

# Tester av Stated Preference-teknikker og samvalgdesign

## En metodestudie av kollektivtrafikanterers verdsetting av kvalitetsfaktorer

Nils Fearnley og Kjartan Sælensminde

Transportøkonomisk institutt (TØI) har opphavsrett til hele rapporten og dens enkelte deler. Innholdet kan brukes som underlagsmateriale. Når rapporten siteres eller omtales, skal TØI oppgis som kilde med navn og rapportnummer. Rapporten kan ikke endres. Ved eventuell annen bruk må forhåndssamtykke fra TØI innhentes. For øvrig gjelder [åndsverklovens](#) bestemmelser.

# Forord

Kollektivtrafikanternes verdsetting av kvalitet er et sentralt element i alle nytte-kostnadsanalyser og planleggingsverktøy som benyttes når nye tilbud skal vurderes. Det er i all overveiende grad *Stated Preference* (SP) analyser, og da spesielt samvalganalyser, som benyttes for å beregne denne verdsettingen. SP-undersøkelser er en samlebetegnelse på en rekke analyser med hypotetiske valg.

Flere undersøkelser viser at designet i stor grad kan påvirke resultatene, og at noen kvalitetsfaktorer er mer følsomme for dette enn andre. Dette kan bety at skjevheter i SP-design kan føre til betydelige feilkilder når nye tilbud skal utvikles. Formålet med prosjektet er derfor å rette et kritisk søkelys på design av SP-analyser for å kunne forbedre analyseverktøyet til beslutningsformål.

I dette metodeprosjektet har vi undersøkt og sammenlignet fire ulike SP-metoder, med hovedfokus på ulike designvarianter av samvalganalyser (parvise valg). Analysene er basert på datamateriale fra en pilotundersøkelse for Oslo Sporveier blant kollektivtrafikanter i Oslo. Vi vil understreke at denne metoderapporten forutsetter en viss kjennskap til SP-analyser.

Prosjektet inngår i Vegdirektoratets etatsprosjekt for kollektivtransport. Prosjektet er finansiert av Vegdirektoratet, mens Oslo Sporveier har finansiert omkostningene til datainnsamlingen. Kontaktpersoner har vært James Odeck og Rune Gjøs fra Vegdirektoratet og Jo Inge Kaastad fra Oslo Sporveier.

Forsker Nils Fearnley har vært prosjektleder fra TØI. Han har skrevet kapitlene 1–5, 7 og 9, med unntak av kapittel 5.8. I tillegg har forsker Kjartan Sælensminde arbeidet på prosjektet og skrevet kapitlene 5.8, 6 og 8. De øvrige kapitlene har forskerne skrevet i fellesskap. Forskningsleder Bård Norheim har vært kvalitetssikrer, og avdelingssekretær Kari Tangen har stått for tekstbehandling og layout.

Oslo, desember 2001  
TRANSPORTØKONOMISK INSTITUTT

*Knut Østmoe*  
instituttssjef

*Ingunn Stangeby*  
avdelingsleder

# Innhold

## Sammendrag

## Summary

<b>1 Bakgrunn og problemstilling .....</b>	<b>1</b>
<b>2 Stated Preference-undersøkelser.....</b>	<b>3</b>
2.1 Ulike SP-metoder .....	3
2.2 Design av samvalganalyser .....	4
2.3 Analysen.....	5
<b>3 Vanlige feilkilder og skjevheter i SP-undersøkelser.....</b>	<b>6</b>
3.1 Utvalgsstrategi.....	7
3.2 Datainnsamling.....	7
3.3 Intervjusituasjonen .....	8
3.4 Leksikografiske svar.....	9
3.5 Beskrivelse av egenskapene .....	10
3.6 Eksempler på testing av ulike design.....	11
3.7 Oppsummering.....	16
<b>4 Metode og datagrunnlag.....</b>	<b>17</b>
4.1 Rekruttering og datainnsamling .....	17
4.2 SP-metoder .....	18
4.3 Struktur på spørreskjemaet.....	21
4.4 Tester av design og verdsetting .....	22
4.5 Frafall og utvalgsskjevheter .....	24
4.6 Utvalget .....	25
4.7 Oppsummering.....	28
<b>5 Samvalganalyser (SC).....</b>	<b>30</b>
5.1 Verdsetting av kvaliteten på reisen .....	30
5.2 Leksikografi .....	33
5.3 9 versus 15 valgonganger.....	36
5.4 Små og store prisvariasjoner .....	36
5.5 Balansering ved hjelp av CV/TP .....	37
5.6 Avvikende svar i SC-sekvensene.....	38
5.7 Læringseffekt og tretthet .....	39
5.8 Preferanseusikkerhet i samvalg .....	40
5.9 Oppsummering samvalganalyser.....	43
<b>6 Samvalg og kontekst: Blokkdesign.....</b>	<b>45</b>
6.1 Test av kontekst og blokkdesign ved hjelp av tre valgsekvenser .....	46
6.2 Resultater fra de ulike valgsekvensene .....	46
6.3 Sammenligning av verdsetting funnet i ulike kontekster.....	47
6.4 Konklusjoner basert på tester av blokkdesign .....	49

<b>7 Betinget verdsetting og likeverdspris (CV/TP)</b> .....	<b>50</b>
7.1 Forverring vektlegges mer enn forbedring.....	50
7.2 Respondentene avrunder svarene.....	51
7.3 Høy andel 0-verdsettinger i CV/TP.....	52
7.4 Overbud i CV/TP.....	52
7.5 Oppsummering.....	53
<b>8 Frisch-metoden (FM)</b> .....	<b>54</b>
8.1 FM kan korrigere leksikografiske svar.....	54
8.2 Inkonsistente valg er vanlig i SC-sekvenser, men umulig innen den enkelte FM-sekvens.....	54
8.3 Resultater ved bruk av Frisch-metoden.....	55
8.4 Respondentenes valg i FM-sekvensene.....	55
8.5 Avveininger mellom reisetid, tid mellom avgangene og takst.....	57
8.6 Tester av konsistens mellom ulike FM-sekvenser.....	59
8.7 Forslag til videre testing av FM.....	59
<b>9 Fleksibelt pakkevalg (POG)</b> .....	<b>61</b>
<b>10 Tester av konsistens mellom de ulike SP-metodene</b> .....	<b>62</b>
10.1 CV/TP versus samvalganalyser.....	62
10.2 FM versus samvalganalyser.....	63
10.3 POG versus samvalganalyser.....	64
10.4 Frisch-sekvensene versus CV/TP.....	64
10.5 Frisch-metoden, CV/TP og samvalganalyser.....	64
10.6 Oppsummering.....	66
<b>11 Oppsummering, vurdering av SP-metodene og konklusjoner</b> .....	<b>67</b>
11.1 Testing av utvalget.....	68
11.2 Tester av samvalgsekvensene.....	68
11.3 Tester av CV/TP-sekvensene.....	70
11.4 Tester av Frisch-metoden.....	70
11.5 POG virker uegnet for verdsettingsstudier.....	70
11.6 Noe konsistens mellom SP-metodene.....	71
11.7 Struktur på spørreskjemaet.....	71
11.8 Sammenlignende vurdering av metodene.....	72
11.9 Egenadministrasjon fungerer tilfredsstillende.....	73
11.10 Oppsummering og konklusjoner.....	74
<b>Litteratur</b> .....	<b>75</b>
<b>Vedlegg 1: Eksempler på rekruttering av respondenter via Trafikanten og Oslo Sporveiers hjemmesider</b> .....	<b>79</b>
<b>Vedlegg 2: Frafallsregistrering: Telefonrekruttering for egenadministrerte intervjuer</b> .....	<b>81</b>
<b>Vedlegg 3: Frafallsregistrering: Telefonrekruttering for hjemmeintervjuer</b> .....	<b>83</b>
<b>Vedlegg 4: Frafall i felt, dvs. frafall ved hjemmebesøk for hjemmeintervjuer</b> .....	<b>85</b>
<b>Vedlegg 5: Frekvenstabeller spørreskjemaet</b> .....	<b>87</b>
<b>Vedlegg 6: Beregninger av samvalgene</b> .....	<b>103</b>
<b>Vedlegg 7: Oversikt over leksikografiske svar</b> .....	<b>113</b>

Sammendrag:

# Tester av Stated Preference-teknikker og samvalgdesign

## En metodestudie av kollektivtrafikanter verdsetting av kvalitetsfaktorer

Dette prosjektet har hatt som formål å kritisk gjennomgå design av *Stated Preference* (SP) undersøkelser for å avdekke måter designet påvirker resultatet og gi anbefalinger til hvordan problemet kan reduseres.

Hovedproblemstillingen har vært todelt:

1. I hvilken grad er det mulig å utvikle tester som avdekker hva som er god eller dårlig design på SP-undersøkelser, og kan disse testene benyttes for å vurdere hvilke design som fungerer best?
2. Vil egenadministrerte PC-intervjuer fungere like bra som hjemmeintervjuer?

Det er blitt gjennomført en pilotstudie blant Oslos kollektivreisende, der 10 ulike design for undersøkelsen er blitt testet og analysert. Med *design* forstår vi:

- Utvalgs- og rekrutteringsmetode
- Metode for datainnsamling
- Struktur på spørreskjema
- SP-metoder
- Utforming av de enkelte SP-metodene

Undersøkelsen omfattet fire SP-metoder: Parvise samvalg (SC), likeverdpris (CV/TP), Frisch-metoden (FM) og Fleksibelt pakkevalg (POG).

I SC verdsettes flere goder i forhold til hverandre, og disse godene kan både være markedsgoder og ikke-markedsgoder. I undersøkelsene blir respondenter stilt overfor valg mellom reisealternativer som består av en "pakke" av egenskaper (f.eks. reisetid, pris, regularitet og komfort) ved en bestemt reise. I undersøkelsen inngikk 6 ulike SC-design.

CV/TP er mer direkte spørsmål om betalingsvillighet av typen "*Hva er det maksimale du ville betalt for reisen med økt reisetid?*"

FM er en metode som sirkler inn en respondents avveining mellom to og to egenskaper. I en serie valg

mellom to reisealternativer blir forskjellene i nivåene på egenskapene gradvis redusert i retning av respondentens preferanser.

POG lar respondenten velge den kombinasjonen av godeforbedringer som gir størst nytte. Samtidig må respondenten ta hensyn til at dette påvirker prisen.

## Datainnsamlingsmetode og frafall

Datainnsamlingen har vært fordelt mellom "tradisjonelle" hjemmeintervjuer og egenadministrerte, PC-baserte intervjuer på diskett som respondentene har fått tilsendt.

Det er en viss forskjell mellom de to utvalgene, som i noen grad skyldes at forskjellige grupper har tilgang til PC og vegrer seg mot hjemmeintervju. I et hovedprosjekt er det derfor nødvendig å supplere egenadministrerte intervjuer med hjemmeintervjuer blant underrepresenterte respondentgrupper.

I telefonrekrutteringen av respondenter var om lag 12 % villige til å delta i egenadministrerte intervjuer, og oppfylte kriteriene for å delta i undersøkelsen. Til sammenligning var andelen 7 % for rekrutteringen til hjemmeintervjuer.

Frafallet blant dem som fikk tilsendt diskett for egenadministrerte intervjuer ligger jevnt på ca 50 %. Andelen kan reduseres ved å utvikle mer brukervennlige løsninger som ikke avhenger av operativsystem, f.eks. internett-baserte løsninger.

## Tester av designet

I designtestene er hovedformålet å påvise måter designet har påvirket respondentenes svar, og måter dette problemet kan reduseres. Vi har sett på forhold ved designet som påvirker

- størrelse, signifikans og fortegn på verdsettingene.

- leksikografi, dvs når respondenten sorterer alle alternativene i SC etter én (eller flere) egenskaper gjennom hele samvalgsekvensen. Leksikografi er et alvorlig metodeproblem i samvalganalyser, og kan skyldes manglende balansering av egenskapsnivåene (ev. spesielle preferanser), eller at respondenten forenkler beslutningsrutinen. Det er også en viss sannsynlighet for at det skjer tilfeldig.
- forekomsten av avvikende svar, dvs. svar det ut fra modellene er svært liten (i denne studien definert som hhv. 0,05 og 0,1 prosent) sjans for å velge.
- 0-svar og overbud: 0-svar omfatter 0 i verdsetting og 0 i betalingsvilje, mens overbud innebærer ekstremt høy verdsetting av en kvalitetsfaktor.
- tretthetstegn, som fører til mindre konsistent svargivning.
- konsistens i svargivningen på tvers av SP-metodene
- respondentenes vurdering av vanskelighetsgrad ved metodene.
- frafall.

## Samvalgene

I alt seks ulike samvalgsdesign er blitt utarbeidet:

1. Standarddesign: Fire kvalitetsegenskaper med tre nivåer hver og opptil 9 valg. Preferanseusikkerhet (kan velge "*Absolutt alternativ X*" eller "*Trolig alternativ X*").
2. Balansering av nivåene ved hjelp av CV/TP.
3. CV/TP-balansering, samt svaralternativ "Vet ikke" og "Like gode".
4. Standarddesign, og i tillegg et eget blokkdesign med grupper av faktorer på flere nivåer.
5. CV/TP-balansering. Ta med opptil 15 valg for å studere tretthetseffekter.
6. Redusere antall faktorer fra 4 til 3 for å se hvordan dette påvirker leksikografi og verdsetting.

I samvalgsekvensene er det til dels store sprik i verdsettingen av kvalitetsfaktorer mellom designvariantene. Tabell S.1 viser et utdrag av beregningene. Noe av forskjellene skyldes tilfeldig variasjon i de små utvalgene, men tester viser at forskjellene også skyldes egenskaper ved designet.

- Bedre balansering av nivåene i SC-designet gir lavere andeler leksikografi og forbedrer samvalgsekvensene betraktelig. Med bakgrunn i dette er det forsøkt å forbedre balanseringen ved å la variasjonen rundt basisprisen bli bestemt av respondentenes svar i en forutgående CV/TP-sekvens. Dette gir lavere andeler leksikografi. Et slikt opplegg med å la SP-metodene spille sammen har med andre ord hjulpet til med å redusere et viktig metodeproblem.
- Samvalgsekvenser med 15 valgoganger oppfattes av respondentene som vanskeligere å besvare enn sekvenser med opptil 9 valgoganger. De har fungert relativt dårlig med hensyn til de vanligste indikatorene for forklaringskraft, og har høyere andeler avvikende svar. Lange samvalgsekvenser skaper derfor usikkerhet rundt beregningene snarere enn å gi oss mer informasjon. Våre analyser indikerer videre at lange samvalgsekvenser ikke bare påvirker den enkelte samvalgsekvensen, men også senere sekvenser i samme intervju.
- Ved å åpne for preferanseusikkerhet i svargivningen kan vi hente mer informasjon fra hver observasjon, ved at det blir mulig å vekte valgene i analysene.
- Blokkdesign, som fanger opp pakkeeffekter (at verdsettingen av en pakke med forbedringer er forskjellig fra summen av de individuelle forbedringene), viser at faktorene relatert til reisen (sitteplass, reisetid og frekvens) verdsettes opptil 35 % lavere når de forbedres samtidig enn når de bare verdsettes enkeltvis.

Tabell S.1: Beregnede SC-verdsetninger (kroner pr time eller pr reise) av egenskapene fordelt på ulike design og datainnsamlingsmetoder

Kvalitetsfaktor	Design 1	Design 2	Design 3	Design 4	Design 5	Design 6	Hjemme-intervju	Egen-adm.	Alle
Reisetid, kr pr time	7	9	20	17	10	25	8	26	23
Ståplass, tillegg/time	15	17	17	15	11		16	12	13
Ventetid, kr pr time	77	88	69	61	48	44	82	66	68
Vektore, kr pr reise	2	1	-1	-1	0	2	1	0	0
Nødtelefon, kr pr reise	1	1	0	1	-3	-0	1	0	0
Overbygg, kr pr reise	1	0	-0	0	-5		1	-1	-1
Leskur, kr pr reise	2	2	1	1	1		2	1	1
Realtid, kr pr reise	1	0	1	1	3	2	1	2	1
Rutekart, kr pr reise	1	1	-1	0	-4	-1	1	-1	-1
Ant. respondenter	35	31	53	52	44	55	66	298	364

## CV/TP-sekvensene

Vi har påvist følgende sammenhenger mellom egenskaper ved designet og CV/TP-verdsetninger av kvalitetsfaktorer:

- Respondenter har en tendens til å avrunde svarene sine. Dette påvirker resultatene og skaper skjevheter i verdsettingene.
- I spørsmål om kompensasjonskrav for en *forverring*, kan verdsettingen bli overdreven stor, fordi det ikke er spesifisert noen nedre grense for respondentens *betalingsvilje* for den forverrede reisen. Dette er en viktig årsak til at forverring tillegges større vekt enn forbedringer i CV/TP.
- Spørsmålsstillingen i CV/TP er direkte og har ingen elementer av avveining, noe som gjør metoden utsatt for protestvalg og strategiske svar. Dette er årsaken til flere observasjoner av 0 i *betalingsvilje* (gir høy verdsetting av en egenskap som forverres), en stor andel med 0 i *verdsetting*, og en del *overbud* som gir ekstremt høye verdsetninger av forbedringer.

## Frisch-metoden

Mellom 14 og 18 % av respondentene valgte alternativet med lavest pris i alle FM-valgene, dvs. ville ikke godta høyere pris for å få hhv. redusert reisetid eller høyere frekvens. Vi har antatt at disse har 0 i betalingsvillighet. Sammenlignet med andelen respondenter med 0 i betalingsvilje i CV/TP, og som sorterer leksikografisk etter prisen i SC, er problemet med 0 i betalingsvillighet betraktelig redusert i FM.

Mellom 82 og 95 % av respondentene har en eller flere ganger valgt alternativet der de måtte avgi noe for å få en forbedring. Disse har altså veid endringer i godene mot hverandre. En stor andel respondenter foretar altså faktiske avveining i FM.

