Eliciting Patient & Public Preferences in Health Care: The Use of Stated Preference Approaches

Terry Flynn PhD

Senior Research Fellow,
Head of Social Policy and Economic Evaluation
The Centre for the Study of Choice (CenSoC)
Public demands improved transport

ANDREW WEST
February 12, 2010

ALMOST two thirds of Sydneysiders support "high investment in public transport" - and they are prepared to pay for it.

One of the most extensive surveys of commuter needs finds the public is desperate for improved train, bus, light rail and ferry services - and even drivers, who rarely use mass transit, strongly favour better public transport.

Have your say

The survey of 2400 randomly selected people - conducted by the Centre for the Study of Choice at the University of Technology, Sydney - reflected an overwhelming preference for public transport solutions to the city's congestion crisis.

A subgroup of 1200 people, who were asked about long-term solutions, said they were prepared to pay a mixture of slightly higher fares, limited congestion charges and even annual household levies, which would raise almost $36 billion over the next 25-30 years to improve public transport. But the public was strongly against paying more for the current levels of service, which they consider inadequate.

The results of the research come as the Herald-backed independent public inquiry into Sydney's transport needs - headed by Ron Christie, a former chief of the State Rail Authority and the Roads and Traffic Authority - prepares to...
Stated preference approaches – why?

• Revealed preference data not available
  – Market mechanism not used
  – Market not yet exist
  – Possibility of refining design of good or service – need to inform product development

• Disentangle effects of particular features
  – Avoid multicollinearity by design

• Patient reported outcomes
  – Need to value different outcomes

• Individual patient preferences
Why discrete choice experiments (DCEs)?

- Organizations must estimate demand for new products with new attributes/features - no prior RP data.
- Little variability in explanatory variables in real markets.
- Highly correlated explanatory variables - ill-conditioned RP data
- New variables explain choices - new features/variants.
- RP data violate assumptions &/or contain statistical nasties.
- Time consuming/expensive to collect RP data.
- Product(s) not traded in real market.
Features of preference data sources:

- **RP data**
  - Depict world as it is now (current market equilibrium),
  - Capture inherent attribute relations (technology constraints fixed),
  - Have only existing alternatives as observables,
  - Embody market & personal constraints of decision-makers,
  - Have high reliability & face validity,
  - Yield one observation per respondent at each observation point.
Features of data sources, 2

- DCE or “stated preference” (SP) data
  - Hypothetical or virtual decision contexts (flexibility),
  - Control attribute relationships - can map utility for new technology levels into existing levels,
  - Include existing &/or proposed &/or generic options,
  - Not easy (maybe impossible) to estimate changes in markets & personal constraints ("Now, assume you’re female" ...),
  - Reliable if folks understand, are committed to & can complete tasks,
  - (Usually) give several observations/person/observation.
Discrete Choice Experiment

[Image of various soda cans with checkboxes]
How often I choose Coca-Cola over Pepsi provides an estimate of how much I value Coca-Cola over Pepsi

Choice frequencies indicate value
A second bit of theory….

If we manipulate the features (attributes) of the goods/services in a systematic way we can allocate total “value” to each of the possible features

How sensitive are you to:

- Brand?
- Sugar/diet?
- Flavour?
Describing a good or service by its attributes

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flavour</td>
<td>Orange</td>
<td>Lemon</td>
<td>Cola</td>
</tr>
<tr>
<td>Diet</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Brand</td>
<td>Tango</td>
<td>7up</td>
<td>Pepsi</td>
</tr>
</tbody>
</table>
History of stated preferences

• Early work (Thurstone)

• Theory of conjoint measurement
  – Debreu, Luce, Tukey
  – No proper theory of the errors people observed to make
  – It is a theory about the STATISTICAL manipulation of numbers and their decomposition into “part-worths”, NOT a theory of how humans make choices

