

Summary:

Testing Stated Preference Techniques and Stated Choice Design

A method study of public transport passengers' valuation of quality

The aim of this project is to critically examine stated preference design in order to expose ways in which the design affects the result, and to provide recommendations for SP design improvements.

There are two main objectives:

1. To develop tests that identify well functioning designs.
2. To test whether self-administered, computer based interviews perform as well as traditional home interviews where interviewers visit the interviewees.

The project is based on a pilot survey among public transport users in Oslo. 10 different designs were developed, tested and analysed. We define *SP design* as

- Sampling and recruitment methods,
- data collection method,
- questionnaire structure,
- SP method, and
- design of individual SP method.

Four different SP methods were used in the survey: Stated choice (SC), Contingent Valuation / Transfer Price (CV/TP), the Frisch Method (FM) and Package Option Grid (POG).

SC gives the respondent tradeoffs between two and two alternatives where a number of attributes are described in a way that ensures systematic variation in attribute levels. For this study we have developed 6 different SC designs.

The CV/TP method involves direct questions about willingness to pay, such as, "*How much are you willing to pay for the same journey but with increased travel time?*"

FM is a method to circle in respondents' trade-offs between attributes in a sequence of choices between pairs of alternatives.

The WinMINT Package Option Grid (POG) lets the respondent choose combinations of improvement and worsening of a number of attributes according to their preferences and budget constraint.

Data collection and attrition

Data collection has been divided between 'traditional' home interviews and self-administered, computer-based interviews on a floppy disk that is posted to the respondents.

There occurred to be some differences between the two groups. A 'real' survey must therefore be supplemented by home interviews in order to secure representation by all socio-economic groups.

Recruiting respondents by telephone, about 12 percent of the people called were willing (and met the criteria to participate) to participate in the self-administered interviews. In comparison this figure was 7 percent in the telephone re

Response rates for self-administered interviews lie around 50 percent. This can be improved by developing more user-friendly solutions that do not depend on e.g. operating system. We recommend an internet based solution be developed.

Testing design

The aim of the design tests has been to demonstrate ways in which the design affects respondents' answers, and to establish methods to reduce this problem. We have considered the effects of design on:

- Size, sign and significance level of valuations
- Lexicography, i.e. when respondents sort their all their answers in according to one attribute only, throughout the SC sequence. Lexicography is a

major problem in SC and is caused either by insufficient balancing of the attribute levels (alternatively extreme preferences) or simplification of the choice task by the respondent. There is also a small chance that this happens accidentally.

- Incidence of deviant answers, i.e. choices which, according to the models, there is a very small probability of choosing (in this study we used probabilities of less than 0.05 and 0.1, respectively)
- '0-answers' and overbids: A *0-answer* is either 0 in valuation or 0 in willingness to pay (WTP), An *overbid* is an extremely high valuation of an attribute
- Fatigue effects, which lead to inconsistent answers
- Consistency of respondents' answers across the SP methods
- Respondents' reported degree of difficulty in answering the different SP methods
- Attrition

Stated Choice

Six different SC designs were developed:

1. Standard design: 4 attributes, each with 3 levels. Up to 9 replications/choices. Allowance for uncertainty in choices (can choose "Certainly alternative X" or "Probably alternative X").
2. Price levels balanced according to respondents' WTP in a preceding CV/TP question
3. CV/TP balancing. Include answer categories "Don't know" and "The alternatives are indifferent".
4. An additional SC sequence with 'block design' with groups of attributes.
5. CV/TP balancing. Up to 15 replications to test for fatigue effects.
6. Reduce number of attributes from 4 to 3

There is considerable variation between valuations of quality factors between the different designs. This is illustrated in Table S.1. Some of the differences are caused by random variation, but our tests show that the design has also affected the results.

- A better balancing of the attribute levels will reduce the problem of lexicography. We have sought to do this by balancing the variation of the price levels according to respondents' WTP in a preceding CV/TP question ("CV/TP balancing"). This reduces the proportions lexicographic choices. In other words, by making use of information provided in other SP methods we have managed to reduce an important problem in SC analyses.
- SC games with 15 replications (choices) are regarded by respondents as more difficult than games with up to 9 choices. The long games produce models with poor estimates and poor explanatory power, and have a higher proportion of deviating answers. Instead of providing more information long games reduce the quality of the data. Our analyses show that long SC games not only affect individual games, but also subsequent games.
- Allowing uncertainty in answers we can elicit more information from each observation, and improve the models by attaching different weights to the answers when analysing them.
- We designed a 'block' SC design in order to check for effects of packing attributes. Our estimates show that the value of a package of improvements (seating, travel time and frequency) is up to 35 percent lower than the sum of individual improvements.