## POG virker uegnet for verdsettingsstudier

Vi har ikke klart å gjøre nytte av materialet fra POG-sekvensene. POG er muligens egnet til å kartlegge hvilke kvalitetsfaktorer trafikantene prioriterer, men vi har ikke identifisert noen metode for å analysere POG-dataene som er tilfredsstillende for verdsettingsstudier.

## Respondentene svarer delvis konsistent

Det er generelt godt samsvar mellom metodene når det gjelder hvilke egenskaper som er *viktigst* for de enkelte respondentene. Trafikantene har altså relativt klare prioriteringer som kommer til syne uavhengig av SP-metode.

Tabell S.2 viser et eksempel på verdsetninger som er beregnet i FM, CV/TP og samvalgene. Resultatene fra disse metodene er i samme størrelsesorden, men likevel med prosentvise store sprik.

Tabell S.2: Verdsetting av (reduisert) reisetid og (reduisert) tid mellom avgangene fra FM, CV/TP og SC. N=289-298<sup>1)</sup>

	FM <sup>1)</sup>	CV/TP <sup>1)</sup>	SC
(Redusert) reisetid (kr/time)	32 – 39	19 – 20	23
(Redusert) tid mellom avgangene (kr/time)	32 – 34	22 – 28	34

<sup>1)</sup> Nedre grense er avveiningene når respondenter som tilkjenner ga betalingsvillighet på mer enn tre ganger taksten er ekskludert. I øvre grense er alle respondentene er inkludert.

Tester viser også at respondenter med spesielle preferanser svarer relativt konsistent på tvers av SP-metode. For eksempel har respondenter med høy betalingsvilje i CV/TP også tilkjenner en høyere verdsetting i samvalgene. Samtidig er det lite sammenheng mellom 0-verdsettinger i CV/TP og FM og leksikografiske svar i samvalgene. Vi har konkludert med at egenskaper ved spørsmålstillingen i de ulike metodene, og spesielt CV/TP har forårsaket dette.

## Egenadministrasjon fungerer tilfredsstillende

Med bakgrunn i våre tester mener vi at egenadministrerte intervjuer er forsvarlig både faglig sett og ikke minst økonomisk. Vi anbefaler at det utarbeides mer brukervennlige løsninger for egenadministrerte intervjuer, f.eks. internett- eller windows-baserte løsninger, i den videre utviklingen av egenadministrerte intervjuer. Og vi vil igjen understreke nødvendigheten ved å supplere egenadministrerte intervjuer med hjemmeintervjuer for å sikre representasjon fra hele befolkningen.

## Anbefalinger

- Leksikografi er et alvorlig problem i samvalgundersøkelser. Andelen leksikografi gir en god indikasjon på hvor godt samvalgdesignet har fungert, og bør alltid rapporteres.
- 0-svar og ekstreme verdsettinger er tilsvarende alvorlige problemer i CV/TP og FM. Disse andelen må derfor også rapporteres. FM har vist seg å ha lave andeler slike svar og er derfor funnet å være en bedre verdsettingsmetode enn CV/TP.
- Det er et vesentlig forbedringspotensial i å utvikle mer fleksibelt design av samvalgene, f.eks. ved å balansere nivåene i samvalgsekvensen ut fra betalingsvilje i andre typer SP-metoder. I denne studien er CV/TP brukt til dette. Dette gir en klar tendens til lavere andeler leksikografi. På grunn av metodiske svakheter ved CV/TP anbefales FM i stedet for CV/TP til å balansere samvalgsekvensen i fremtidige undersøkelser.
- Hjemmeintervjuer er svært ressurskrevende. Våre undersøkelser viser at egenadministrasjon, som er langt rimeligere, har fungert godt. Datainnsamlingsmetoden har hatt svært liten innvirkning på resultatene. Imidlertid må egenadministrerte intervjuer suppleres med hjemmeintervjuer for å sikre at alle grupper av befolkningen blir representert.
- Egenadministrerte intervjuer bør gjøres mer brukervennlige for å redusere frafallet. Vi anbefaler at det utvikles Windows- eller internettbaserte løsninger.
- Det anbefales økt bruk av kontrollspørsmål i alle typer SP-undersøkelser. Kontrollspørsmål kan avdekke årsaker til at respondentene tilkjenner ekstreme preferanser, og forekomsten av protestvalg og forenklete beslutningsrutiner.



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=289-298<sup>1)</sup>

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

<sup>1)</sup> 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.

# 1 Bakgrunn og problemstilling

Kollektivtrafikanternes verdsetting av kvalitet er et sentralt element i alle nytte-kostnadsanalyser og planleggingsverktøy som benyttes når nye tilbud skal vurderes. Det er i all overveiende grad SP-analyser, og da spesielt samvalganalyser, som benyttes for å beregne denne verdsettingen. Samtidig viser flere undersøkelser at designet i stor grad kan være med på å påvirke resultatene i undersøkelsene, og at noen kvalitetsfaktorer er mer følsomme for dette enn andre. Dette kan bety at skjevheter i design eller ”kopiering” av resultater fra andre undersøkelser kan føre til betydelige feilkilder når nye tilbud skal utvikles.

I prosjektet har vi definert "design" i vid forstand. Det vil si at vi legger følgende aspekter inn i betraktninger rundt designbegrepet:

- Utvalgs- og rekrutteringsmetode
- Metode for datainnsamling
- Struktur på spørreskjema,
- SP-metoder
- Utforming av de enkelte SP-metodene

I utgangspunktet var dette prosjektet ment som et samarbeid mellom TØI og høyskoler i Norge. TØI skulle yte bistand til studenter som ville bruke SP-analyser i oppgavene sine, og som skulle gjøre datamaterialet tilgjengelig for våre analyser. Ingen studenter viste interesse for å bruke SP-analyser, og prosjektet ble redefinert til å inngå som en pilotundersøkelse for en større verdsettingsstudie blant kollektivtrafikanter i Oslo-regionen som er planlagt i 2002. På mange måter har dette vært en heldig utvikling for prosjektet. Blant annet har det medført at vi har fått en langt bedre kontroll med de ulike designene som skal testes, og også et mer homogent utvalg i undersøkelsene.

Formålet med prosjektet er å rette et kritisk søkelys på design av SP-analyser for å kunne forbedre analyseverktøyet til beslutningsformål. Som et grunnlag for en slik drøfting har vi:

1. Gjennomgått hovedprinsippene for en verdsetting av kvalitet ved SP-analyser og hovedgruppene for ulike SP-analyser og design.
2. Oppsummert de forskningsrapportene som har testet alternative design.
3. Foretatt en konkret testing av alternative design.
4. Oppsummert hovedresultater for pilotundersøkelsen av Oslo-trafikanternes verdsetting av kvalitet.

5. Oppsummert sentrale retningslinjer og kjennetegn på en ”god” design.

SP-analyser er en relativt ny metode innenfor transportforskningen, og det er først de siste ti - femten årene at metoden for alvor har kommet i praktisk anvendelse. Dette skyldes i all hovedsak EDB-utviklingen, som har gjort det mulig å foreta mer skreddersydde undersøkelser og mer avanserte statistiske analyser og tester av svarene. Det er en rekke SP-analyser som er gjennomført både i Norge og internasjonalt, men relativt få som konsentrerer seg om alternative design. Dette skyldes i første rekke at det er relativt kostnadskrevenne å gjennomføre datainnsamling og at tradisjonelle oppdragsgivere som regel ikke er interessert i økte kostnader til metodetesting.

Den mest omfattende testingen av alternative design ble gjennomført av Staffan Widlert i Sverige (1994) hvor en del studenter gjennomførte intervjuene basert på ulike typer design, både PC-baserte og papirbaserte intervju i tillegg til ranking, parvise valg, skala osv. Konklusjonene på denne analysen var blant annet at designet på SP-undersøkelsene i stor grad kunne påvirke resultatene i undersøkelsen og at det var viktig å utvikle tester som avslørte hvordan intervjuobjektene taklet valgsituasjonen. Det var samtidig ingen endelig konklusjon på hva som var et ”riktig” design, selv om PC-baserte intervjuer fungerte bedre enn de papirbaserte.

Det er grunn til å understreke at designet vil påvirke alle typer undersøkelser av transportmarkedet og en sentral del av metodeutviklingen innenfor reisevaneundersøkelser er for eksempel relatert til sammenhengen mellom design og mål på transportmiddelfordelingen (Stopher og Jones 2001). Mål på kollektivandeler vil f.eks bare kunne sammenliknes hvis en kjenner designet på undersøkelsen og hvordan utvalget er rekruttert med mer. Det spesielle med SP-analyser er at dette er en relativt ny metode og at disse sammenhengene i mindre grad er testet og dokumentert.

Tidligere tester har vist at *Stated Preference* kan være godt egnet til å kartlegge trafikantenes preferanser. Det er imidlertid et klart skille mellom de "tunge" kvalitetsfaktorene, som ulike reisetidsfaktorer, bytte m.m., og mer marginale faktorer som utrygghet, lavgulvvogner og informasjon på holdeplassen. Dette skyldes i all hovedsak at trafikantene verdsetter de marginale faktorene svært forskjellig og at en gjennomsnittsverdi dermed blir mer usikker (Norheim 1996).

Utformingen av spørreskjemaet kan spille en stor rolle for de svarene som avgis. Det er i hovedsak to hovedproblemer som er avdekket i tester knyttet til designet av undersøkelsen:

- Dominante kvalitetsfaktorer, dvs. at noen kvalitetsfaktorer "overdøver" de andre faktorene.
- Utvalgsstørrelse og segmentering, dvs. at en person foretar flere valg og dermed "tilslører" den reelle utvalgsstørrelsen.

Vi vil i denne undersøkelsen definere et "godt design" som en undersøkelse hvor det er mulig å avdekke om intervjupersonene har foretatt "reelle valg" og om det er mulig å generalisere resultatene til en større gruppe av trafikanter. Dette betyr at det ikke er et spørsmål om vi "tror på" resultatene fra SP-undersøkelsen, men om undersøkelsen gir grunnlag for å avdekke om det er konsistente og pålitelige resultater. Målsettingen med prosjektet er derfor å utvikle noen "kjøreregler"/tester for SP-analyser som gjør at beslutningstakere og de som skal benytte resultatene i planleggingsformål kan ha større

tillit til resultatene. Hovedproblemstillingen i dette prosjektet er todelt:

1. I hvilken grad er det mulig å utvikle tester som avdekker hva som er god eller dårlig design på SP-undersøkelser, og kan disse testene benyttes for å vurdere hvilke design som har fungert best?
2. Vil egenadministrerte PC-intervjuer fungere like bra som hjemmeintervjuer, og i hvilken grad vil dette gi skjevheter i intervjuet som ekskluderer sentrale personer i målgruppen?

Testingen vil konsentrere oppmerksomheten om:

- Ulike design av samvalgundersøkelser (parvise valg)
- Utvikling av overlappende tester for å avdekke gode/dårlige design og konsistensen i svarene som er avgitt
- Betydningen av dominante faktorer og etablering av design for å redusere dette problemet
- Effekter av alternativ datainnsamling

## 2 *Stated Preference*-undersøkelser

SP-undersøkelser kjennetegnes ved å ha en eksperimentell design der egenskapsnivåene (ved f.eks. en reise) varierer systematisk (som regel statistisk uavhengig av hverandre). Respondentene blir stilt overfor hypotetiske situasjoner (alternativer) som de skal vurdere om det var virkelige situasjoner. Pearmain mfl.(1991:16f) sammenfatter kjennetegn ved SP-undersøkelser slik:

- SP er et redskap for markedsundersøkelser for å avdekke hvordan folk vil reagere på ulike hypotetiske reisesituasjoner.
- Disse situasjonene består av ulike kombinasjoner av egenskaper som er relevant for reisevalget.
- Forskeren konstruerer reisesituasjoner på en slik måte at effekten av hver faktor kan måles kvantitativt.
- Forskeren må forsikre seg om at respondenten forstår den hypotetiske situasjonen helt klart, at situasjonen oppfattes troverdig og realistisk, og at den relateres til respondentens erfaringer.
- De som blir valgt til å intervjues må være representative for den befolkningsgruppen som er i fokus for undersøkelsen.
- Svarene blir analysert slik at den relative vektleggingen av hver av egenskapene kan måles – dette forutsetter bruk av riktige statistiske modellerings-teknikker.
- Resultatene gir forskeren muligheten til å analysere, prioritere og måle effekten av ulike tiltak.

### 2.1 Ulike SP-metoder

De tre vanligste metodene for SP-undersøkelser er:

- Betinget verdsetting (*contingent valuation*)
- Samvalganalyse (*stated choice*)
- Likeverdspris (*transfer price*)

I tillegg har vi i denne studien benyttet metodene *Fleksibelt pakkevalg* og *Frisch-metoden*. Disse er beskrevet i kapittel 4.

Selv om alle disse tilnærmingene går under samlebetegnelsen *Stated Preference*, er det innenfor samferdselsforskningen vanligst å bruke SP som betegnelse for samvalganalyser. Den følgende beskrivelsen av de ulike metodene bygger i stor grad på Sælensminde (1995a).

*Betinget verdsetting.* Mer direkte spørsmål om betalingsvillighet går gjerne under fellesbetegnelsen *contingent valuation method* (CVM), som på norsk oversettes med betinget verdsetting. I denne gruppen finnes både direkte og mer indirekte spørsmålsformuleringer, som ”hva er det høyeste beløpet du er villig til å betale for X?” eller ”vil du godta en økt skatt på Y% hvis Z blir K% bedre?”. En fordel med denne metoden er at spørsmålene og valgene er enkle, og at analysen er forholdsvis problemfri. Betinget verdsetting har særlig vært brukt i forbindelse med verdsetting av ulike natur- og miljøfaktorer, men metoden har også vært benyttet til å finne betalingsvillighet for redusert risiko. På grunn av den direkte spørsmålsstillingen er det lett å forstå at metoden kan kritiseres for å være sårbar for taktisk svargivning. Hvis f.eks. respondentene frykter at undersøkelsen skyldes at myndighetene vurderer å øke billettprisene eller skattene, er det rimelig å anta at de svarer deretter.

*Samvalganalyse.* Samvalganalyse brukes særlig til markedsanalyser og analyser av trafikanters verdsetting av forbedringer i transporttilbudet. Også miljøgoder blir ofte verdsatt ved samvalganalyser. Ved samvalganalyser verdsettes flere goder i forhold til hverandre, og disse godene kan både være markedsgoder (f.eks. reisemiddel) og ikke-markedsgoder (f.eks. luftkvalitet). I undersøkelsene blir respondenter stilt overfor alternativer som består av en ”pakke” av egenskaper (f.eks. reisetid, pris, regularitet og komfort) ved en bestemt reise. En slik pakke vil bli referert til som et *alternativ*. Den vanligste typen samvalganalyse er *parvise valg* (*stated choice*), men *rangering* (*SP rank*) og *rating* (*SP rate*) er også brukt i ulike undersøkelser. Ved *parvise valg* velger respondenten den av to alternativer som har de mest attraktive egenskapene. Ved *rangering* skal et antall alternativer rangeres i rekkefølge etter hvor attraktive de er for respondenten. *Rating* innebærer å gi de enkelte alternativene poengsum/karakter på en skala. Ifølge Wardman (1998) har det i Storbritannia de siste 20 årene blitt en større og større andel av samvalgundersøkelser som bruker metoden parvise valg for tidsverdiundersøkelser. I dette prosjektet har vi konsentrert oss om parvise valg, og i det videre er det denne metoden vi henviser til når vi snakker om samvalg.

*Likeverdspris.* Denne metoden fremstår som en mellomting mellom betinget verdsetting og samvalganalyse. Spørsmålsstillingen dreier seg om betalingsvillighet for

goder og villighet til å akseptere kompensasjon for forverring. Likeverdprismetoden kan brukes til å verdsette egenskaper som respondenten kjøper i et virkelig marked, dvs. at spørsmålene eksplisitt refererer seg til respondentens nåværende beslutningssituasjon og valgmuligheter. En slik anvendelse av likeverdprismetoden vil i stor grad begrense hvilke goder en kan verdsette og i sterkere grad relatere verdsettingen til "bruksverdier". Som med betinget verdsetting er det klart at den direkte spørsmålsstillingen åpner for mulige taktiske svar.

## 2.2 Design av samvalganalyser

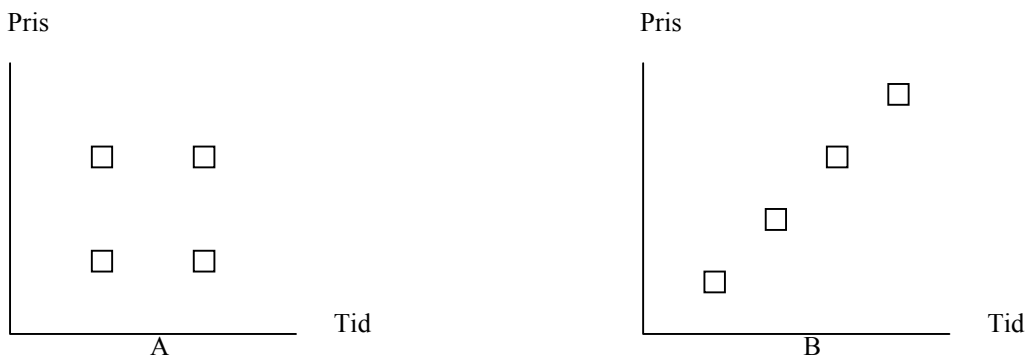
Ortogonal design er det mest vanlige opplegget for å variere egenskapsnivåene i spørsmålene. Ortogonalitet innebærer at variablene varierer uavhengig av hverandre. Med henvisning til regresjonsteori kan det vises at de statistiske egenskapene ved modellen forbedres, fordi den partielle effekten av hver egenskap lettere kan isoleres. Figur 2.1 illustrerer forskjellen mellom ortogonalt design, der variablene er ukorrelerte, og et design der variablene er korrelerte. I det første tilfellet varierer variablene uavhengig av hverandre, og i estimeringen av de individuelle effektene av endringer i tid og pris oppstår det ikke korrelasjonsproblemer.