• Need for an error theory
  – Work in mathematical psychology
  – Luce & Marley (& others)
  – McFadden (1974)
• Louviere & Woodworth (1983) proposed DCEs
  – Integrated experimental design with discrete multivariate statistical methods for contingency tables
    • Choice data consistent with McFadden’s MNL model.
  – Prior work used quasi-designs (all pairs, ad hoc pairs, triples, quads, etc) for multiple choices (esp in psych & stats; eg, Kendall & Smith, 1940; Bradley & Terry, 1952; Wells, 1991).
    • L & W gave systematic ways to create choice sets - resulting choices are consistent with statistical choice models; designs simulate real markets.
    • DCEs called “just another form of conjoint analysis” (incorrect), so the term “choice-based conjoint analysis” describes what we now call “DCEs”.
      – Responsible for much confusion to this day.
• Little progress on DCE design until mid-90s.

• Design efficiency work appears mid-90s.
  – Highly-cited paper (Bunch et al. 1996) motivates “shifted designs” - later shown to be optimal (Street & Burgess, 2007).
A view of real markets

Awareness → Interest → Capability → Choose now? → Choose later? → Choose never?

Choice-Based Conjoint & DCEs

Unintegrated questions:
1. Volumes
2. Interpurchase times
3. Pre-choice expectations
4. Updating expectations

Per week:
- A
- B
- ...
- J

As expected?

update
Knowledge required for DCEs

• Judgment & decision making (psychology, marketing)
• Discrete choice models (economics, psychology, marketing)
• Experimental design (statistics, economics, marketing)
• Discrete multivariate & Bayesian statistics

• Cross-disciplinary TEAMS needed for DCEs.
  – eg, CenSoC includes academics in economics, engineering, finance, marketing, physics, psychology, statistics, transport & IT.
  – Problems too complex & important to leave to one field.
Discrete choice experiments are NOT conjoint analysis

• Just because people with industry links in North America use the CA term is not a good justification
• DCEs are backed by a THEORY of human decision-making
• CA has NO SUCH THEORY
• CA includes a bunch of atheoretical techniques/tasks
  – What theory explains what 6 out of 7 on a rating scale MEANS?
  – What real life behaviour does 6 out of 7 explain?
• World experts in several fields use terms DCE/DCM
  – Transport (Louviere, Hensher, Swait, Rose)
  – Environmental economics (Carson)
  – Health Economics (Ryan, Louviere, Flynn)
• “Choice-based conjoint” is a term driven by the north american marketing community
  – It uses the name of an established academic technique – the statistical technique of “conjoint measurement” in a misleading manner
Choice based measurement

• Thurstone (1927)
  – who is also the father of psychometrics
  – But Random Utility Theory is his work that is of interest to us

• Options compete
  – Force discrimination between options
  – Discrete choices – don’t ask respondents for numbers
  – Respondents make errors: the size of underlying differences on some latent scale relative to errors gives us importance

• Choice frequencies indicate value
DCE History

• Based on integrated, sound & tested theory.
  – Thurstone’s (1927) random utility theory (RUT).

• DCEs owe much to McFadden (1974) - extended Thurstone’s pairs to multiple choices.
  – RUT: latent construct called “utility” in individual’s heads cannot be observed by researchers.
    • Unobservable (latent) utilities for choice options.
    • Latent utilities decomposable into 2 components - systematic/explainable + random/unexplainable component.
      – Systematic component - attributes of choice options that can be identified and measured, and factors that explain differences in individuals’ choices.
      – Random component includes all unidentified factors that impact choices.
Random utility theory

• It assumes (REQUIRES) respondents to make errors
• The size of the errors, relative to the size of the true underlying differences in utilities on the latent scale (noise-to-signal) enables us to get our numbers

• Choices must be probabilistic
  – There must be a non-zero probability that a person will change his/her decision for a given choice on a different occasion
  – Structural versus sampling zeros