Table S.1: Estimated SC valuations (NOK per hour or per trip) of various attributes according to design and data collection method.

Attribute	Design 1	Design 2	Design 3	Design 4	Design 5	Design 6	Home interview	Self-administered	All
Travel time, NOK per hour	7	9	20	17	10	25	8	26	23
Standing place, additional per hour	15	17	17	15	11		16	12	13
Waiting time per hour	77	88	69	61	48	44	82	66	68
Security guards, per trip	2	1	-1	-1	0	2	1	0	0
Emergency phones, per trip	1	1	0	1	-3	-0	1	0	0
Station with roofing, per trip	1	0	-0	0	-5		1	-1	-1
Shelter, per trip	2	2	1	1	1		2	1	1
RTI, per trip	1	0	1	1	3	2	1	2	1
Rout map, per trip	1	1	-1	0	-4	-1	1	-1	-1
N	35	31	53	52	44	55	66	298	364

CV/TP

We have established the following relationship between design and CV/TP valuations of attributes:

- Respondents tend to round off their willingness to pay (WTP). This influences the results.
- When asking for WTP when an attribute is worsened the valuations may be exaggerated because no minimum WTP has been set. A number of respondents choose the lowest possible WTP, which is 0 in this case, and the corresponding extremely high valuations cause quality reductions to be given a higher valuation than improvements in CV/TP games.
- CV/TP gives no trade-off between variables. Hence it is easier to give strategic, or protest, answers. This is the reason for a large number of observations of *zero WTP* (gives high valuations of worsening standards), *zero valuations* and a few *overbids* (gives high valuations of improvements)

The Frisch method

Between 14 and 18 percent of the respondents chose the lowest price alternative in all FM questions, which means they would not accept a higher price for reduced travel time or improved frequencies. We assume that these respondents have 0 WTP. Compared with the problems of 0 WTP in CV/TP and lexicographic choices in SC this is a great improvement.

Between 82 and 95 percent of the respondents have chosen an alternative that indicates the fact that they have given up something in exchange for an improvement in

another attribute. In other words, a large proportion of the respondents have made actual trade-offs within the Frisch framework.

POG seems unsuited for valuation studies

We have not been able to make use of the Package Option Grid data for valuation purposes. POG may be used to learn about travellers' priorities in some respects, but we have not managed to transfer this information into relative valuations. Within POG there are no trade-offs or possibilities for different valuations than those set by the questionnaire designer.

Respondents answer fairly consistent

The different SP methods correspond quite well in determining the *most important attributes* for different respondents. This means that respondents have relatively clear priorities, which they communicate independent of method chosen.

Table S.2 shows an example of valuations calculated in FM, CV/TP and SC. We see that the results are in the same order of magnitude, but with rather large percentual differences.

Table S.2: Valuations of (reduced) travel time and (increased) frequencies in FM, CV/TP and SC. N=289-298¹⁾

	FM ¹⁾	CV/TP ¹⁾	SC
(Reduced) travel time, NOK/hour	32 – 39	19 – 20	23
(Reduced) headway, NOK/hour	32 – 34	22 – 28	34

¹⁾ Lower limit excludes respondents whose WTP exceeds three times their ticket price. In upper limit all respondents are included

Tests also show that respondents with specific preferences give relatively consistent answers across the SP methods. For example, respondents with valuations in CV/TP have also a high valuation of the same attributes in SC. However, there is not much consistency between zero valuations in CV/TP and FM and lexicographic choices in SC. We have concluded that this is due to quite different properties of the methods, and problems with CV/TP in particular.

Self-administration works well

Based on our tests self-administered interviews can be recommended both from a financial and a methodical point of view. We do however recommend the development of more user-friendly solutions, e.g. based on Windows or internet. And, importantly, in order to secure a good representation of the population self administered interviews must be supplemented by home interviews.

Recommendations

- Lexicography is a major problem in stated choice surveys. The proportion lexicographic answers is a good indication of how well the survey has functioned, and should always be reported.
- Zero WTP, extreme WTP and zero valuations in CV/TP and FM are equally problematic. In FM the proportion of such answers is low, and therefore the FM approach is found to be superior to CV/TP.
- There is a great potential for improving SC design by making it more flexible. We have found that by balancing the price level according to the stated WTP in a CV/TP question we have managed to reduce the problem of lexicography. As we have found FM to be superior to CV/TP, we recommend FM be used for this purpose in future surveys.
- Home interviews are expensive. We have found that self-administered interviews, which are less costly, work well. The effect of data collection method on the results is minimal. In order to secure proper representation of the population the data must be supplemented by home interviews.
- Self administered interviews must be made more user friendly in order to improve response rates. We recommend Windows or internet based solutions.
- Increased use of control questions in all types of SP surveys may reveal more information about why respondents express extreme preferences.