I samvalganalyser der tre-fire egenskaper skal testes, blir det fort mange spørsmål hvis man bruker et fullt faktorisk design. Blant de vanligste måtene å redusere antallet alternativer er *fraksjonert design* (se f.eks. Olsson (1997:44) og *differansedesign*. Sistnevnte kan f.eks. redu-

sere de fire egenskapene 1. reisetid buss, 2. reisetid tog, 3. pris buss og 4. pris tog reduseres til to "differanser": differansene tid (tog *minus* buss) og pris (tog *minus* buss) i designet.

Bruk av computer ved gjennomføringen av intervjuene gjør det enklere å bruke adaptiv design. Den enkleste formen for adaptiv design er å relatere egenskapsnivåene til den reisen respondenten foretar i intervjuøyeblikket (evt. respondentens siste kollektivreise). Ved å innledningsvis stille spørsmål om egenskaper ved den reisen respondenten foretar, kan computerprogrammet variere egenskapene rundt disse. Dette kan i prinsippet gjøres ved alle typer datainnsamling. En mer avansert form for adaptiv design er når dataprogrammet fortløpende sjekker svarene respondentene gir og tilpasser egenskapsnivåene i de neste spørsmålene etter disse. Dette er en verdifull egenskap med computerbaserte SP-undersøkelser fordi det gir langt større grad av fleksibilitet, og man kan avdekke eventuelle ekstreme verdsettinger av spesielle egenskaper.

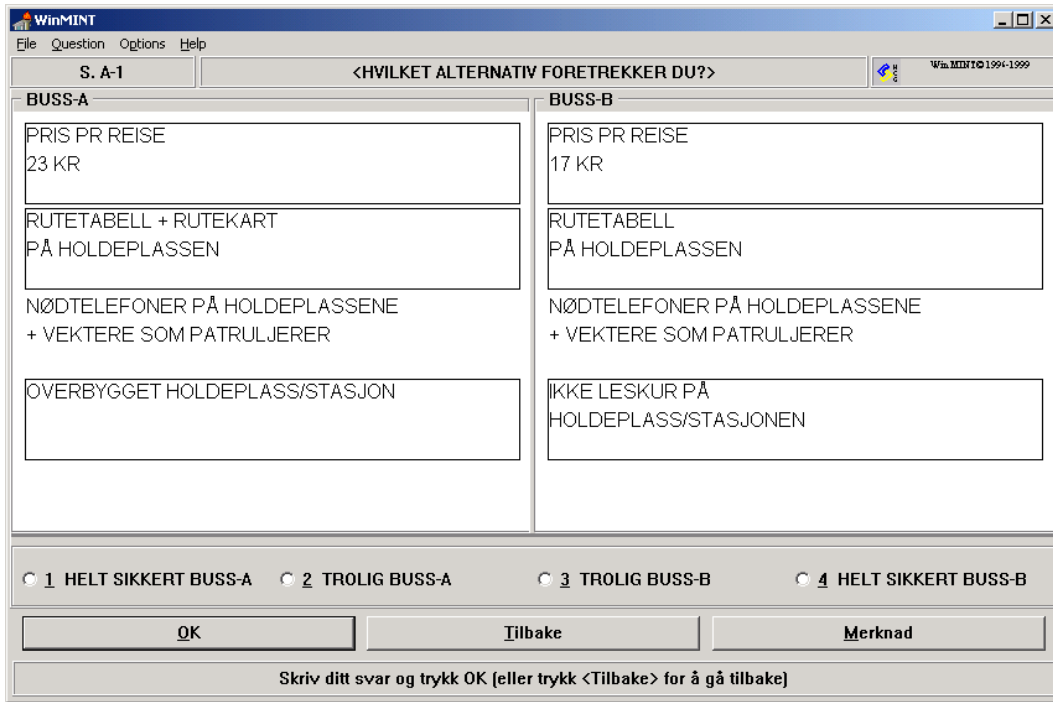
Figur 2.2 viser et eksempel på oppsettet for en samvalganalyse. Her ser vi at respondenten har valgt mellom "BUSS-A" (venstre side) og "BUSS-B" (høyre side). Reisealternativene er her beskrevet med de fire egenskapene (kvalitetsfaktorene) pris, informasjon, trygghet og holdeplassen. Nivåene på disse egenskapene varieres slik at det i hver av valgomgangene (som regel ni valgomganger) blir nye avveiinger.



©TØI 2001

Figur 2.1: Eksempler på ukorrelerte (A) og korrelerte (B) variabler. Eksemplene er tatt fra Olsson (1997:45)





©TØI 2001

Figur 2.2: Eksempel på skjermbilde av samvalgspørsmål

### 2.3 Analysen

Samvalganalysen blir brukt til å estimere parametrene i nyttefunksjonene for en logitmodell ved hjelp av *maximum likelihood*-estimering. Logitmodellen har formen:

$$P_i = \frac{e^{U_i}}{\sum_n e^{U_n}}$$

der

$P_i$  er sannsynligheten for å velge alternativ  $i$   
 $n$  er de alternativene det kan velges mellom  
 $U_n$  er nyttefunksjon for alternativ  $n$

Nyttefunksjonen kan ha flere former, men vanligvis blir en lineær additiv funksjon brukt, som f.eks.

$$U_n = \beta_0 + \beta_1(\text{Pris}) + \beta_2(\text{Reisetid}) + \mathcal{E}$$

Den relative verdsettingen kan dermed beregnes som forholdet mellom parametrene for reisetid og pris. I dette eksemplet er verdsettingen av marginal endring i reisetid (VoT) lik:

$$\text{VoT} = \frac{\partial U / \partial(\text{reisetid})}{\partial U / \partial(\text{pris})} = \beta_2 / \beta_1.$$

På samme måte kan alle egenskapene ved reisen gis en marginal (relativ) verdsetting.

Statistiske indikatorer for hvor god forklaringskraft modellen har, er:

- Log likelyhood
- $\rho^2$  (konstant)
- t-tester for parameterestimaten og for de relative verdsettingene

## 3 Vanlige feilkilder og skjevheter i SP-undersøkelser

Dette kapitlet vil legge vekt på tre sentrale problemstillinger innen SP-undersøkelser, nemlig:

1. Hva er de vesentligste kildene til feil i SP-analyser?
2. I hvilken grad, og hvordan, påvirker designet resultatet?
3. Hvilke metoder er utarbeidet for å påvise og redusere disse problemene?

Mange feilkilder i SP-undersøkelser har vært kjent lenge, og er utførlig behandlet i litteraturen. Som vi skal se, omfatter de i hovedsak statistiske svakheter, praktiske problemer og problemer knyttet til intervjusituasjonen. I dette kapitlet vil vi ikke gå inn på problemer knyttet til selve dataanalysen. Det følgende bygger på Fearnley (2000).

Den mest grunnleggende kritikken av SP-metoder er det faktum at det er en eksperimentell øvelse der respondenter blir stilt i en hypotetisk situasjon og må svare på hypotetiske spørsmål (jf. første punkt i Pearmain mfl. (1991) sin definisjon). Det er derfor flere grunner til å anta at det folk sier de vil gjøre og det de faktisk gjør, kan være to forskjellige ting. Sælensminde (1997) understreker dette problemet. Han rapporterte om forsøk med å minne respondentene på at de må forholde seg til sine budsjettbetingelser. Resultatene av den undersøkelsen viser, for å sette det på spissen, at respondentene forandret svarene sine da de ble bedt om å ta hensyn til virkeligheten. Dette underbygger i stor grad den kritikken av SP-metoder som går ut på at den hypotetiske spørsmålsstillingen forringer resultatenes troverdighet.

En viktig antakelse som også lett kan kritiseres, er at respondenten har forstått den hypotetiske situasjonen helt klart og at situasjonen oppfattes som troverdig. Sælensminde (1995b) illustrerer hvordan respondenten kan feiltolke spørsmålene. Da respondenten ble stilt overfor valg som involverte lengre reisetid enn den faktiske reisetiden på den virkelige reisen, ble det ikke oppgitt noen grunn til den forlengede reisetiden. Det var derfor sannsynlig at mange oppfattet lengre reisetid som *for-sinkelse*, som verdsettes mye høyere enn reisetid. Resultatet ble dermed en situasjon hvor beregningen av befolkningens verdsetting av reisetid ble altfor høy fordi spørsmålsstillingen var uklar.

Problemet med troverdighet varierer fra SP-analyse til SP-analyse. Alle alternativer i en SP-sekvens må virke troverdige for at respondentene skal kunne forstå dem som sannsynlige. Det kan lett være fristende å la troverdigheten til alternativene vike for å forbedre statistiske egenskaper eller forbedre balanseringen av egenskapsnivåene. Dette kan gjøre valgsituasjonen enda mer fremmed for respondenten og dermed være med på å kaste tvil over resultatene av undersøkelsen. Det er derfor vanlig at designet gjennomgås med "friske øyne" av en som ikke har sett seg blind på statistiske og andre betraktninger før (pilot)undersøkelsen settes ut i live. På denne måten kan usannsynlige og fornuftsstridige alternativer unngås.

Kravet om at forklaringsvariablene skal være uavhengige, har vært gjenstand for noe debatt i det siste tiåret. Wardman (1999a) lister noen grunner til å avvike kravet om ortogonalt design. Han påpeker at avvik fra ortogonalt design er *nødvendig* i tilfeller der:

- Det forårsaker at alternativene virker urealistiske.
- Man tar med variabler som bestemmes utenfor eksperimentet (f.eks. "som nå").
- Man kombinerer forskjellige design.

Videre kan avvik fra ortogonalt design være *ønskelig* i noen sammenhenger:

- Dekningsområdet for de relative avveiningene (ofte referert til som "strålediagram", se f.eks. Swanson mfl. 1993) kan forbedres ved å avvike ortogonalt design.
- Ortogonalitet, og dermed ingen korrelasjon, minimerer ikke variansen til de relative verdsettingene. Avvik fra ortogonalitet kan med andre ord forbedre egenskapene til modellene.

Watson mfl. (1996) har argumentert for dette siste punktet. De kritiserer den allmenne oppfatningen at ortogonalt design er det beste designet med hensyn til statistiske egenskaper. Denne ideen er ifølge dem hentet fra regresjonsteorien, som stiller beslektede krav til egenskapene ved forklaringsvariablene. Watson mfl. viser at kravet om ortogonalt design i flere tilfeller med fordel kan lempes, både ved beregning av relative verdsettinger

(som tidsverdier) og ved beregning av prognosemodeller. Avvik fra ortogonalt design kan i begge tilfellene forbedre de statistiske egenskapene til de estimerte parameterne. Den optimale korrelasjonen mellom forklaringsvariablene er altså ikke nødvendigvis 0.

### 3.1 Utvalgsstrategi

Utvalget av respondenter skal være mest mulig representativt for befolkningen eller for den bestemte brukergruppen man studerer. I den grad man mislykkes med å få et representativt utvalg vil dette skape skjevheter i data-materialet som igjen forårsaker gale anslag for befolkningens preferanser.

Valget av kontaktform kan ekskludere ulike grupper. I seg selv er ikke dette noe overraskende, men det er likevel viktig å vurdere i hvor stor grad det forårsaker skjevhet i resultatene. Intervjuer om bord på bussen ekskluderer ikke-brukere, f.eks. dem som er forhindret fra å bruke bussen pga. manglende lavgulvpåstigning, eller de som føler at det er for utrygt å bruke T-banen. Intervjuer med PC kan virke truende for dem som ikke behersker datamaskiner og dermed ekskludere denne gruppen fra undersøkelsen. Telefonintervjuer utelater bl.a. dem som ikke står oppført i katalogene, døde, hjemløse osv. Axhausen (1998) rapporterte f.eks. at valget av telefon som kontaktmedium førte til at kvinner og eldre var overrepresentert i utvalget. Utvalget måtte derfor vektas for å gjøre det representativt mht alder, kjønn og andel månedskortbrukere.

Respondenter som har erfaring fra egenskapene det spørres om, legger vanligvis en annen vekt på egenskapene enn dem som ikke har erfaring. Dette innebærer at et utvalg med *brukere* vil kunne ha andre preferanser enn *ikke-brukere*. For eksempel beskriver Sjöstrand (1999) at flybussreisende som ikke var vant til å måtte bytte buss, vurderte et bytte tre ganger verre enn dem som har erfaring med det. Derimot vurderte de *ventetiden ved bytte* langt lavere enn dem som bytter. Dette kan skyldes at reisende som ikke må bytte legger mindre vekt på tid, siden de er vant med å ha det komfortabelt hele veien.

Ortúzar og Willumsen (1996) anbefaler at det gjøres om lag 75-100 intervjuer for hvert markedssegment (aldersgruppe, bileiere, reiseformål osv) man er interessert i å undersøke, hvis man skal være sikret å få et godt utvalg. Norheim (1996) mener at antallet intervjuer som bør gjennomføres avhenger av hvilke egenskaper som undersøkes. Han skriver:

Våre tester viser at særlig de mer marginale kvalitetsfaktorene, eller de hvor det er store variasjoner i trafikantenes preferanser, er mest følsomme for

utvalgsstørrelsen. For disse faktorene kan utvalg under 100 personer gi stor usikkerhet i resultatene. De mer tunge kvalitetsfaktorene (reisetid, sitteplass mm) gir relativt stabile estimater helt ned til et utvalg på 30 personer. Som en hovedregel må en derfor være forsiktig med å dele opp utvalg under 100 personer, og alle delutvalg må tolkes med forsiktighet. Samtidig er det viktig å ha en klart definert målgruppe for utvalget, slik at en eventuell segmentering kan foretas gjennom rekruttering. (s 38)

Problemet med utvalget kan illustreres med funnene til Norheim (1996). I gjennomgangen av en undersøkelse av kollektivtrafikantenes preferanser i Oslo viste det seg at hele 12 % av respondentene mente at lavgulvbusser var den aller viktigste faktoren ved kollektivreisene. Imidlertid kom ikke dette til syne i de aggregerte analysene. Faktisk ble verdien av lavgulvbusser beregnet til kr -0,50 pr. tur for hele befolkningen, altså en negativ verdsetting. I undersøkelsen med lavgulvbusser ble 425 personer intervjuet. Blant disse var det 51 som la aller størst vekt på lavgulvbusser. I den grad disse personene tilhørte et segment (eldre, uføre, småbarnsforeldre eller lignende) ble det derfor gjort altfor få intervjuer med denne gruppen for å påvise deres spesielle preferanser.

En pragmatisk tilnærming til dette spesielle problemet ville være å beregne verdsettingen av lavgulvbusser blant f.eks. bevegelseshemmede, og justere den aggregerte verdsettingen i forhold til andelen bevegelseshemmede trafikanter totalt.

### 3.2 Datainnsamling

Det er ingen gitt metode for datainnsamling for de forskjellige typer SP-undersøkelser. De ulike SP-metodene stiller likevel ulike krav til presentasjonen av valg-situasjonen, slik at f.eks. samvalganalyser ikke kan gjøres over telefon. I den grad det er mulig med flere kontaktformer, er det de økonomiske rammene eller behovet for et representativt utvalg som bestemmer kontaktformen. Generelt kan vi si at:

- Det i varierende grad er mulig å tilpasse spørsmålene til respondentens aktuelle situasjon, såkalt adaptiv design<sup>1</sup>. Slik tilpasning vil f.eks. kunne foregå

<sup>1</sup> Adaptiv design har først og fremst vært brukt som begrep på en situasjon der respondentens første valg i sekvensen av samvalg brukes til å fastsette egenskapsnivåene for de senere valgene. Motivet for å bruke en slik adaptiv design er å raskere få tak i respondentens preferanser. Dette kan imidlertid skape problemer dersom respondenten pga. vanskelige valg og manglende opplæring velger inkonsistent (dvs. i strid med sine preferanser) i de første valgene. En slik form for adaptiv design bør en derfor være forsiktig med. En adaptiv design der respondentens valg tilpasses

fortløpende under intervjuet dersom respondentens svar registreres direkte i et PC-program. Både telefonintervju og intervju ved direkte møte kan i prinsippet administreres ved hjelp av PC.

- Spørreskjemaer er billig å produsere og distribuere og kan gi en grei presentasjon av alternativene. Problemet er at spørreskjemaer er lite fleksible og vanskelige å tilpasse til den reisendes situasjon. Det er vanskelig å rettlede respondentene hvis de har spørsmål utover veiledningen som følger med. Man kan heller ikke så lett forandre rekkefølgen på spørsmålene tilfeldig, slik det er ønskelig for å forbedre statistiske egenskaper (Bradley og Daly 1994) ved undersøkelsen.
- Computerbaserte intervjuer gjør det enkelt å endre rekkefølgen, og hver respondent kan få presentert et tilfeldig utvalg av spørsmålene. Datamaskinen kan sjekke svarene fortløpende, og eventuelt stille kontrollspørsmål. Siden respondenten taster inn svarene direkte på PC-en, er det lettere å unngå feilkoding. Videre tillater computere bruk av grafikk for å presentere alternativene, og de åpner for adaptiv design. Hovedproblemer med computerbaserte SP-undersøkelser er at de kan skremme bort folk som er uvant med bruk av PC og dermed forrykke utvalgsbalansen, og at det kan være relativt dyrt i form av innkjøp av maskiner og dataprogrammering.
- Telefonintervjuer er dyrere å gjennomføre, men til gjengjeld får man svarene direkte. Telefonundersøkelser kan være PC-baserte, slik at adaptiv design og randomisering er mulig. Det er også mulig å gi fortløpende veiledning.
- Intervjuer om bord kan være en hyggelig avveksling på lengre reiser, men stressende på korte reiser. Slike undersøkelser er relativt billige og greie å administrere, fordi svarprosenten er ofte høy og man når ut til mange på en gang (Sjöstrand 1999). Et problem er at man vet lite om dem man ikke når frem til.
- Hjemmeintervjuer er ressurskrevende, men kan til gjengjeld være både omfattende og detaljerte. Det er f.eks. mulig å stille utdypende spørsmål. Problemet er at intervjuerens nærvær kan påvirke hvordan respondentene svarer.
- "Hall-test" er en datainnsamling hvor respondentene blir invitert inn til et intervjulokale, ofte nær sentrale plasser/veier og hvor selve intervjuet tar relativt kort tid. Dette er en relativt "billig" metode, men kan ha det problemet at rekrutteringen blir skjev i forhold til

dem som har god tid, og det kan være problemer med å kontrollere frafallet.