• Choices cannot be deterministic
  – “Choose lowest cost option” – NOT ALLOWED
  – “I don’t believe in states worse than dead” – NOT ALLOWED

• You CANNOT aggregate probabilistic and deterministic respondents
  – You have variance heterogeneity (heteroscedasticity) on the latent scale
  – Heteroscedasticity leads to BIAS in limited dependent variable models
Random utility theory

- You CANNOT aggregate probabilistic and deterministic respondents
  - You have variance heterogeneity (heteroscedasticity) on the latent scale
  - Heteroscedasticity leads to BIAS in limited dependent variable models

- What if I do aggregate them?
  - Biased estimates
  - Willingness to pay figures wrong (might be too high or too low)
  - You are misusing (public) funds in your research giving misleading policy advice!
Random utility theory and the logit model

- Conditional logit derived from iid EV1 errors
  - $U_{jn} = V_{jn} + \varepsilon_{jn}$
  - $\varepsilon_{jn} \sim$ as iid extreme value Type 1; covariances = 0, variances constant & equal
  $\Rightarrow$ conditional logit model:

$$Pr(i) = \frac{e^{\lambda V_i}}{\sum_{J} e^{\lambda V_j}} \quad j = 1, \ldots, J$$

$$V_{jn} = \lambda \left[ \beta_{0jn} + \beta_{1jn} f \left( X_{1jn} \right) + \ldots + \beta_{Kjn} f \left( X_{Kjn} \right) \right]$$

$\lambda$ is a scale constant, inversely proportional to $\sigma_{\varepsilon}$.
DCE example – housing choice

Set 1 of 16:

Choose a home to rent

- **Click** to enlarge images to see more details
- Images are EXAMPLES ONLY of a particular property type and size
- All homes are the same MEDIUM STANDARD QUALITY

Monthly rental payments

- **Home A**: $948
- **Home B**: $1240
- **Home C**: $948
- **Home D**: $948

Assume you are considering RENTING one of the following homes to live in, which would you MOST prefer to live in? (tick one)

- [ ] Home A
- [ ] Home B
- [ ] Home C
- [ ] Home D
DCE example – wine choice
Designing DCEs - Intro

- Discrete choice experiments (DCEs)
  - Conjoint Analysis ≠ DCEs
    - Inconsistent with theory of demand
    - Primarily statistical & ad hoc
    - Incapable of dealing with full range of behaviors
    - And many other reasons
  - Pioneered by Louviere & Woodworth (JMR, 1983) & Hensher & Louviere (1982)
  - Widely used in marketing, applied economics, transportation and other fields
Principles of design

• Alter levels (amounts) of attributes (features) in a systematic way
• Observe how most preferred choice changes in response
• Repeated questioning
  – Rely on choice frequencies so can’t get just one observation per person!
  – Observe how consistent respondents are in choices
• Minimise the opportunity for respondents to make (different) assumptions about the study and its aims
  – Do not have particular options appear much more often than others if at all possible
Reducing the number of choice sets

• Unlikely to ask people to evaluate full factorial – other options?
  – In order of increasing complexity:
    • OMEP (resolution 3) only
    • OMEP with main effects independent of 2-way interactions
    • Main effects + selected 2-way interactions (resolution 4)
    • Main effects + all 2-way interactions (resolution 5)
    • Higher resolution designs
    • Full factorial blocked
How many choice sets?

- Decide UP FRONT whether you need to estimate two-way and higher interactions
  - If the disutility of pain is different, depending on whether the respondent is depressed, that is a two-way interaction between pain and depression
  - You can’t go back and test for this if you didn’t plan for it in design – why?
  - We get smaller (e.g. MAIN EFFECTS) designs by DELIBERATELY CONFOUNDING interactions with main effects

So, the beta coefficient for “extreme pain” is in fact the beta for “extreme pain main effect + two-way interaction between x & y + three-way interaction between x & y & z”

- Those two interactions must be ZERO if your beta is giving you what you really want – the main effect of extreme pain
Assume we made no mistakes regarding main effects, interactions etc. Which type of design?

