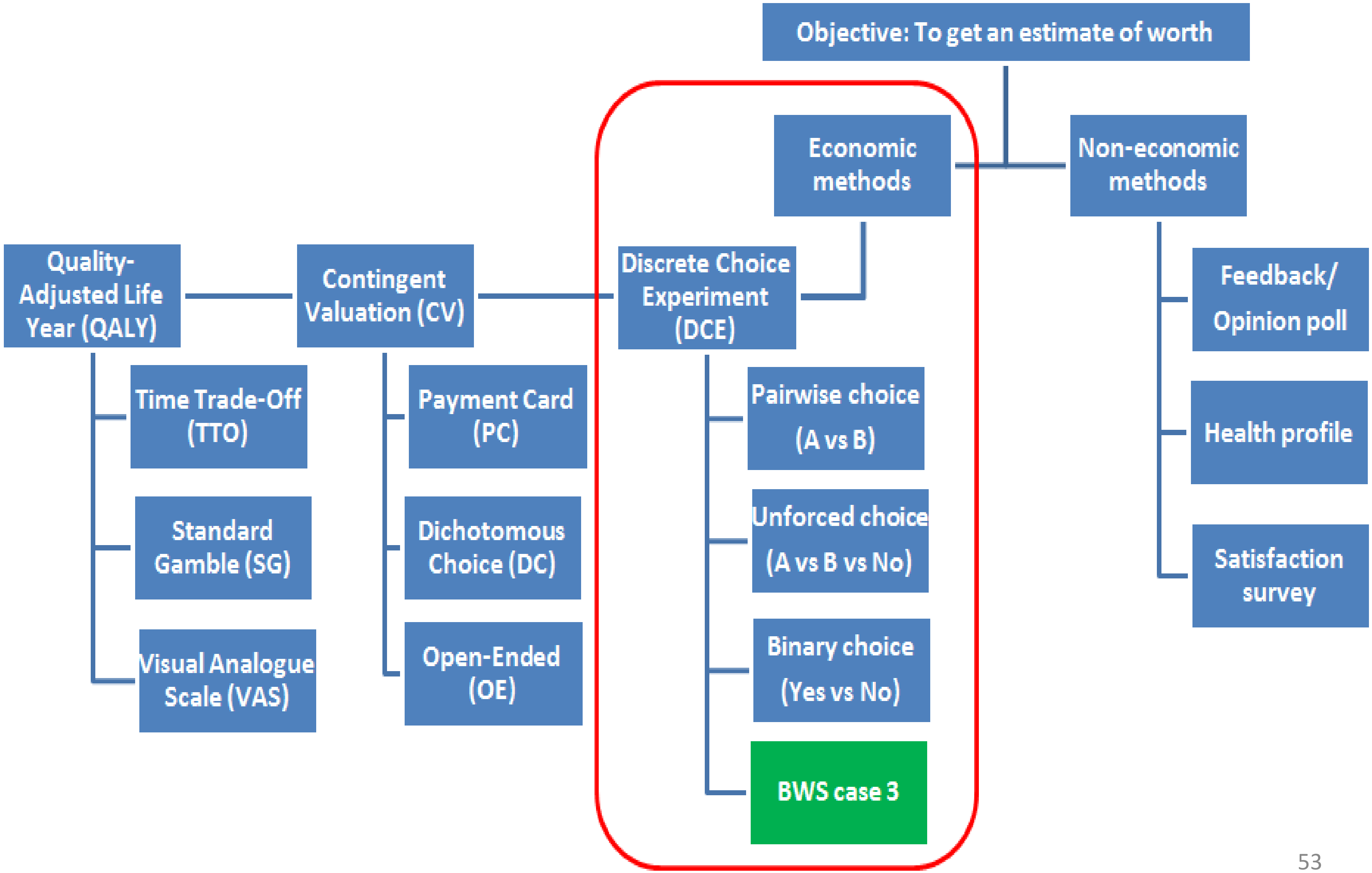
$$P(AB) = P_{\{ABC\}}(Best = A) \times P_{\{BC\}}(Worst = C)$$

- Remark:
  - Best using a model in line with the 'true' choice process underlying BW decisions, but it is not necessary.
  - As before, these models are compatible with RUT

Limitations of BWS-3

- **Errors variability**: Different choice contexts are associated with different error scales and/or different determinants of the choices
- **Cognitive difficulties**: People might have difficulties processing many alternatives at the same time (Increase simplifying heuristics? Attributes non-attendance?)
- Difficult to find the 'right' **choice process**: Maybe asking for {1st best; 2nd best} would be better than a 1-round or 2-rounds Best-Worst format
- **Sophisticated modelling**: Given the (high) influence of errors, it is necessary to analyse the data with at least Heteroskedastic Logit models. What about heterogeneity in preferences? Far from being straightforward …

Conclusion for BWS-3



Application: BWS-1

- Identify a topic and describe it with items/attributes
- Use the ‘design guide’ to create your own choice experiment
- Answer the BWS tasks
- Analyse the results
- What are your thoughts?