Generelt sett må alle intervjumetoder ha gode rutiner for frafallsregistrering og en god dokumentasjon av hvem som var målgruppen for undersøkelsen.

### 3.3 Intervjusituasjonen

I en SP-undersøkelse er respondenten pr. definisjon satt i en hypotetisk situasjon, og svarene blir nødvendigvis også hypotetiske. Det er derfor en rekke ulike måter svarene kan bli påvirket på. Nedenfor er en liste med bare et utvalg av disse. Den bygger bl.a. på Wardman (1999a), Swanson (1998), Pearmain og Kroes (1990).

- Strategiske svar. Når respondentene tror de kan påvirke beslutninger ved å svare strategisk. Dette vil kunne oppstå hvis det går frem hva som er hensikten med undersøkelsen.
- Fokusering. Når det legges kunstig høy vekt på spesielle egenskaper, som sikkerhet. Leksikografiske svar er et beslektet problem som oppstår spesielt i valgsituasjoner der mange egenskaper varieres.
- Informasjon. Spørsmålene kan inneholde ufullstendig informasjon om egenskapene, som respondentene baserer valgene på. De egentlige preferansene kan derfor være annerledes.
- Rettferdiggjøring. Det er veldig typisk at respondentene forsvarer de valgene de allerede har tatt. For eksempel tror ofte bilister at det er dyrere å reise kollektivt enn det egentlig er, og kollektivtrafikanter at bilisme er dyrere og mer forurensende enn det i virkeligheten er.
- Pakkeeffekter (fokuseringseffekter). Når respondenten tillegger en av variablene i undersøkelsen mer innhold enn det som er ment, kalles det gjerne fokuseringseffekt. For eksempel kan "trafikkstøy" oppfattes som "alt negativt ved veitrafikk" og blir dermed gitt uforholdsmessig høy prioritet. Dette gjør at man ikke uten videre kan summere miljøkostnader som er beregnet i ulike SP-undersøkelser, fordi sannsynligheten for dobbelttelling er stor. (Wardman (1999b) bruker begrepet pakkeeffekt om et annet fenomen, nemlig når man forsøker å verdsette egenskaper som alle bidrar til å forbedre det samme aspektet ved reisen. De beregnede verdiene til hver enkelt av forbedringene summerer til noe helt annet (mindre) enn verdien til *pakken* av forbedringene. Wardman fant en "pakke-ratio" (pakke/summen av individuelle verdier) på 0,3–0,9. Interak-

---

den reiselengde og kostnad som han/hun hadde på den aktuelle reisen er imidlertid en form for tilpasning til den aktuelle konteksten som må anses som en styrke ved PC-administrerte SP-intervjuer.

sjoner, budsjettbetingelser og avtagende grensenytte er mulige forklaringer på fenomenet.)

- Spørsmålsformulering. Formuleringen av spørsmålene, eller rekkefølgen, kan lede respondenten i retning av en spesiell type svar.
- Middel. Ulike kostnader oppfattes forskjellig. For eksempel billettpris, veiprising og miljøavgifter blir gjerne møtt med ulik reaksjon, selv om kostnaden kan være den samme og pengene finansierer det samme prosjektet.
- Fortrolighet. Erfaring med egenskapene det spørres om, påvirker svarene. Dette er diskutert i avsnittet om utvalgsstrategi.
- Avrunding. I direkte betalingsvillighetsspørsmål er verdier som for eksempel slutter på 5 eller 0 overrepresentert (0, 5, 10, 15, 20 osv).
- Status quo. Man vet hva man har men ikke hva man får. Derfor ønsker ikke alle at det skal skje noen forandring.
- Innlæringseffekt. Den hypotetiske situasjonen er uvant for respondenten, og de første svarene kan derfor være avvikende som følge av at respondenten trenger tid til å venne seg til situasjonen. Både Norheim (1996) og Laitila og Westin (1997) har påvist en innlæringseffekt i undersøkelsene sine, der analysene viser tydelige tegn på at de første spørsmålene i uforholdsmessig stor grad avviker fra preferansene som går frem av de senere svarene. Norheim mener at den uvante valgsituasjonen gjør at det første spørsmålet i SP-undersøkelsen ofte er vanskelig. Hvis dette er et generelt problem, foreslår Laitila og Westin at det kan løses ved at det legges inn flere øvelsesspørsmål.
- Tretthet. For mange spørsmål utmatt respondenten og forårsaker feil i svarene. Dette problemet kan lett oppstå i SP-undersøkelser, hvor hver person ofte må svare på mange lignende spørsmål. Wardman (1999a) viser i en oversikt at trenden over tid er mot færre og færre spørsmål pr respondent. I 1980-årene var det gjennomsnittlig 12,5 parvise valg og 9,4 rankingalternativer. I midten av 1990-årene var tallene hhv. 10,1 og 9. Dette viser at man nå legger mer vekt på tretthetsproblemet i designet av undersøkelser.
- Små endringer. Disse kan ha lett for å bli oversett. Her gjelder det at egenskapsnivåene må være godt balanserte slik at respondenten gjør reelle avveininger.

Selv om denne listen er lang, kan den lett gjøres mye lengre. Det som er viktig her, er å skille mellom tilfeller hvor disse feilene oppstår tilfeldig og de tilfeller hvor det

skjer systematisk. Det er et mindre alvorlig problem hvis feilene opptrer usystematisk, fordi når dataene aggregeres, kan vi regne med at feilene oppveier hverandre. Samtidig vil det svekke estimatenes forklaringskraft. Hvis feilene derimot skjer systematisk, har vi å gjøre med et alvorlig, strukturelt problem. Hvis det har seg f.eks. sånn at respondentene i altfor stor grad tillegger én egenskap størst vekt (f.eks. prisen) vil dette gi en uforholdsmessig høy relativ verdsetting av denne egenskapen, og en tilsvarende lavere relativ verdsetting av de andre egenskapene. Dette er et eksempel på såkalte leksikografiske svar (se under).

### 3.4 Leksikografiske svar

Når respondenten sorterer alternativene etter én (eller flere) egenskaper gjennom hele samvalgsekvensen, kalles dette leksikografisk svargivning. Dette er et viktig problem, fordi forutsetningen for en samvalgundersøkelse er at respondentene foretar avveininger. Det er i hovedsak to forhold som kan forårsake leksikografiske svar (Widlert 1994).

For det første kan en egenskap være så viktig for respondenten at nivåene på alle de andre egenskapene blir uten betydning i forhold. Dette kan skyldes dårlig balanse mellom egenskapsnivåene i designet av SP-undersøkelsen, men det kan også hende at den ene egenskapen er det eneste som teller for respondenten (ekstreme preferanser). Swanson (1998) understreker at det er viktig å supplere resultatene fra en SP-undersøkelse med informasjon om nivåene på egenskapene som er brukt. Hvis nivåene virker ubalanserte kan det kaste tvil over resultatene. I tilfellene der verdsettingen av den ene egenskapen er veldig høy, kan det være problematisk å beregne den faktiske verdsettingen til respondenten.

For det andre kan det være at respondentene finner intervjusituasjonen så vanskelig at de forenkler beslutningsrutinene sine og konsentrerer seg om en (eller noen) egenskaper. Når dette er tilfelle, vil denne egenskapen få en kunstig høy prioritet, og betydningen av andre egenskaper blir tilsvarende undervurdert. Ifølge Ortúzar og Willumsen (1996) oppnår man best resultat hvis respondenten vurderer endringer i opp til tre egenskaper samtidig, men selv det kan bli for komplekst i noen sammenhenger.

Motivasjonen for å svare leksikografisk har derfor stor betydning for verdsetting av egenskapene. I det første tilfellet risikerer man å undervurdere betydningen av egenskapen det sorteres etter, og i det andre tilfellet overvurderes denne egenskapen. Det er derfor vesentlig å lage et design for SP-undersøkelsen slik at flest mulig av

respondentene faktisk foretar avveininger der alle egenskapene tas i betraktning.

Betydningen av leksikografiske svar kan variere. Norheim (1996) fant at leksikografiske svar påvirker resultatene. Men ved å ekskludere de leksikografiske svarene forble de tunge kvalitetsfaktorene i stor grad uendret, mens de mer marginale faktorene fikk en sterkere vektlegging. Dette skyldes hovedsakelig at leksikografiske svar oftest fokuserer på prisen, som er en tung kvalitetsfaktor.

Sælensminde (1997) påpekte at andelen leksikografiske svar faller med antall spørsmål respondenten svarer på. Dette stemmer intuitivt, fordi jo færre spørsmål som skal besvares jo større er sjansen for at svarene tilfeldigvis forekommer leksikografiske. Andelen leksikografiske svar må derfor tolkes i forhold til den "naturlige" (forventede, tilfeldige) andelen leksikografiske svar.

PC-baserte, adaptive SP-teknikker kan lettere eliminere problemet med leksikografiske svar ved at nivået på faktorene avhenger av tidligere svar (Pearmain and Kroes 1990). Man kan dermed sirkle inn respondentens avveininger selv om denne har ekstreme preferanser.

### 3.5 Beskrivelse av egenskapene

I SP-undersøkelser ønsker man å finne frem til en relativ verdsetting av ulike egenskaper ved reisen. Disse kan grovt skiller inn i "tunge" kvalitetsfaktorer som pris, reisetid og bytte, og andre, "mykere" faktorer som informasjon, trykghetsfølelse, reisekomfort osv. De tunge kvalitetsfaktorene har respondentene som regel god kjennskap til, og kan lett forholde seg til. De mykere faktorene er gjerne vanskeligere å forholde seg til av flere grunner, og kan derfor være vanskelig å vurdere for respondenten.

Nelson og Towris (1995) understreker hvor viktig det er med presentasjonen av egenskapene i en SP-undersøkelse: Det er viktig å forsikre seg om at respondenten har den samme forståelsen av den hypotetiske situasjonen som forskeren har, for hvis ikke, kan situasjonen lett bli at respondenten gir svar som ikke bygger på den informasjonen som er gitt i intervjuet.

Sælensminde og Hammer (1993) spurte respondentene i hvor stor grad de hadde hatt problemer med å forstå betydningen av ulike faktorer. De skriver at "Bare 16,7 % av respondentene har rapportert at de synes det var vanskelig å forstå betydningen av en eller flere av faktorene som inngikk i spillene." (s. 50). Blant disse faktorene var CO<sub>2</sub>-utslipp (drivhuseffekt) og drivstoffkostnad/billettpris de egenskapene som hyppigst forårsaket problemer. Det siste er urovekkende, fordi den

relative verdsettingen av en egenskap er et forholdstall der parameteren for pris/kostnad inngår i nevneren.

Wardman (1999a) nevner noen grunner til at beskrivelsen av egenskaper og egenskapsnivåer kan by på problemer. Dette er kun noen utvalgte eksempler på hvor mange problemer som kan oppstå i en spørreundersøkelssituasjon:

- Meningsløs målestokk og størrelser som det er vanskelig å forholde seg til. I mange tilfeller der det er mulig å kvantifisere egenskapene, men folk fleste kan ikke relatere seg til måleenheten. Eksempler kan være støyendring i decibel, antall dekar skog bevart, antall fugleliv reddet, endring i ulykkesrisiko, cm benplass, forurensing i  $\mu\text{g}/\text{m}^3$ , osv. I de aller fleste tilfeller er det bedre å bruke sammenligninger ("som nå", "som i Kirkeveien") enn å risikere at egenskapene misforstås ved bruk av uforståelige målestokker.
- Nelson og Towris (1995) beskriver bl.a. problemet de hadde med å beskrive forskjeller i regularitet. For en ikke innvidd kan man ikke presentere et reisealternativ der reisetiden er 60 minutter og standardavviket er 15 minutter. De konkluderer med at det kreves nøye planlegging og pilotundersøkelser for å finne frem til en forståelig måte å beskrive slike egenskaper på.
- Ingen målestokk. Dette er kvalitative aspekter ved reisen som ikke kan kvantifiseres. Eksempler er komfort, design, dekor, atmosfære (stemning/hygge), setekvalitet, ventilasjon etc. Igjen er den beste tilnærmingen å sammenligne egenskapene med andre kjente situasjoner ("som på turbussene"). Ofte blir også illustrasjoner brukt for å beskrive egenskapen, men som vi skal komme tilbake til, er ikke dette helt uproblematisk heller.
- Egenskapen eksisterer ennå ikke. Dette er ting som ennå ikke er introdusert på (det lokale) markedet, som sanntidsinformasjon, kaffeservering osv.

Sælensminde og Hammer (1993) påviste hvordan verdsettingen av miljøgoder i sterk grad påvirkes av informasjonen som blir gitt til respondenten under intervjuet. De fant en sterk tendens til at respondenter oppgir høyere betalingsvilje for miljøegenskaper når det gis tilleggsinformasjon om dem.

Wardman (1999b) skriver at når respondenten har begrenset kunnskap om egenskapene som testes i SP-undersøkelsen, er presentasjonen avgjørende. Alle undersøkelser bør rapportere hvordan egenskapene er presentert (foto, verbal beskrivelse osv.) fordi dette kan ha avgjørende betydning for hvordan de verdsettes.

### 3.6 Eksempler på testing av ulike design

Med design mener vi her alt som har med utarbeidelsen og gjennomføringen av SP-undersøkelsen. Dette inkluderer valg av SP-metode (betinget verdsetting, likeverdspris, SP rank, SP rate, SP choice), utvalgsmetode, metode for innhenting av data, valg av egenskaper og nivåer osv. En såpass vid definisjon av design åpner for en bred gjennomgang av problemstillingene og kan belyse ulike aspekter ved forholdet mellom design og resultat.

Swanson (1998) har hevdet at designet påvirker resultatet fordi folk går ikke rundt med en fast nyttefunksjon. Det er ifølge ham en myte at folk besitter en fastlagt preferansestruktur, og at man alltid velger likt ut fra en gitt preferanseramme. Preferanser varierer med omstendigheter, tid, osv, og man skal derfor ikke la seg overraske hvis man observerer at preferanser varierer. Swanson viser i et eksempel hvordan resultatene avhenger av designet. Han beskriver et tidlig forsøk (fra 1971) der to utvalg er blitt stilt overfor nøyaktig de samme valgsituasjon, bare med forskjellig spørsmålsstilling. Resultatene viste seg å avhenge fullstendig av spørsmålsstillingen.

I samvalganalyser er spørsmålsstillingen generelt ikke så ledende som f.eks. i betinget verdsetting, men det er klart at også her avhenger kvaliteten på svarene og analysene i stor grad av kvaliteten på spørsmålene.

Denne seksjonen vil konsentrere seg om arbeider der forholdet mellom design og resultat er utforsket, og som har gjennomført tester og sammenligninger av ulike design. Tre viktige forskningsprosjekter der dette er blitt studert, er Staffan Widlert (1994), Sjöstrand (1999) og Laitila og Westin (1997). Disse vil bli utdypet nedenfor. I tillegg vil resultater fra bl.a. forskningsaktiviteten på Transportøkonomisk institutt bli brukt til å belyse og komplementere de ovennevnte arbeidene.

#### 3.6.1 Designet påvirker resultatet: Staffan Widlert

Blant testingene av forskjellige SP-design som er rapportert, er Staffan Widlerts (1994) omfattende undersøkelse av forholdet mellom design og resultat, er det oftest omtalte. Formålet med eksperimentet til Widlert var å belyse forholdet mellom design og resultat i samvalg-analyser:

1. Sammenligne parvise valg på computer med rangering av kort
2. Sammenligne rangering av kort med rating av alternativer
3. Undersøke betydningen av å tilpasse egenskapsnivåene til respondentens erfaringsgrunnlag

4. Undersøke betydningen av ulike antall alternativer
5. Undersøke betydningen av ulike prisnivåer
6. Sammenligne absolutte egenskapsnivåer med relative
7. Undersøke betydningen av intervjuerens nærvær

Widlert gjennomførte 15 ulike typer intervjuer med forskjellig design og datainnhentingsmetoder. Hver av øvelsene omfattet de samme fire egenskapene (pris, tid, frekvens og ulike typer togsett) med 3–4 nivåer. Egenskapene tid, pris og frekvens ble beskrevet med ord, mens togsettene ble beskrevet med bilder og ord. Intervjuene foregikk på togstrekningene Stockholm–Sundsvall og Stockholm–Göteborg og ble gjennomført av studenter. I alt 5700 intervjuer ble gjennomført.

Som et eksempel på resultatene av de ulike SP-designene er forskjellene i tidsverdier ganske illustrerende. Se Tabell 4.1. De varierer med en faktor på hele 4 mellom laveste og høyeste verdi. Dette til tross for at spørsmålene er nærmest identiske, og at utvalgene er forholdsvis like. Widlert tilskriver mye av forskjellene til hvordan respondentene har reagert på de ulike typene design.

I første rekke konsentrerte Widlert seg om å identifisere sammenhengen mellom design og andelen leksikografiske svar. Leksikografiske svar er, som vi har sett, opphav til viktige metodeproblemer, spesielt fordi man ikke kan vite om egenskapen overvurderes eller undervurderes av respondenten som svarer leksikografisk. Tabell 3.1 er en oppsummering av de viktigste funnene til Widlert.

Det går relativt klart frem av tabellen at ranking-øvelsene ga de største andelen leksikografiske svar, med spill nr. 10 med 59% leksikografiske svar som det tydeligste eksemplet. Den høye andelen i spill nr. 10 skyldes at det også ble testet for virkningen av uforholdsmessige store prisvariasjoner. SP rating ga den laveste andelen leksikografiske svar. Vi ser også hvordan høye andeler svar sortert etter pris i stor grad assosierer med lave estimerte tidsverdier, og omvendt.