• Statistically and econometrically it largely won’t matter (Rose) with following caveats
  – Ensure all parameters are estimable (no massive correlations)
  – If sample size is small, an efficient design is more likely to identify significant effects (better precision and power)
  – If you have no priors on estimates, efficient design might “home in” on “wrong” part of utility space – orthogonal design may be better

• Psychologically and conceptually it MIGHT matter
  – Not all designs are equally easy
  – Some designs might induce people to use a heuristic that ISN’T their true decision rule, and NO AMOUNT OF ECONOMETRICS OR STATISTICS WILL SAVE YOU!
    • Remember James Tulsky’s point about the “default” option?
    • There are all sorts of possibilities regarding “where you start”, “path dependency” etc
Large versus small

• **Blocking designs**
  
  – eg, randomly assign 2048 treatments to blocks of 16 (128) without replacement.
  
  – Randomly assign 4 people to each block (512).
  
  – Can test all main effects & interactions, but you must assume all folks are preference clones.

• **If you use smaller designs, you must make STRONG assumptions – see next slide.**
Binary response design example

• **Typical design objectives are:**
  
  – Identification - what form(s) of utility functions can be estimated.
    
    • Some designs allow only main-effects; others allow non-additive models.
  
  – Precision - estimate confidence intervals (given specification & sample size).
  
  – Cognitive complexity - task complexity &/or difficulty
    
    • Little consensus/empirical evidence on optimum levels (pilot tests inform this).
  
  – Market realism - how closely experiment/task simulates real market.
    
    • The more they resemble actual markets, the higher face validity.
Design – brief history

• Orthogonal designs for linear models
  – Each attribute is independent of every other’s
  – Multicollinearity is ZERO by DESIGN
  – Span the utility space – so potentially inefficient

• $2^J$ designs
  – Options present/absent
  – Choice set size vary (potentially bad)

• Efficient designs
  – Good statistical properties
  – Possible problems cognitively
Design – orthogonal designs

- Zero correlations across attributes
- Each attribute is experimentally manipulated completely independently of other attributes
- Make analysis easy, summary statistics less likely to be misleading
# Design – $2^j$ designs

<table>
<thead>
<tr>
<th></th>
<th>Object W</th>
<th>Object X</th>
<th>Object Y</th>
<th>Object Z</th>
<th>Choice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set 1</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>W</td>
</tr>
<tr>
<td>Set 2</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>x</td>
<td>W</td>
</tr>
<tr>
<td>Set 3</td>
<td>✓</td>
<td>✓</td>
<td>x</td>
<td>✓</td>
<td>W</td>
</tr>
<tr>
<td>Set 4</td>
<td>✓</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>W</td>
</tr>
<tr>
<td>Set 5</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
</tr>
<tr>
<td>Set 6</td>
<td>✓</td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td>W</td>
</tr>
<tr>
<td>Set 7</td>
<td>✓</td>
<td>x</td>
<td>✓</td>
<td>x</td>
<td>W</td>
</tr>
<tr>
<td>Set 8</td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>W</td>
</tr>
<tr>
<td>Set 9</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>x</td>
<td>X</td>
</tr>
<tr>
<td>Set 10</td>
<td>x</td>
<td>✓</td>
<td>x</td>
<td>✓</td>
<td>X</td>
</tr>
<tr>
<td>Set 11</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>Y</td>
</tr>
<tr>
<td>Set 12</td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>W</td>
</tr>
<tr>
<td>Set 13</td>
<td>x</td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td>X</td>
</tr>
<tr>
<td>Set 14</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>x</td>
<td>Y</td>
</tr>
<tr>
<td>Set 15</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>Z</td>
</tr>
<tr>
<td>Set 16</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>-</td>
</tr>
</tbody>
</table>
Design – efficient designs