Ved sammenligning av spill nr. 1 og 2, der pris og reisetid ble presentert som hhv. relative endringer og i faktiske størrelser (adaptiv), og som kun prosentvise endringer (ikke-adaptiv), er det indikasjoner på at respondenter legger mer vekt på prisen i spill som ikke er adaptive. I spill 2 får respondenten mindre informasjon om egenskapsnivåene, og det er derfor trolig at prisen tillegges større vekt jo mindre informasjon det gis om egenskapsnivåene. I det adaptive spillet er antallet leksikografiske svar som er sortert etter prisen lavere enn for det ikke-adaptive spillet, og verdsettingen av tid blir derfor langt høyere.

Tabell 3.1: Andel (%) leksikografiske svar i de forskjellige spillene sortert etter egenskapene pris, tid og annet, samt beregnede tidsverdier i de forskjellige spillene. Kilde: Widlert 1994

Spill	Pris	Tid	Annet	Total	Tidsverdi <sup>2</sup>
Basis: Computer <sup>1</sup> , adaptiv	7	10	5	22	0,78
Computer, ikke-adaptiv	15	6	1	22	0,40
Rank, 9 kort, med info om hva de relative endringene vil si for en typisk reise	35	4	0	39	0,32
Som 3, men kun info om relative endringer	38	4	0	42	0,30
Rank, 5 av 16 kort	45			613	0,36
Rank, 9 av 16 kort	30			343	0,27
Rank, 16 av 16 kort	27			303	0,25
Rating, 9 alternativer 0–100 poeng	6	1	0	7	0,58
Absolutte endringer	27	7	1	35	0,68
Rank, store endringer i pris	59	1	1	61	0,51
Rank, hovedsakelig prisreduksjoner	26	8	0	34	0,32
Rank, kun prisøkning	47	3	0	50	0,26
Computer, store prisendringer	35	3	4	42	0,83
Laptop computer, intervjuer til stede	n/a	n/a	n/a	n/a	0,91
Computer, ingen intervjuer til stede	n/a	n/a	n/a	n/a	0,85

<sup>1</sup> Med computer menes små Poquetmaskiner uten opplyst skjerm

<sup>2</sup> Tidsverdi regnet som prosent endring i pris / prosent endring i tid

©TØI 2001

Spill 3 ga under halvparten så lave tidsverdier som spill 1. Dette til tross for at de samme egenskapene og egenkapsnivåene er brukt i begge spillene. Den høyere andelen leksikografiske svar, som sorterer etter prisen i rank-øvelsen, forklarer mye av forskjellene i tidsverdiene. Ifølge Widlert er det opplagt at respondentene har brukt forenklete beslutningsrutiner når de er blitt bedt om å rangere kort.

Resultatene i spill 3 og 4 ga ingen indikasjoner om at resultatene av rankingspillene endrer seg når det gis mer informasjon om egenkapsnivåene. I spill 3 ble de relative endringene forklart som prosentvis endring samtidig som det ble informert om hva det faktisk ville si for størrelsen på egenkapsnivåene. I spill 4 ble det bare informert om prosentvis endring i egenkapsnivåene. Andelen leksikografiske svar og de beregnede tidsverdiene er bare marginalt forskjellige i de to spillene.

Andelen leksikografiske svar faller med antallet parvise spørsmål. Dette er illustrert i spillene 5, 6 og 7. Grunnen i for så vidt intuitiv, som nevnt over, at jo færre spørsmål som stilles, jo større er sjansen for at svarene (tilfeldigvis) forekommer å være sortert etter én egenskap. Widlert undersøkte hvordan andelen leksikografiske svar ble endret når han trakk ut ett og ett av respondentenes svar tilfeldig. Det viste seg at i spill 7, som hadde lavest andel leksikografiske svar, ville andelen faktisk være høyere enn i spill 5 og 6 hvis respondentene

bare hadde rangert bare hhv. 5 og 9 tilfeldig utvalgte kort. Dette kan forklare hvorfor de estimerte tidsverdiene *faller* når andelen leksikografiske svar *øker*.

Spillene 5–7 viste også at andelen leksikografiske svar som sorteres etter pris øker når antallet alternativer øker. Mens 74% av de leksikografiske svarene i spill 5 er sortert etter prisen, er så mange som 90% sortert etter prisen i spill 7.

Spill 8, rating-øvelsen ga veldig få leksikografiske svar. Her skulle respondentene karakterisere 9 ulike alternativer på en skala fra 0–100. Widlert konstaterer at rating-øvelsen ikke inviterer til forenklete beslutningsrutiner i samme grad som ranking-øvelsene gjør det.

I spill 9 ble det eksperimentert med å endre egenkapsnivåene med *absolutte* størrelser i stedet for relative endringer som er brukt i de andre spillene. Uansett hva respondentene betalte for reisene sine, ble de forespeilet reisekostnader som var -20, +20 og +40 SEK. For alle som foretok kortere/billigere reiser, ville disse prisendringene være tungtveiende. Derfor er det sannsynlig at de som svarer leksikografisk gjør det fordi prisegenskapen er dominerende, og ikke fordi de bruker forenklete beslutningsrutiner. Widlert fant imidlertid ingen indikasjoner på dette. I stedet konkluderer han med at kan være lettere for respondenter å forstå absolutte endringer enn relative endringer.



Det er generelt viktig å balansere egenskapsnivåene i forhold til hverandre i SP-spill. Dette fordi selve grunn-tanken i SP er at respondenten skal gjøre avveininger mellom alternativene, og derfor må endringene i egenskapsnivåene stå i forhold til hverandre. I spill 10, 11 og 12 ble effekten av ubalanserte egenskapsnivåer testet. En egenskap, prisen, ble variert på tre ulike måter i de tre spillene, nemlig med hhv. store prisvariasjoner, hovedsakelig *prisreduksjoner* og med *prisøkninger*.

Som det fremgår av Tabell 3.1 er spill 10 det spillet som ga den høyeste andelen leksikografiske svar etter prisen. Store prisvariasjoner gir med andre ord direkte utslag i leksikografiske svar. Dette skyldes ikke, ifølge Widlert, at respondentene forenkler beslutningsrutinene ved å fokusere på priselementet, men rett og slett at prisvariabelen blir den dominerende faktoren.

Selv om de estimerte tidsverdiene i spillene 11 og 12 ligger nokså nær hverandre, fant Widlert at andelen leksikografiske svar er langt høyere når respondentene kun blir gitt alternativer med *prisøkninger* (spill 12) enn når de blir gitt alternativer med hovedsakelig *prisreduksjoner* (spill 11).

Spill 13 tilsvarende spill 1, men har langt større variasjoner i prisen. Andelen leksikografiske svar er langt større, som forventet, men de estimerte tidsverdiene ligger nær hverandre. Widlert konkluderer derfor at selv om designet kan være skjevt, gir computerbaserte SP-undersøkelser mer stabile resultater.

I de to siste spillene ble det undersøkt om intervjuerens tilstedeværelse hadde noen betydning. Resultatene i spillene er forholdsvis like, og ingen av de estimerte parametrene i modellene er signifikant forskjellige fra hverandre. Tidsverdiene ligger heller ikke langt fra hverandre. Dette indikerer derfor at man gjerne kan la respondentene gjennomføre spillet på egenhånd. På denne måten kan det spares penger på datainnsamlingen.

Resultatene av Widlerts undersøkelser og sammenligninger viser at designet i stor grad påvirker resultatene. Dette skyldes i første rekke at respondentene benytter forenklete beslutningsrutiner når situasjonen blir vanskelig å håndtere. En annen viktig årsak til skjevheter i spillene, er manglende adaptiv design. Ifølge Widlert skyldtes mange av feilene i de ikke-adaptive spillene at egenskapsnivåene ikke sto i forhold til respondentenes faktiske situasjon. Derfor er computerbaserte, adaptive spill langt å foretrekke fremfor manuelle, ikke-adaptive spill.

### 3.6.2 Flere faktorer påvirker resultatet: Laitila og Westin

Laitila og Westin (1997) presenterer resultatene fra et prosjekt som tar for seg forholdet mellom SP-designet og analysen av egenskapenes betydning. Spesielt ser de på effekten av antallet egenskapsnivåer i designet, antallet spørsmål og på egenskaper ved respondentene, som alder, yrkesstatus og erfaringsbakgrunn.

Tidligere undersøkelser som Laitila har gjort indikerer at man generelt kan si at den relative vekten for en egenskap øker når antallet egenskapsnivåer øker. Laitila har vist at dette til dels kan forklares ved at når antallet egenskapsnivåer øker, forringes det statistiske presisjonsnivået i de relative verdsettingene. Ved å forbedre den statistiske modellen som analysen bygger på, kan imidlertid effekten av antallet egenskapsnivåer reduseres betraktelig.

I en konkret SP-undersøkelse gjennomgår Laitila og Westin resultatene fra en studie av togreisendes holdninger og relative verdsetting av pris, plasstype (sovevogn eller liggevogn) og reisetid. Undersøkelsen har tre hovedmål, nemlig å analysere betydningen av forskjellige rating-skalaer, å finne ut om passasjerer med ulike karakteristika hadde ulike preferanser og å analysere hvordan ulike egenskaper vurderes av de reisende og hva dette vil si i praksis.

I alt seks ulike SP-spill ble designet, med ulike rating-skalaer og ulike antall egenskapsnivåer. De tre forskjellige rating-skalaene var henholdsvis fra 1 til 9, fra -2 til +2 og fra -4 til +4, der de høyeste verdiene er de beste vurderingene. Valget av disse typene skalaer bygger på indikasjoner fra andre forskningsarbeider om at respondenter gjerne forholder seg annerledes til sentrerte (med 0 som 'verken/eller'-vurdering) enn til skalaer som starter fra 0 eller 1 og går oppover. I prinsippet er de to skalaene fra 1 til 9 og fra -4 til +4 identiske, idet begge har 9 grader. For å teste effekten av ulike antall egenskapsnivåer ble det designet to forskjellige spill der hhv. prisen og reisetiden hadde flere nivåer. I alle spillene oppga respondenten kjønn, alder og yrkesstatus.

En modell ble bygget for å undersøke hvordan de sosioøkonomiske egenskapene og de ulike designene påvirket koeffisientene i nyttefunksjonen. Det går klart frem av resultatene at betydningen av de ulike egenskapene varierer med sysselsetting og alder. For eksempel ble en reise på sovevogn istedenfor i liggevogn verdsett til SEK 45,10 for en yrkesaktiv 20-åring, og til SEK 180,14 for en yrkesaktiv 50-åring. Det viste seg også at yrkesaktive setter større pris på plasstypen og reisetiden enn studenter og vernepliktige gjør.

Det er sterke indikasjoner på at antallet nivåer av en egenskap som inngår i spillet påvirker denne egenskapens relative verdsetting. I et eksempel illustreres hvordan

forskjellene i tidsverdiene påvirkes av antallet nivåer av egenskapene reisetid og pris. I spillet med tre nivåer for reisetid og to nivåer for pris ligger de beregnede tidsverdiene langt høyere enn i spillet der reisetiden bare hadde to nivåer og prisen hadde tre. Siden tidsverdien beregnes som forholdet mellom  $\beta_{tid}$  og  $\beta_{pris}$  i nyttefunksjonen, kan dette indikere at jo flere nivåer en egenskap presenteres med, jo større vekt vil denne egenskapen tillegges. Dette resultatet støtter hypotesen som er fremkommet i andre forskningsarbeider.

Valg av responsskala gir også utslag i bedømmelsen av egenskapene. I forhold til 0–9 skalaen, fant Laitila og Westin langt høyere betydning av egenskapene når de brukte skalaen fra -2 til +2 i spillene. Skalaen fra -4 til +4 ga motsatt effekt. I skalaen fra -2 til +2 lå respondentenes vurdering av alternativene i større grad i den øvre halvdel av responsskalaen, og spillet med 5 grader ga lavere tidsverdier enn 9-gradspillene. Alt i alt ga skalaen fra -4 til +4 høyere tidsverdier, og skalaen fra -2 til +2 lavere tidsverdier enn 0–9-skalaen.

Laitila og Westin konkluderer med at de har funnet indikasjoner på at det fins mange ukjente faktorer i SP-undersøkelser som forskeren ikke har kontroll over, og som i vesentlig grad påvirker resultatene. Det er derfor fortsatt viktig å forske på disse faktorene for å videreutvikle SP-metoden til et mer troverdig planleggingsverktøy. Spesielt påpeker de at det er mangelfull kunnskap rundt de psykologiske aspektene ved undersøkelsen, f.eks. hvordan respondentene påvirkes av ulike design. Resultatene til Laitila og Westin peker i retning av at både valget av responsskala og antallet egenskapsnivåer vil påvirke respondentenes svar. I tillegg påpeker de at det er behov for å studere nærmere hvordan de mange gjentatte spørsmålene påvirker respondentene.

### 3.6.3 Samvalganalyser pr brev: Sjöstrand

Sjöstrand (1999) studerte i hvilken grad SP-undersøkelser pr brev kan benyttes for å estimere busspassasjerers verdsetting av kvalitetsfaktorer. Et mål var å undersøke om slike undersøkelser kan gjøres via brev, i motsetning til telefon- eller hjemmeintervjuer, som ifølge Sjöstrand har vært vanlig tidligere, og i så fall hva slags type SP-design som er best egnet. Et annet mål var å undersøke i hvilken grad respondentenes erfaring har betydning ved verdsetting av kvalitetsfaktorene.

Til det første delen av undersøkelsen designet Sjöstrand 6 ulike SP-design der de samme faktorene og de samme endringene i nivåene inngikk. Disse var:

*To ulike spill med parvise valg:*

1. Seks parvise valg på ett ark
2. Seks parvise valg på hvert sitt kort

*To forskjellige rating-spill:*

3. Seks alternativer på ugradert skala fra verst til best
4. Seks alternativer på kontinuerlig skala med 0 og 100 som ekstremverdier

*To ulike rankingspill:*

5. Rangere 6 klistremerker som hver representerer ett alternativ
6. Rangere seks kort som festes sammen

Fordi det er viktig å bruke de reisendes egne erfaringer til grunn for de egenskapsnivåene som presenteres, ble respondentene rekruttert ved at de ble kontaktet på holdplasser. De ble da stilt noen kontrollspørsmål for å sikre at utvalget var i målgruppen og spurt om de ville være med på undersøkelsen, som de ville få tilsendt senere. De som var villige til å delta ble spurt om egenskaper ved bussturen de tok. Spørsmålene i spillene kunne dermed tilpasses de faktiske egenskapene ved respondentenes reiser.

Spillene ble vurdert ut fra flere kriterier. Sjöstrand skiller mellom *ekstern* og *intern* validitet. Førstnevnte viser til anvendbarheten av resultatene, som hvor godt resultatene er representative for befolkningen, andelen leksikografiske svar, osv, og sistnevnte viser til signifikansnivå og statistiske presisjon. Det viste seg at de ulike spillene hadde til dels svært ulike egenskaper og presisjonsnivå.

Med hensyn til responsen, kom spillene 1 og 2 best ut med hhv. 90 og 88 i svarprosent. Øvelsene 3 og 4 ga lavest svarprosent med 71–74. Og blant svarene som kom inn var andelen brukbare svar også lavest i øvelse 3 og 4. Det virker derfor som om rating-øvelser gir det mest pålitelige utvalget, idet både frafallet og andelen forkastede svar er relativt lave. I så måte er det sannsynlig at resultatene fra spillene med parvise valg representerer befolkningen best.

Alle spillene ble testet for leksikografiske svar. Her viste det seg at spillene 3 og 4 kom bedre ut enn de andre, med nær 0% leksikografiske svar sortert etter prisen. De andre typene spill hadde opptil 10% leksikografiske svar sortert etter pris, noe som også i og for seg er relativt lavt. Ifølge Sjöstrand er det vanskeligere å svare leksikografisk i rating-spill fordi man bedømmer ett og ett alternativ i stedet for å sammenligne dem som man gjør i parvise valg og ranking. Det er interessant at i ingen av spillene var det prisen som dominerte de leksikografiske svarene.

De fleste leksikografiske svarene var derimot sortert etter gangtid, bytte eller frekvens.

T-testene i modellene viste at parvise valg gjennomgående gir sikrere estimater for parametrene. Med unntak av egenskapene informasjon og lavgulv var alle parametrene (pris, reisetid, gangtid, frekvens, bytte, byttetid, sitteplass) signifikante på 10% nivå. Både ranking- og rating-spillene ga mer varierende t-verdier.

Ifølge Sjöstrand vil store parametre i absolutt forstand forklare betydningen av egenskapen i større grad enn små parametre. Sjöstrand indekserte parametrene i forhold til de som ble beregnet i spill 1, og kunne på den måten lett påvise hvilke typer SP-undersøkelser som gir de største parametrene. Hun kaller forholdet mellom de beregnede parametrene og indeksparametrene for ”skalafaktor”. De parvise valgene hadde de høyeste skalafaktorene og ga med andre ord de største parametrene i nyttefunksjonene.

Sjöstrand bruker standardfeilen for verdsettingen av egenskapene som en indikasjon på hvor sikre beregningene er. Dette avhenger ikke bare av hvor godt designet er, men også av faktorer som antallet svar og hvor likt respondentene svarer. Det er derfor ikke noen direkte vurdering av designet, men kan være en indikasjon på hvor godt de ulike spillene gjør det i forhold til hverandre.

For hvert av spillene beregnet Sjöstrand standardfeilen for verdsettingen av alle egenskapene. For tidsverdier (reise-, vente- og gåtid), ga ranking-undersøkelsene de største standardfeilene, altså de minst sikre beregningene. For verdsetting av kvalitetsegenskaper (info, sitteplass og lavgulv) kom ranking derimot best ut, mens parvise valg og rating hadde høyere standardfeil.

En telefonoppringing til et utvalg av respondentene etter at de hadde svart på brevundersøkelsen ga viktig informasjon om hvordan respondentene hadde opplevd spørresituasjonen. Mens de som hadde blitt gitt parvise valg i overveiende grad var positive til undersøkelsen og rapporterte få problemer, var det ingen positive kommentarer til spill 3 (rating best–verst). I spillene 4, 5 og 6 var respondentene delt nokså likt mellom positive og negative kommentarer.

Sjöstrand konkluderer med at brevbaserte SP-undersøkelser bør designes som parvise valg på ett papir, altså som i spill 1. Dette designet ga høy svarfrekvens, og scoret godt på statistiske tester. Rating på en ugradert skala fra verst til best fungerte veldig dårlig, mens skalaen fra 0 – 100 fungerte relativt bra og kan benyttes i noen sammenhenger. SP rank bør ikke brukes. Undersøkelsene og metodetestene viser med andre ord at brevbaserte SP-undersøkelser kan gi relativt sikre resultater, men at designet må velges med omhu.

### 3.6.4 Andre metodetester

Flere arbeider ved TØI beskriver tester av SP-design. Disse testene tilsvarer stort sett de testene som er beskrevet over og vil derfor ikke bli behandlet om igjen her.

Sælensminde (1997) studerte virkningen av å gi respondentene en budsjettpåminnelse når de skulle svare på spørsmålene. Han fant en relativt entydig indikasjon på at verdsettingen av egenskapene faller når respondentene blir bedt om å ta hensyn til sine totale budsjettbeskränkninger eller til sine transportbudsjett. Men disse forskjellene var ikke signifikante. En regresjonsmodell viste imidlertid signifikant utslag på betalingsvilligheten av en budsjettpåminnelse for en pakke bestående av kun to egenskaper. Dersom pakken besto av tre faktorer, ga en budsjettpåminnelse ikke signifikant utslag på betalingsvilligheten. Sælensminde (1998) konkluderer med at respondentene har lettere for å ta hensyn til budsjettbetingelsen når de får en ”komplett” pakke å verdsette.

I en samvalganalyse i Oslo testet Norheim (1996) en rekke måter designet kan påvirke resultatet. Blant dem er:

- Betydningen av nivåene på egenskapene. Norheim fant at spill med større differanse i prisnivåene ikke har gitt noen entydig økning av leksikografiske svar. Ulike variasjoner i prisnivåene har heller ikke påvirket relative verdsettinger i noen signifikant grad.
- Betydningen av utvalgsstørrelsen. Ved å teste om en tilfeldig oppdeling av utvalget i undergrupper påvirker estimatene, identifiserte Norheim forskjeller mellom egenskapene. Verdsettingen av de ulike tidsfaktorene viste seg å være nokså stabile selv med små delutvalg, mens verdsettingen av de mer spesielle egenskapene varierte mer;
- Betydningen av spørreskjemaets omfang. Tretthets-tegn hos respondenten kan identifiseres ved å sammenligne  $\rho^2$  (konstant) og t-verdien for første og andre halvdel av spørsmålene. Norheim fant ingen enhetlig tendens til forverring i den siste halvdel, selv om det var noe variasjon i signifikansnivå mellom de to halvdelene.

Innenfor testing av SP-design er det også blitt gjort undersøkelser av hvordan bruken av bilder og grafikk i SP-undersøkelsene påvirker resultatet. Nelson og Towris (1995) fant indikasjoner på at måten egenskapene presenteres på har stor innvirkning på resultatene. De sammenlignet spill der egenskapene ble presentert verbalt med spill der de ble presentert visuelt. Olsson (1997) påviste det samme i en SP-undersøkelse som kartla bilisters reaksjoner på ulik veikvalitet, som veiunderlag, vedlikehold og rasteplasser. I undersøkelsen førte bruken av illu-

strasjoner av disse egenskapene til at de ble gitt høyere verdsetting.

### 3.7 Oppsummering

Denne gjennomgangen har vist at det er mange fallgruver å passe seg for i designet av SP-undersøkelser. Det er lite som tyder på at forholdet mellom designet og resultatet er nøytralt.

Kommunikasjonen med respondentene er et svært kritisk punkt og er blant de viktigste metodeproblemene. Problemene som kan oppstå i intervjusituasjonen er mange og kan forårsakes både av spørsmålsstillingen, omstendighetene rundt intervjuet, respondentenes problemer med å sette seg inn i spørsmålene og deres problemer med å forutsi sin egen atferd i en hypotetisk situasjon.

Tre viktige arbeider som har testet ulike SP-design er blitt presentert. Alle tre har påvist at resultatene i stor grad påvirkes av designet av undersøkelsen.

Den mest omfattende av undersøkelsene ble foretatt av Staffan Widlert (1994), som fikk gjennomført og sammenlignet resultatene fra 15 SP-undersøkelser. De ulike SP-metodene *rank*, *rate* og *choice*, hver med flere ulike spørsmålsdesign, ga svært ulike resultater. Den

viktigste årsaken til dette er, ifølge Widlert, at respondentenes tilbøyelighet til å benytte forenklede beslutningsrutiner (gi leksikografiske svar) varierer med de ulike designene. Widlert fant at adaptive, computerbaserte undersøkelser basert på parvise valg ga de mest stabile resultatene.

Laitila og Westin (1997) sammenlignet design som hadde forskjellige rating-skalaer og ulike antall egenskapsnivåer. De fant også at de forskjellige designene ga ulike resultater. Samtidig fant de at resultatene avhang i stor grad av egenskaper ved respondentene, som alder, yrke og erfaring med egenskapene det spørres om.

Sjöstrand (1999) utførte 6 ulike SP-undersøkelser pr brev. Hver av metodene *rank*, *rate* og *choice* ble prøvd ut, hver med to ulike design. Også Sjöstrand påpekte at de store variasjoner i resultatene avhang av designet. De forskjellige designene hadde forskjellige egenskaper med hensyn til statistiske indikatorer og andre mer praktiske vurderinger. Sjöstrand fant at SP-undersøkelser lar seg gjøre via brev, men at valget av SP-metode og design er vesentlig for et sikkert resultat.

De arbeidene som er blitt presentert her, har brukt til dels svært ulike tilnærminger for å utforske sammenhengen mellom design og resultat, og til å vurdere de ulike designene opp mot hverandre.

## 4 Metode og datagrunnlag

I dette kapitlet vil vi presentere de ulike designvariantene som er brukt i undersøkelsen og de ulike testene som skal gjøres. Her vil vi også oppsummere datagrunnlaget i form av beskrivelser av utvalget og rekrutteringen.

Vi har definert "design" slik at det inkluderer disse aspektene ved SP-undersøkelsen:

- Rekrutteringsmetode og metode for datainnsamling
- SP-metoder
- utforming av de enkelte SP-metodene
- struktur på spørreskjema

### 4.1 Rekruttering og datainnsamling

Utgangspunktet for de vurderingene som er gjort i dette prosjektet er at en ønsker å beholde fordelene med en fleksibel, adaptiv design der en ved hjelp av et PC-program kan skreddersy valgene for den enkelte respondent samtidig som en ønsker å få ned kostnadene ved at respondentene administrerer intervjuet selv. Vi har derfor testet hvordan det fungerer når vi sender spørreskjemaet til respondentene som vanlig post vedlagt instruksjoner og en diskett med nødvendige filer – altså egenadministrerte intervjuer.

Rekrutteringen av disse respondentene er fordelt mellom:

- a) Telefonoppringning fra Gallup med invitasjon til å delta i undersøkelsen og få tilsendt diskett eller bli oppsøkt av intervjuer hjemme (ikke begge tilbud)
- b) Invitasjon på Trafikantens og Oslo Sporveiers hjemmesider på internett

Sistnevnte rekrutteringsmetode er benyttet kun for å få et større utvalg personer som tester det egenadministrerte skjemaet, og er ikke planlagt som rekrutteringsmetode i en hovedundersøkelse. Illustrasjoner av hvordan rekrutteringen har foregått er gjengitt i Vedlegg 1.

Resultatene fra variantene med egenadministrert diskett blir testet mot resultatene fra et utvalg som har gjennomført den "tradisjonelle" varianten med telefonrekrutterte, PC-baserte hjemmeintervju.

En annen mulighet, som i utgangspunktet var tenkt testet, er å sende e-post med instruksjoner og nødvendige filer som vedlegg. Dette har vi imidlertid vurdert slik at det stiller så store krav til respondentene (med hensyn til nedlasting av filer, installering av programmet og tilbakesending av den ferdige datafilen), at frafallet trolig ville blitt svært omfattende.

En tredje metode som også kunne vært testet er å legge spørreskjemaet ut på internett. Dette krever imidlertid at den adaptive oppdateringen av sidene kan foregå raskt, slik at respondentene slipper å vente og dermed går lei og avbryter intervjuet før det er ferdig. En slik nettvariant krever mer ressurser enn vi har tilgjengelig i dette prosjektet til utvikling og utprøving av grensesnitt, men vil i høyeste grad være interessant for senere studier dersom egenadministrasjonen av spørreskjemaene, som vi får testet i denne omgang, viser seg å fungere tilfredsstillende. En ferdigutviklet nettvariant vil både gjøre det enklere for respondentene og redusere kostnadene sammenlignet med diskettvarianten.

Det er med andre ord tre kategorier for rekruttering og datainnsamling som inngår i studien:

1. Egenadministrert, forhåndsrekruttert på telefon
2. Egenadministrert, rekruttert på internett
3. Hjemmeintervju, rekruttert på telefon

Vi har stilt følgende krav til respondentene:

- Må ha reist kollektivt med Oslo Sporveier siste måned
- Må være 16 år eller eldre
- Må være villig til å motta intervjuer hjemme (for hjemmeintervjuene) eller disponere PC og være villig til å svare på spørsmålene på egenhånd og returnere disketten

Med et budsjett på ca 100.000 kroner til datainnsamling har vi fått til 66 hjemmeintervjuer fordelt på to designvarianter. Innenfor rammen fikk vi også forhåndsrekruttert 427 navn via telefon av Norsk Gallup. Rekruttering via internett er en billig metode, men denne rekrutteringen gikk relativt sent. I løpet av om lag en måned hadde 193 personer latt seg rekruttere.

## 4.2 SP-metoder

De ulike SP-metodene som er kombinert og sammenlignet i denne studien, er:

- Betinget verdsetting (*Contingent valuation*, CV) og likeverdspris (*Transfer Price*, TP).
- Frisch-metoden, er en tidligere uprøvd, men intuitiv metode for å "sirkle inn" en respondents avveining mellom to og to egenskaper. I en serie valg mellom to reisealternativer blir forskjellene i nivåene på egenskapene gradvis redusert i retning av respondentens preferanser.
- Samvalg, *Stated Choice* (SC).
- "Fleksibel pakkesammensetning", (*Package Option Grid*, POG) er en ny mulighet i programpakken MINT der respondenten selv kan velge den kombinasjonen av godeforbedringer (både hvilke goder og hvilket nivå) som gir størst nytte. Samtidig må respondenten ta hensyn til den totale kostnaden for pakken av goder som velges.

### 4.2.1 Betinget verdsetting og likeverdspris

I CV/TP-sekvensene stilles spørsmål av typen: "Hva er det maksimale du er villig til å betale for en reisetidsforbedring på X minutter?" eller "Hva er det maksimale du er villig til å betale for den samme reien, men uten sitteplass?"

Utgangspunktet for disse beregningene er prisen pr reise som respondenten faktisk har betalt. I eksemplene over fungerer den faktiske billettprisen som henholdsvis nedre og øvre grense for betalingsvillighet. Med andre ord, hvis respondenten oppgir det samme beløpet som reisen i utgangspunktet kostet, tilkjenner hun null betalingsvilje for, eller verdsetting av, disse egenskapene. Verdsetting av forbedringer i tilbudet er det samme som *betalingsvillighet*, mens verdsetting av forverring er uttrykk for *kompensasjonskrav*.

#### Eksempel på CV/TP-sekvens

Bussreisen du beskrev tok 20 minutter og kostet kr 15. Tenk deg nå at denne reisen tar 25 minutter. Hva er det maksimale du ville betalt for reisen med økt reisetid? Dvs. ved hvilken pris er de to reisene likeverdige?

Reise 1: Reisetid = 20 minutter og kostnad = 15 kr  
Reise 2: Reisetid = 25 minutter og kostnad = ? kr

Gi ditt svar i intervallet 0-15 kroner: \_\_\_\_\_

Over har vi gjengitt et typisk eksempel på hvordan CV/TP-metoden er brukt i undersøkelsen. Nivåene for både reisetiden og kostnaden er relatert til respondentens faktiske reise. Hvis respondenten svarer 12 kroner i dette eksemplet, er det et uttrykk for verdsetting av reisetid på:

$$(15-12 \text{ kr}) / (25-20 \text{ minutter}) = 3/5 \text{ kr pr min} = 60 \text{ øre/min eller } 36 \text{ kr/time}$$

Hvis respondenten derimot svarer 15 kroner, som er det høyeste mulige valget, gir denne respondenten til kjenne en betalingsvillighet på 0 kroner for tidsbesparelsen, og dermed ingen verdsetting av reisetid. Men hvis respondenten svarer 0 kroner, blir regnestykket:

$$(15-0 \text{ kr}) / (25-20 \text{ minutter}) = 15/5 \text{ kr pr min} = 3 \text{ kr/min eller } 180 \text{ kr/time}$$

Et svar på 0 kroner er med andre ord ikke et uttrykk for lav verdsetting, men snarere det motsatte – et overbud.

Den øvre grensen for betalingsvillighet i spørsmål med forbedringer er satt vilkårlig til 99 kroner for å sikre at respondentene ikke gir ekstreme svar ved f.eks. å forveksle kroner og øre eller feiltasting.

Det er dermed to viktige problemstillinger knyttet til CV/TP-metoden. Den første går på 0-svar, altså når respondenten ikke tilkjenner verdsetting av en kvalitetsegenskap. Den andre knytter seg til overbud. Overbud skjer enten når respondenter oppgir svært høy betalingsvillighet for forbedringer eller ikke er villige til å betale noe for en reise som innebærer en kvalitetsforverring.

Vi har bedt respondentene verdsette følgende endringer i tilbudet:

- Forbedret reisetid
- Forverret reisetid
- Forbedring fra sitte halve reisen til sitte hele reisen\*
- Forbedring stå hele reisen til sitte hele reisen\*
- Forverring fra sitte halve reisen til stå hele reisen\*
- Forverring fra sitte hele reisen til stå hele reisen\*
- Redusert ventetid
- Pakke 1: Realtidsinformasjon, overbygg/tak, nødtelefon og vektore
- Pakke 2: Kort reisetid, sitteplass, høy frekvens\*

\* Avhengig av om respondenten satt eller sto hele eller deler av veien.

#### 4.2.2 Frisch-metoden

Både ved CV/TP og samvalg er det metodiske svakheter som i stor grad kan påvirke de verdsettingene som fremkommer fra analysene. Generelt kan en si at problemet med disse metodene er relatert til om valgsituasjonene som presenteres er slik at de ikke bare gjør det *mulig* for respondentene å tilkjenne sine preferanser for godene som presenteres, men at de også gjør valgene *overkommelige* (om ikke akkurat *lette*) for respondentene. Dersom valgene/spørsmålene er for vanskelige/upresise, kan en oppleve at respondenter som forsøker å gjennomføre valgoppgaven etter intensjonen ikke klarer å velge konsistent og i tråd med sine preferanser.

For om mulig å redusere omfanget av denne typen problemer i SP-data, vil vi prøve ut en SP-metode foreslått av Ragnar Frisch i sitt Nobelprisforedrag (Frisch 1972). Prinsippene for denne SP-metoden, som vi helt enkelt har kalt "Frisch-metoden" (og forkortet FM) er presentert i neste seksjon. Hvordan Ragnar Frisch brukte metoden, er beskrevet i Sælensminde (2001c).

Prinsippene for Frisch-metoden (FM) er at respondenten presenteres for to alternativer der f.eks. to faktorer inngår og at respondenten skal avveie forbedringer i den ene faktoren mot forverring i den andre. Til forskjell fra andre SP-metoder vil en i FM gradvis forsøke å nærme seg respondentens nivå for indifferens der forbedringen i den ene faktoren veier like mye som forverringen i den andre. Denne gradvise tilnærmingen til indifferensnivået oppnår en ved å gjøre forverringen (betalingen for forbedringen) i det første valget så stor at den langt overgår det de fleste (helst alle) respondentene er villig til å avgi (betale) for å oppnå den skisserte forbedringen.<sup>2</sup> I tabell 4.1 er det vist et eksempel der en respondent avveier 25% redusert reisetid mot høyere takst. For å få en stor nok forverring i første valg er taksten satt tre ganger så høy som den taksten respondenten i utgangspunktet hadde betalt for sin reise.

Eksemplet i tabell 4.1 viser at for hver gang respondenten velger reise 1 blir taksten i reise 2 gjort mindre i neste valg. Og når respondenten velger reise 2 blir taksten i reise 2 gjort større i neste valg. Ved at takstforskjellen mellom reise 1 og 2 (ev. mellom et valg respondenten har sagt nei til, reise 2 i valg IV, og

et valg han/hun har sagt ja til, reise 2 i valg V), blir halvert nærmer man seg raskt de nivåer der respondenten vil være indifferent mellom alternativene. Her er det bare brukt hele kroner og hele minutter i alternativene. Hvor nøyaktig en skal være mht å forsøke å fastslå folks preferanser kan diskuteres, men både for lite og for stort detaljnivå kan skape problemer.

Ved bruk av FM får vi fra sekvensen i tabell 4.1 vite at respondenten vil betale 18 kroner for reisen med redusert reisetid. Prisforskjellen mellom 15 og 18 kroner tilsvarer altså tidsforskjellen på 5 minutter. Omregnet i kroner pr minutt får vi altså 3 kr/5 min = 60 øre.