• Obtain high levels of statistical efficiency
  – More info per choice set
• Do this by having minimal overlap
  – Every attribute varies across alternatives (profiles)
  – E.g. in alternative one pain will be “none”, in alt 2 “some”, and in alt 3 “severe”
• The respondent must trade across all attributes
  – Potentially high cognitive burden
  – Can’t “ignore attributes x, y, and z because they’re constant”
Design – bayesian efficient designs

- Don’t attempt to “span the whole utility space” which orthogonal designs do
- Therefore potentially don’t “waste” questions asking about areas of utility space that are not relevant to a respondent
- “home in” on key area of utility space that are relevant to respondents
- Compare the “crucial” alternatives that give maximal information about the parameters of interest
Current state of DCE art

• Design theory still young; many unresolved problems & issues.
  – Some serious design-induced artifacts have been detected,
  – work is underway to resolve them.
Those were the basics of design.

What about analysis?
Analysis (1)

• LOOK at your data
• THINK about how heterogeneity might manifest itself
  – Is it reasonable to assume there is a continuous (standard) distribution of preferences for a given parameter?
  – Might the population’s preferences be “lumpy”? Different segments (cluster/classes) valuing different things?
• Statistical models make ASSUMPTIONS about how the data were generated and collected
  – Use a statistical model that is appropriate to the psychological model you believe people followed
  – To come later – why you should not call any of your methods/models “maxdiff”
### Analysis (2): marginal means

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dementia</strong></td>
<td>383</td>
<td>1,949</td>
<td>2,332</td>
</tr>
<tr>
<td></td>
<td>16.42</td>
<td>83.58</td>
<td>100.00</td>
</tr>
<tr>
<td><strong>Cog ok</strong></td>
<td>721</td>
<td>1,611</td>
<td>2,332</td>
</tr>
<tr>
<td></td>
<td>30.92</td>
<td>69.08</td>
<td>100.00</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>1,104</td>
<td>3,560</td>
<td>4,664</td>
</tr>
<tr>
<td></td>
<td>23.67</td>
<td>76.33</td>
<td>100.00</td>
</tr>
</tbody>
</table>
• Analysis is part art, part science

• If you think you can follow a set of tests from the typical economist’s/statistician’s “quantitative toolbox” THINK AGAIN....

• The FUNDAMENTAL problem of interpretation

Estimated beta is NOT an estimate of the true beta (preference) – it is a PERFECTLY CONFOUNDED estimate of beta and a function of the variance on the latent scale...
Analysis (4)

```
. logit choice  cog_ok stroke antibiotics tubefeeding ventilation

Iteration 0:  log likelihood =  -5070.581
Iteration 1:  log likelihood =  -4427.5438
Iteration 2:  log likelihood =  -4390.2633
Iteration 3:  log likelihood =  -4390.198
Iteration 4:  log likelihood =  -4390.198

Logistic regression

Number of obs    =        9328
LR chi2(5)       =      1360.77
Prob > chi2      =    0.0000
Pseudo R2        =      0.1342

Log likelihood =  -4390.198

| choice          | Coef.  | Std. Err. |    z  |   P>|z| | [95% Conf. Interval] |
|-----------------|--------|-----------|-------|-------|----------------------|
| cog_ok          | 0.4830 | 0.0272    | 17.72 | 0.000 | 0.4295944 0.5364611 |
| stroke          | 0.7567 | 0.0286    | 26.40 | 0.000 | 0.7005058 0.8128387 |
| antibiotics     | 0.6261 | 0.0433    | 14.46 | 0.000 | 0.5412166 0.7109527 |
| tubefeeding     | -0.3386 | 0.0485 | -6.98 | 0.000 | -0.4337163 -0.2434395 |
| ventilation     | -0.5191 | 0.0503 | -10.32 | 0.000 | -0.6177216 -0.4205404 |
| _cons           | -1.4344 | 0.0298 | -48.11 | 0.000 | -1.492787 -1.375928  |
```
Implications of scale in CL

- CL parameters not identified unless $\lambda$ fixed.
  - Recall $\lambda$ inversely proportional to $\sigma_\varepsilon$
    - Larger error variances $\Rightarrow$ smaller $\beta$’s
    - Smaller error variances $\Rightarrow$ larger $\beta$’s
  - Suppose $\lambda$ not constant, but varies:
    - Across individual?
    - Within individuals?
    - With covariates ...?
    - With factors in choice experiments (eg, attributes/levels) ...?