Valgene som gis til respondentene i Frisch' metode er mye enklere enn for enkelte andre SP-metoder. Bruk av PC til å administrere nivåene for faktorene gjør dessuten oppgaven svært enkel for eventuelle intervjuere og egenadministrerte intervjuer. Det burde derfor ligge vel til rette for å kartlegge befolkningens preferanser ved hjelp av Frisch-metoden.

Tabell 4.1: Eksempel på valgsekvens i Frisch-metoden. Respondenten skal her velge mellom reise 1 og 2 der takstnivåene i reise 2 varierer. Respondenten har anledning til å si at reise 1 er like god som reise 2 når de to reisen oppfattes som like gode

Valg nr	Reise 1 (respondentens faktiske reise)		Reise 2		Hva som ble valgt
	Reisetid	Takst	Reisetid	Takst	
I	20 min	15 kr	15 min	45 kr	Reise 1
II	20 min	15 kr	15 min	30 kr	Reise 1
III	20 min	15 kr	15 min	23 kr	Reise 1
IV	20 min	15 kr	15 min	19 kr	Reise 1
V	20 min	15 kr	15 min	17 kr	Reise 2
VI	20 min	15 kr	15 min	18 kr	Reise 1 = Reise 2
VII	20 min	15 kr	15 min		

©TØI 2001

#### 4.2.3 Samvalgdesignene

Standarddesignet av samvalgsekvensen består av to spill/valgsekvenser. Det ene spillet inneholder de "lettere" kvalitetsfaktorene informasjon på holdet, plass, leskur, trygghet og pris. Det andre spillet inneholder de tyngre faktorer reisetid, sitteplass, frekvens og pris. Hver av faktorene har tre nivåer.

Basisnivået for prisfaktoren er fastsatt ut fra det billettslag respondenten har brukt på den aktuelle reisen. Følgende priser pr reise er lagt til grunn:

<sup>2</sup> Dersom en respondent, til tross for stor forverring i det første valget, likevel sier seg villig til å avgi/betale den skisserte forverringen, kunne en også forsøkt å finne denne respondentens indifferensnivå ved å gi alternativer med ytterligere forverringen. Dette har vi ikke gjort i dette prosjektet fordi en da lett kommer til "urealistiske" nivåer, der en kan ha respondenten mistenkt for å utnytte at undersøkelsen er hypotetisk og at den tilkjennegitte betalingsvillighet ikke er reell. Jfr tilsvarende problemer i CV/TP der slike "overbud" kan være et stort problem.

Billetttype	Gjennomsnittspris
Enkeltbillett	20 kroner
Flexikort	16 kroner
Dagskort	14 kroner
7-dagerskort	15 kroner
Månedskort	13 kroner
3-månederskort	11 kroner
Annet	15 kroner
Ungdom og honnørreisende	50% av pris

©TØI 2001

Respondentene får i standarddesignet av både valgsekvens 1 og 2 valget mellom to ulike reiser (reise A og B) med det transportmiddelet de brukte på den aktuelle reisen (enten buss, trikk, T-bane eller lokal-tog). I hvert enkelt valg varierer nivåene på de fire faktorene som inngår innenfor de nivåene som er skissert i tabell 4.2 og 4.3. I standarddesignet får respondenten fire valgalternativer:

1. "Helt sikkert alternativ A"
2. "Trolig alternativ A"
3. "Trolig alternativ B"
4. "Helt sikkert alternativ B"

Dette gir oss muligheten til å analysere dataene både med hensyn til preferanseusikkerhet og å slå sammen

Tabell 4.2: Samvalgsekvens 1: Lettere (mindre tungtveiende) kvalitetsfaktorer relatert til holdeplassen

	Nivå 1	Nivå 2	Nivå 3
Informasjon på holdeplassen	Rutetabell	Rutetabell + rutekart	Rutetabell + rutekart + informasjon om faktisk ankomsttid
Leskur*		Leskur	Overbygget stasjon
Trygghet		Nødtelefoner på holdeplassene	Nødtelefoner på holdeplassene + vektere som patruljerer
Pris/reise	-3 kr	Basert på respondentens billetttype	+3 kr

\* Leskur tas ut i designvariant VI der valgsekvensen har kun tre faktorer

©TØI 2001

Tabell 4.3: Samvalgsekvens 2: Tungtveiende faktorer relatert til kollektivreiser

	Nivå 1	Nivå 2	Nivå 3
Reisetid	-25%	Basis	+25%
Frekvens	-50%	Basis	+ 10 minutter
Sitteplass*	Ingen del av reisen	Halve reisen	Hele reisen
Pris/reise	-5 kr	Basert på respondentens billetttype	+5 kr

\* Sitteplass tas ut i designvariant VI der valgsekvensen har kun tre faktorer

©TØI 2001

valg av alternativ 1 og 2 (hhv. 3 og 4) og analysere som om respondentene bare hadde fått to valgalternativer (A eller B).

En sentral målsetting i prosjektet er å bedre kvaliteten på dataene ved å ta i bruk ny design for å redusere andelen leksikografiske og inkonsistente valg i samvalgstudier. En måte å gjøre dette på er å vurdere hvordan ulike SP-metoder kan kombineres og hvilke muligheter dette gir for bedre å balansere nivåene i samvalgene. Vi har derfor forsøkt å balansere egenskapsnivåene i samvalgsekvensen ut fra respondentenes svar i CV/TP-sekvensen. I stedet for å definere prisvariasjonen som  $\pm 3$  og  $\pm 5$  kroner i forhold til den faktiske prisen respondenten betalte for reisen i henholdsvis samvalgsekvens 1 og 2, har vi brukt betalingsviljen for de to pakkene i CV/TP-sekvensen som grunnlag for variasjonen rundt billettprisen.

Problem med leksikografiske valg i samvalgstudier er trolig også avhengig av kompleksiteten i de valgalternativene respondenten blir forelagt. Her kan vi også teste betydningen av:

- Antall valg (mht lærings- og tretthetseffekter)
- Antall faktorer i hvert valg
- Preferanse-/valgusikkerhet (hvor sikker er respondenten på sitt valg?)



Vi har valgt å konsentrere oss om parvise valg (også kalt SP *choice* eller samvalg). Men dersom vi også tar med noen tester av hvor sikker respondenten er på sine valg grenser metoden inn mot rangering/poenggivning. En slik skala kan f.eks. være femdelt med følgende alternativ:

1. "Helt sikkert alternativ A"
2. "Trolig alternativ A"
3. "Alternativ A og B er like gode"
4. "Trolig alternativ B"
5. "Helt sikkert alternativ B"

Det er også mulig å ta med et sjette alternativ: "Vet ikke". Dette sjette alternativet vil gi respondenten en "utvei" dersom et valg blir så vanskelig at han/hun mister oversikten. Dersom det brukes slik av respondentene, vil det trolig bidra til å redusere inkonsistente valg og leksikografiske valg som skyldes "forenkling". Det er tross alt bedre at en respondent som ikke overskuer valgoppgaven velger "vet ikke" i alle valgene som blir presentert enn at han/hun velger konsekvent alternativet med laveste pris eller kortest reisetid.

All verdsetting, både i virkelige markeder og hypotetiske markeder, er kontekstavhengig. Samsvar mellom konteksten i SP-studien og konteksten resultatene senere skal brukes i er derfor viktig for å få mest mulig korrekte verdsettinger i den konteksten verdsettingene skal brukes. Et sentralt spørsmål i denne sammenheng er da:

Vil en *blokkdesign med grupper av faktorer på flere nivåer* kunne gi relative verdsettinger som i større grad kan oppfattes som absolutte verdsettinger og som dermed kan aggregeres uten at dette kommer i konflikt med respondentens budsjettbetingelser?

Slike spørsmål er det mulig å forsøke å besvare i denne designstudien.

Vi har testet følgende varianter av samvalgmetoden:

1. Balansering av nivåene ved hjelp av CV/TP. Størrelsen på variasjonen rundt basisprisen bestemmes av svarene i den forutgående CV/TP-sekvensen
2. Preferanseusikkerhet i SC
3. "Vet ikke"-alternativ i SC

4. Blokkdesign med grupper av faktorer på flere nivåer
5. Ta med opptil 15 valg for å studere lærings- og tretthetseffekter
6. Redusere antall faktorer fra 4 til 3 for å se hvordan dette påvirker leksikografi og verdsetting

#### 4.2.4 Fleksibelt pakkevalg

Fleksibelt pakkevalg, eller POG (*Package Option Grid*) er et nytt verktøy i softwareprogrammet WinMint. Det fungerer slik at respondentene kan sette sammen sin egen "reisepakke" der de selv bestemmer nivåene på de ulike egenskapene samtidig som dette påvirker billettprisen. Figur 4.1 viser hvordan POG er blitt brukt i vår undersøkelse, og hvordan endringene i kvalitetsfaktorene påvirker billettprisen.

Respondentene har hatt muligheten til å velge mellom tre nivåer for reisetid, frekvens, sitteplass, holdeplassinformasjon, leskur og trygghetsforhold. Alle har innvirkning på taksten (her målt takstendring i øre pr reise), som blir presentert nede på skjermbildet.

### 4.3 Struktur på spørreskjemaet

Tabell 4.4 beskriver de 10 ulike designene for spørreundersøkelsen. De fordeler seg på 6 ulike designvarianter for samvalgene, 2 rekrutteringsmetoder og 2 datainnsamlingsmetoder. I tillegg er rekkefølgen på de ulike SP-sekvensene variert. For å skille mellom de ulike designvariantene, rekrutteringsmetodene, datainnsamlingsmetodene og spørreskjemastrukturene, har vi definert unike designnummer for hver av kombinasjonene. Designnumrene gir altså en presis beskrivelse av hvilken designvariant, rekrutterings- og innsamlingsmetode som er brukt for hvert enkelt intervju. Tabellen beskriver disse. Av den kan vi lese for eksempel at design 4 og 14 har samme designvariant, men at respondentene er rekruttert forskjellig. Ved å slå sammen f.eks. intervjuene 13, 14, 15 og 16 kan vi se på de selvrekrutterte intervjuene under ett.

Nivå på forbedringer og endring i billettpris			
Forbedringer	Nivå 1	Nivå 2	Nivå 3
Tid på transportmiddelet <i>Takstendring</i>	25% lengre reisetid -200	Reisetid som nå 000	25% kortere reisetid 200
Tid mellom avgangene <i>Takstendring</i>	Dobbelt så lang tid mellom hver avgang -300	Tid mellom hver avgang som nå 000	Halvparten så lang tid mellom hver avgang 300
Stå-/Sitteplass <i>Takstendring</i>	Ståplass hele veien -200	Sitteplass halve veien 000	Sitteplass hele veien 200
Info på holdeplassen <i>Takstendring</i>	Rutetabell -020	Rutetabell+Rutekart 020	Rutetabell+Rutekart +Info om faktisk ankomsttid 050
Leskur/overbygg <i>Takstendring</i>	Ikke leskur på holdeplass/stasjon -020	Leskur på holdeplass/stasjon 020	Overbygg/tak på holdeplass/stasjon 100
Trygghetsforhold <i>Takstendring</i>	Ikke nødtelefoner eller vektere som patruljerer holdeplassene -050	Nødtelefoner på holdeplassene 010	Nødtelefoner på holdeplassene + vektere som patruljerer 060
Total takstendring -300			

©TØI 2001

Figur 4.1: Eksempel på POG-sekvens

Tabell 4.4: Oversikt over de ulike designnumrene. "Designvariant" viser til varianten av samvalgsekvensen

Designvariant for samvalg	Rekkefølge på SP-sekvensene	a) diskett pr post telefonrekruttert	b) diskett pr post selvrekruttert (internett)	c) hjemmeintervju forhåndsrekruttert
I	SC, POG, FM			1
II	CV/TP, SC, POG			2
III	CV/TP, SC, FM	3	13	
IV	SC, FM, CV/TP	4	14	
V	CV/TP, SC, FM	5	15	
VI	FM, SC, CV/TP	6	16	

©TØI 2001

## 4.4 Tester av design og verdsetting

Det grunnleggende spørsmålet i denne studien er om trafikantene kan verdsette kvaliteten på kollektivtilbudet. Vi har sett at det er utallige mulige feilkilder og problemer med SP-analyser og at det er grunn til å tvile på at en respondent har klart å tilkjenne sine faktiske preferanser. Våre tester omfatter:

- Størrelse, signifikans og fortegn på verdsettingene
- Vurdering av vanskelighetsgrad ved metodene
- Leksikografi
- 0-svar og overbud
- Preferanseusikkerhet
- Avvikende svar
- Tretthetstegn

- Intern konsistens
- Frafall
- Beskrivelser av utvalget og frafall

### 4.4.1 Størrelse, signifikans og fortegn på verdsettingene

De første testene går på selve beregningene for verdsetting av kvalitetsfaktorer. Fortegn og størrelse på verdsettingen må svare til forventningene, som bygger på tidligere studier og logisk forklaring (som at kortere reisetid er en forbedring og ikke en forverring). Tester av beregningene innebærer også signifikanstester både for parameterestimater (for samvalganalyser) og verdsettingene: Er verdsettingen signifikant forskjellig fra null?

#### 4.4.2 Vanskelig valgsituasjon?

Videre er alle de ulike SP-sekvensene etterfulgt av kontrollspørsmål om hvor vanskelig det var for respondenterne å svare på verdsettingsspørsmålene. I den grad noen av metodene er enklere for respondentene, vil de også være attraktive fordi det gir signaler om at respondentene mestrer intervju-situasjonen. Likevel er ikke dette det eneste kriteriet for en vellykket SP-metode, både fordi de enkleste metodene kan være mest utsatte for taktisk svargivning og protestvalg, og fordi de ikke nødvendigvis gir den riktige konteksten for verdsettingene.

#### 4.4.3 Leksikografi

Som tidligere nevnt er leksikografisk svargivning et alvorlig problem i samvalgundersøkelser. Vi vil derfor se på hvordan andelen leksikografiske svar varierer med de ulike designvariantene. Vi vil teste om antallet kvalitetsfaktorer som inngår i samvalgsekvensen, eller om balansering av prisvariasjonen ved hjelp av CV/TP påvirker andelen leksikografiske svar.

#### 4.4.4 0-svar og overbud

Med 0-svar mener vi at respondentent tilkjenner 0 i betalingsvilje for en kvalitetsfaktor i CV/TP eller alltid velger billigste alternativ i Frisch-sekvensen (og dermed tilsynelatende ikke foretar avveininger). Store andeler slike svar er vanskelig å håndtere, metodologisk. Vi antar at fenomenet i stor grad kan forklares med at respondenterne foretar protestvalg, taktiske valg, eller at de har sterke preferanser.

En beslektet problemstilling har vi kalt "overbud" i CV/TP. Overbud er når en respondent tilkjenner ekstrem høy verdsetting av en kvalitetsfaktor og kan igjen skyldes både protestvalg, taktisk svargivning og sterke preferanser.

Testingen av 0-svar og overbud inkluderer en rapportering av andelen slike svar og vurdering av mulige årsaker på grunnlag av designene av CV/TP- og FM-sekvensene.

#### 4.4.5 Preferanseusikkerhet

Alle samvalgsekvensene gir respondentene mulighet til å velge "*Helt sikkert alternativ X*" eller "*Trolig alternativ X*". Vi har sett på hvordan preferanseusikkerhet påvirker verdsettingsestimaterne ved å sammenligne modellkjøringer der bare "*helt sikkert*"-svarene inngår med modeller der "*trolig*"-svarene inngår.

#### 4.4.6 Avvikende svar?

Vi vil også teste om de ulike samvalgdesignene har påvirket andelen avvikende svar. Et avvikende svar innebærer at respondentent gjør et valg som det ut fra logit-modellen er svært liten sannsynlighet for at vedkommende skal velge. En høy andel avvikende svar indikerer således enten at respondentene har hatt problemer med å takle valgsituasjonen, at den beregnede nyttefunksjonen i liten grad forklarer valgene respondentene gjør, eller at respondentene har store avvik i preferanser. Mange avvikende svar vil derfor svekke beregningene av verdsettingene.

#### 4.4.7 Tretthet

For å teste om tretthet hos respondentene har påvirket resultatene av samvalgundersøkelsen vil vi gjøre tre ulike tilnærminger:

1. Beregne separate verdsettinger for første, andre og tredje tredjedel av valgomgangene
2. Vurdere forekomsten av avvikende svar
3. Teste for heteroskedastisitet som følge av tretthet

#### 4.4.8 Intern konsistens

I undersøkelsen ser vi på fire ulike metoder for verdsetting: Samvalg (SC), Betinget verdsetting og likeverdspris (CV/TP), Frisch-metoden (FM) og Fleksibel pakkesammensetning (POG). Dette gir en unik mulighet til å teste datamaterialet for *intern konsistens* mellom SP-metodene, det vil si om en og samme respondent klarer å tilkjenne verdsettinger som i stor grad samsvarer mellom de ulike metodene. Datamaterialet legger til rette for en rekke sammenligninger på tvers av de ulike metodene og kan gi svar på spørsmål som:

- Er det sammenheng mellom høy/lav verdsetting i FM og høy/lav verdsetting i CV/TP?
- Er det samsvar mellom 0-svar i CV/TP og FM, og leksikografi i samvalgene?

I den grad det faktisk *er* samsvar, har vi også et grunnlag for å svare på om trafikantene kan verdsette kvaliteten, og i hvor stor grad designet påvirker resultatet. Når vi tester datamaterialet for intern konsistens gjennomfører vi derfor en svært viktig test av hvordan de ulike verdsettingsekvensene har fungert.

Vi kan også teste for konsistens a) innenfor en SP-metode og b) innenfor en enkelt sekvens i en metode. Vi vil fokusere på den første av disse tilnærmingene i behandlingen av Frisch-sekvensen og komme inn på

problemstillingen i gjennomgangen av CV/TP-sekvensene.