If scale varies over such factors, predicted probabilities may differ a lot!
Why should scale be constant?

• Scale can be impacted by many factors
  – Louviere, et al. (2002) suggested that:
    – $Y | X, Z, C, G, T$, where
      • $Y = \text{behavioral outcomes of interest}$
      • $X = \text{directly observable or manipulable variables}$
      • $Z = \text{characteristics of people whose behaviors are observed}$
      • $C = \text{conditions, contexts, circumstances, or situations}$
      • $G = \text{geographical, spatial, or environmental characteristics can be constant in one place, but vary from place to place}$
      • $T = \text{particular time slices or periods}$
  
• Scale likely to be affected by many variables.
Scale or preference heterogeneity?

• Most researchers assume preference diffs.
  – But, you can’t separate preference differences from scale differences in almost all current models.

• Recall model estimates = $\beta/\sigma_\varepsilon$
  – Distributions of $\beta$ meaningful iff $\sigma_\varepsilon$ is constant.
  – Can estimate higher moments, add more latency to models - little more than mere statistical description.
  – Predicted probabilities depend on predicted utilities; so, these probabilities differ if scale differs.

• Now we show that scales differ.
We KNOW $\sigma_\varepsilon$ is not constant

- Prior evidence that $\sigma_\varepsilon$ varies (Louviere, et al 2001) with:
  - Data sources (Swait & Louviere 1993)
  - Task complexity (Swait & Adamowicz 2002; DeShazo & Fermo 2004; Islam, et al 2008a,b)
  - Geography & time (Severin, et al 2000)
  - Attribute levels (Eagle & Louviere 2006; Islam, Louviere & Burke, 2007; Meyer & Louviere 2007)
  - People (many authors)
  - Model specification (Train & Weeks 2005)
Summary: What if $\sigma_\varepsilon$ not constant?

- Price estimates (elasticities), other effects biased.
  - WTP & other policy estimates likely biased.
- Biased parameters may predict poorly.
- Random parameter models confound scale & “true” preference heterogeneity.
- Hyper-parameters in HB &/or MIXL biased, & LC segment differences misleading.
  - Are preference differences in attribute parameters?
  - Are scale differences in attribute parameters?
  - Most likely some combination of BOTH (or other).
So what do I do????

• Use THEORY
  – That’s another reason why conjoint approaches are problematic
  – they have no theory!

• Use common sense
  – If you get uniformly larger beta values among those with higher literacy skills is it really realistic that they have larger preferences than those with lower literacy (true higher beta)?
  – Or is it more likely the denominator (variance) is smaller and they are merely more consistent in their choices (make fewer and smaller errors) due to better cognitive and reading skills?

• Use MULTIPLE sources of information
  – DON’T use a single design, if you can possibly avoid it
  – Information that is external to the DCE task is ESSENTIAL if you are to make any progress in disentangling (decomposing) the mean-variance confound
Takeaways

• Work with design experts until you have enough experience
• Do NOT assume that experience of using logit/probit models in medical statistics/econometrics qualifies you to do DCEs
• Think about the psychology of the task
• Whenever you read a DCE paper, ask yourself
  – did they consider how choice consistency might manifest itself?
  – How might their reported estimates be explained by variances rather than means on the latent scale?
• If you aren’t worried about how to correctly interpret DCE results you haven’t (yet) understood the issues!
Session 1 – key references