## 4.5 Frafall og utvalgsskjevheter

Rekrutteringen av respondentene pr telefon ble gjennomført av Gallup. Det er viktig å ha gode rutiner for frafallsregistrering og en god dokumentasjon av hvem som var i målgruppen for undersøkelsen. Frafallsrapporteringen fra Gallup er derfor gjengitt i Vedleggene 2, 3 og 4. Av rapporteringen går det frem at totalt 3619 personer ble ringt opp for rekruttering til egenadministrerte intervjuer. Av disse ble det gjort avtaler om å sende undersøkelsen til 429 personer, eller om lag 12 %. For hjemmeintervjuer ble 1036 personer oppringt og det ble gjort avtaler med 7 % av dem. I felten falt 8 respondenter fra. Det er altså en større andel gjennomførte telefonintervjuer blant dem som ble rekruttert til egenadministrerte intervjuer.

Vi kan skille mellom ulike årsaker til frafall:

- Man ikke får kontakt med respondenten
- Respondent ikke er i målgruppen
- Respondent nekter i utgangspunktet å delta
- Respondenten har sagt seg i utgangspunktet villig til å delta da de ble oppringt, men svarer allikevel ikke

Tabell 4.5 viser fordelingen av de tre første frafallskategoriene for den Gallup-baserte telefonrekrutteringen til egenadministrerte intervjuer og hjemmeintervjuer. Den største andelen frafall er skyldes at det ikke er oppnådd kontakt (det ble kun gjort ett forsøk pr respondent) eller at respondentene i utgangspunktet nekter å delta. Det er relativt små andeler som faller fra på grunn av kriteriene om at respondenten må være over 16 år, ha reist kollektivt i Oslo den siste måneden og ha tilgang til PC (for egenadministrerte intervjuer). Til sammen var det om lag 30 % av de oppringte som ble ekskludert på grunn av disse kriteriene i hver av rekrutteringsrundene. Vi merker oss også at bare 14 % falt fra fordi de ikke hadde tilgang til PC.

For å vite hvor stor andel som ikke tilfredsstilte kriteriene blant dem som faktisk ble spurt om det, må vi se bort fra dem det ikke ble opprettet kontakt med og de som i utgangspunktet nektet å være med på en telefonundersøkelse. Tabell 4.6 viser beregningene. For rekruttering til egenadministrerte intervjuer ser vi at om lag en tredjedel av de oppringte som deltok i telefonintervjuet ble ekskludert fordi de ikke hadde tilgang til PC. En like stor andel ble ekskludert fordi de ikke hadde reist den siste måneden. Fordelingen er skjev for rekrutteringen til hjemmeintervjuer, men

skyldes sannsynligvis feil i kodingen fra Gallup. Selv om utvalget er relativt lite, er det lite sannsynlig at hele 80 % ikke har reist kollektivt i et tilfeldig utvalg.

Tabell 4.5: Frafall (prosent) ved telefonrekruttering til egenadministrerte intervjuer

	Rekruttering til egenadministrert (N=3619)	Rekruttering til hjemmeintervju (N=1030)
Kan og vil delta	12	7
Ikke kontakt	37	34
Nekter i utgangspunktet å delta	21	29
<i>Ikke målgruppe</i>		
- Under 16	2	0
- Ikke PC	14	-
- Ikke reist	14	30
<b>SUM</b>	<b>100</b>	<b>100</b>

©TØI 2001

I Tabell 4.6 ser vi nærmere på respondenter som har sagt seg villige til å delta i telefonintervjuet, dvs at vi ser bort fra personer det ikke er oppnådd kontakt med, og personer som i utgangspunktet ikke var villig til å delta i noen undersøkelse. Tabellen inkluderer andelen som nektet å delta da de ble forespeilet hjemmebesøk eller å få tilsendt diskett som skulle returneres. 5 % nektet å delta i egenadministrerte intervjuer, og 1 % nektet hjemmeintervju. Det er altså en noe høyere andel som vegrer seg mot diskett som datainnsamlingsmetode enn som vegrer seg mot hjemmeintervju. Samtidig er andelen som sier seg villig til å delta høyere i rekrutteringen til egenadministrerte intervjuer.

Tabell 4.6: Frafall (prosent) blant dem som deltok i telefonintervjuet

	Rekruttering til egenadministrert (N=1604)	Rekruttering til hjemmeintervju (N=387)
Villige til å delta	27	19
Ikke målgruppe: Under 16	5	
Ikke målgruppe: Ikke PC	32	-
Ikke målgruppe: Ikke reist	31	80*
Nekt pga diskett/hjemmeintervju	5	1
<b>SUM</b>	<b>100</b>	<b>100</b>

\* Sannsynligvis feil i registreringene hos Gallup

©TØI 2001

Når det gjelder respondenter som har sagt seg villige til å delta i undersøkelsen, men som vi likevel ikke får svar fra, kan vi skille mellom de ulike rekrutterings- og data-innsamlingsmetodene. Tabell 4.7 viser at selv om svarprosenten blant de egenadministrerte intervjuene er lavere enn det vi forventet, er den forholdsvis lik for de to rekrutteringsmetodene. Det er altså ingenting som tyder på at rekrutteringsmetoden har påvirket frafallet. En stor del av frafallet (om lag 10 %) har skyldtes tekniske problemer med gjennomføringen på datamaskiner (diskett, operativsystem osv). En mer brukervennlig PC-basert løsning, som internett, ville derfor sannsynligvis ha redusert frafallet.

Når det gjelder frafall i felt, altså hjemmeintervjuene, er frafallet etter at respondentene er rekruttert på telefon, svært lavt. Det lave frafallet er ikke direkte sammenlignbart med de to andre beregningene, fordi det på forhånd er gjort avtaler med respondentene. De registrerte frafallsårsakene for hjemmeintervjuene er sykdom hos respondenten, respondenten ikke hjemme og intervjuer rakk ikke frem i tide. Disse årsakene er mindre sannsynlige for egenadministrerte intervjuer. Derfor kan vi ikke slutte at hjemmeintervjuer har høyere svarprosent enn egenadministrerte, gitt forhåndsrekrutteringen.

Tabell 4.7: Antallet gjennomførte intervjuer fordelt på ulike rekrutterings- innsamlings- og designvarianter. (Frafall etter rekruttering i parentes)

	a)	b)	c)
Design-variant	Diskett pr post forhånds-rekruttert	Diskett pr post internett-rekruttert	Hjemme-intervju
I			35*
II			31*
III	53 (50,9%)	19 (55,8%)	
IV	52 (50,9%)	28 (36,4%)	
V	44 (58,5%)	27 (41,3%)	
VI	55 (48,6%)	20 (56,5%)	
Alle	204 (47,8%)	94 (52,5%)	66 (11%)

\*Frafallet er ikke fordelt på designvariantene.

©TØI 2001

## 4.6 Utvalget

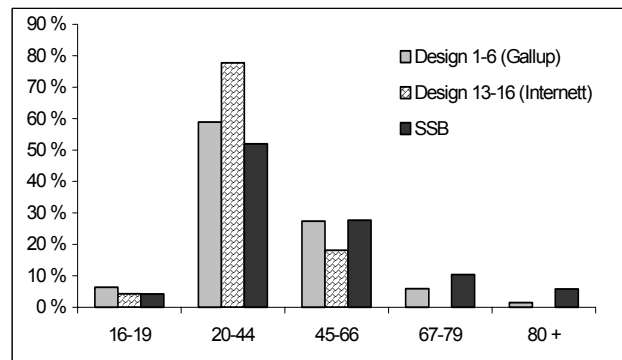
Vår studie innebærer rekruttering av respondenter via internett og gjennomføring av intervjuene ved hjelp av egenadministrerte disketter. Begge deler er nye i Norge i SP-sammenheng. Vi har derfor gått noe i dybden i analysene av utvalget som inngår i studien for å dokumentere

effektene av disse metodene. Dette arbeidet er dokumentert i Nergård (2001).

Vi har sett på variable som alder, kjønn og inntekt for å vurdere om utvalget representerer populasjonen, dvs. Oslos befolkning<sup>3</sup>. Som referanse for alders- og kjønnsfordeling ble det brukt tall i fra Statistisk Sentralbyrå (SSB). For sammenligninger av husholdningenes årlige bruttoinntekt kunne ikke SSBs data brukes som referanse på en tilfredsstillende måte. Vi har derfor brukt tall i fra Reisevaneundersøkelsen 1998.

Totalt deltok 364 personer i undersøkelsen fordelt på 66 personer som deltok i hjemmeintervju (design 1 og 2), 204 personer som ble rekruttert pr telefon og fikk tilsendt diskett (design 3-6) og 94 personer som ble rekruttert på internett og som svarte på tilsendt diskett (design 13-16).

Figur 4.2 illustrerer aldersfordelingen i befolkningen, sammenlignet med aldersfordelingen blant respondentene i design 1-6 (rekruttert av Gallup) og 13-16 (rekruttert på internett).



©TØI 2001

Figur 4.2: Aldersfordeling i befolkningen (SSB), og blant respondentene i design 1-6 (rekruttert av Gallup) og 13-16 (rekruttert på internett)

Av figuren ser vi at aldersfordelingen i Norsk Gallups utvalg er relativt lik SSBs tall. Gjennomsnittsalderen på dem som ble trukket ut av Norsk Gallup er om lag 40 år, mot 45 år som er gjennomsnittet blant befolkningen over 16 år.

Ikke uventet har respondentene som ble rekruttert på internett en langt lavere aldersfordeling. Vi har faktisk ingen observasjoner av personer over 60 år blant disse respondentene. Gjennomsnittsalderen for design 13-16 er om lag 33 år, noe som er 12 år lavere enn gjennomsnittsalderen i befolkningen. Vi kan av dette tydelig se at utvalget som ble rekruttert gjennom internett er skjevare

<sup>3</sup> På grunn av utvalgs-kriteriene om at respondenten skal være over 16 år, ha reist kollektivt den siste måneden, og ha tilgang til PC (for egenadministrerte intervjuer), kan vi ikke uten videre vente at vårt utvalg utgjør et tverrsnitt av befolkningen. Tidligere studier har imidlertid vist at kollektivtrafikanter i stor grad er representative for en befolkning.

enn utvalget som ble rekruttert gjennom Norsk Gallup sammenlignet med SSBs tall.

Tabell 4.8 viser at rekrutteringsmetoden har hatt avgjørende betydning for kjønnsfordelingen blant respondentene. Utvalget som er rekruttert av Norsk Gallup (design 1-6) stemmer relativt godt med den faktiske befolkningen i Oslo, mens rekrutteringen gjennom internett har gitt en signifikant høyere andel menn enn hva som er tilfelle i befolkningen. Når vi i tillegg vet at det er overvekt av kvinner som reiser kollektivt, er det klart at rekruttering på internett har gitt et svært skjevt bilde av kjønnsfordelingen i befolkningen over 16 år og som har reist kollektivt minst én gang den siste måneden.

Tabell 4.8: *Kjønnsfordeling i befolkningen (SSB), og blant respondentene i design 1-6 (rekruttert av Gallup) og 13-16 (rekruttert på internett). Prosent*

	Kvinner	Menn
Rekruttert internett	40	60
Rekruttert Gallup	53	47
Faktisk befolkning (SSB)	52	48

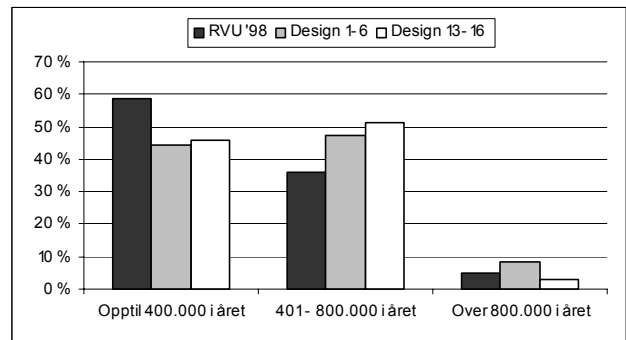
©TØI 2001

Man skal alltid være forsiktig med å trekke for hastige konklusjoner basert på respondentenes oppgitte husholdningsinntekt. Mange ønsker ikke å oppgi denne, og det er rimelig å anta at noen ikke vet svaret for sin familie. Vi likevel valgt å nevne denne variabelen her fordi vi vet at inntekten spiller inn på trafikantenes verdsetting av kvaliteten på kollektivtransporten. Blant andre har Gunn og Sheldon (2000) påvist en klar sammenheng mellom inntektsnivå og verdsetting av tid. Videre har vi sammenlignet vårt utvalg med Reisevaneundersøkelsen (RVU) for 1998 (indeksregulert), noe som skulle gi et relativt godt sammenligningsgrunnlag idet vi kan anta at skjevhetene i besvarelsene vil være de samme.

Figur 4.3 viser inntektsfordelingen i studien vår sammenlignet med RVU '98 (1518 observasjoner). Her ser vi at det er godt samsvar mellom de to rekrutteringsmetodene i vår studie, men at fordelingen er skjev i forhold til RVU '98. Den gjennomsnittlige husholdningsinntekten i vår studie er ca 468.000 kroner (2 ekstreme observasjoner er ekskludert) mot 386.000 kroner i RVU-dataene.

Vurderingen av utvalget viser at internettrekruttering totalt gir en skjev representasjon av befolkningen. Utvalgene som er trukket ut av Norsk Gallup representerer i stor grad populasjonen når det gjelder alder og kjønn. Når det gjelder inntekt har vi observert store avvik i forhold til Reisevaneundersøkelsen 1998, samtidig som det er godt samsvar mellom internettrekrutteringen og rekrutteringen ved hjelp av Gallup. Skjevhetene i utvalget totalt

kommer av at det er stor overvekt av menn i yngre alder som er rekruttert gjennom internett og har ønsket å være med i undersøkelsen. Dette er et forhold som tas i betraktning når de ulike utvalgene sammenliknes senere i denne analysen.



©TØI 2001

Figur 4.3: *Brutto husholdningsinntekt i design 1-6 (rekruttert Gallup) og 13-16 (rekruttert internett) sammenlignet med RVU 1998*

#### 4.6.1 Små utvalg kan gi tilfeldige variasjoner

Det er til dels svært små utvalg i de enkelte designvariantene. Det er derfor viktig å undersøke om tilfeldige variasjoner har forårsaket store skjevheter i forhold til egenskaper ved respondentene og ha dette in mente i de videre undersøkelsene. Slik kan vi unngå å forveksle påvirkning av SP-designen med påvirkning som skyldes tilfeldige utvalgsvariasjoner.

Tabell 4.9 sammenstiller en del gjennomsnittstall for utvalgene. Vi merker oss at i design 3 har respondentene oppgitt en langt høyere gjennomsnittsinntekt enn de andre designene. Dette skyldes i hovedsak én ekstrem observasjon. Gjennomsnittlig ventetid er høyest i design 13 (21,8 minutter mellom avgangene) og lavest i design 16 (12,7 minutter). Reisetiden varierer mellom 21,7 minutter for design 2 og 31,5 minutter for design 14. Reisetiden er også forskjellig mellom de tre utvalgsgruppene: Den er noe kortere for hjemmeintervju-utvalget (1 og 2) enn for utvalget som ble telefonrekruttert for egenadministrerte intervjuer (3-6), men de lengste reisetidene forekommer i det selvrekrutterte utvalget (13-16). Når det gjelder respondentenes alder, er det en klar tendens til at utvalgene til hjemmeintervjuene er eldre enn de øvrige utvalgene. Videre er det selvrekrutterte utvalget det yngste. Et lignende mønster ser vi i kjønnsfordelingen: Det er flest kvinner blant hjemmeintervjuene, og færrest blant de selvrekrutterte. Dette er ikke bare tilfeldige variasjoner på grunn av små utvalg, men uttrykk for egenskaper ved rekrutterings- og datainnsamlingsmetodene.

Tabell 4.9: Et utvalg gjennomsnittsverdier for de ulike designvariantene

Design	Tid mellom avgangene	Pris pr reise	Reisetid (min.)	Ant. personer i husholdningen	Husholdningens bruttolinntekt i 2000	Respondentens alder	Andel kvinner
1	15,1	13,9	22,1	2,06	435	43	54%
2	16,8	12,8	21,7	2,87	414	46	68%
3	15,0	13,8	23,5	2,47	811	40	42%
4	14,1	14,5	23,4	2,60	544	36	62%
5	15,8	14,5	26,9	2,41	442	38	41%
6	17,2	14,5	23,1	2,51	525	40	55%
13	21,8	13,5	24,0	2,37	512	30	37%
14	18,0	13,9	31,5	2,57	451	34	43%
15	17,6	14,0	30,0	2,11	370	33	41%
16	12,7	14,2	28,0	2,10	464	34	40%

©TØI 2001

Verdsetting av kvalitetsfaktorer varierer med reisemiddel (Norheim, 1996). Tabell 4.10 viser fordelingen av hovedkollektivtransportmiddel fordelt på de ulike designvariantene. De største variasjonene i fordelingen er i lokaltog, der design 2 har 0 % og design 13 har 11 %, og for trikk der andelen varierer mellom 7 og 27 %. Design 13 har den høyeste andelen buss og den laveste andelen t-banebraukere, noe som sannsynligvis vil påvirke verdsettingene for dette utvalget.

Tabell 4.10: Fordeling (prosent) av hovedkollektivtransportmiddel fordelt på designvariantene

Design	Buss	Lokaltog	T-bane	Trikk
1	34	3	51	11
2	29	0	48	23
3	30	9	36	25
4	27	10	46	17
5	23	7	55	16
6	31	7	44	18
13	42	11	21	26
14	39	4	43	14
15	37	7	48	7
16	30	5	40	25

©TØI 2001

Formålet med reisen kan også ha avgjørende betydning for verdsetting av ulike kvalitetsfaktorer. Tabell 4.11 sammenstiller fordelingen av et utvalg reiseformål mellom de ulike designvariantene. Vi ser at andelen arbeids-/ skolareiser i stor grad er bestemt av rekrutterings- og datainnsamlingsmetode. Design 1 og 2 har under 50 % slike reiser, design 3-6 noen flere, mens de høyeste andelen er i det selvrekrutterte utvalget. Blant

hjemmeintervjuene er det lave andeler arbeidsreiser både i design 6 og 16, som har samme utforming men ulik rekruttering). Vurderinger av design 6/16, der antallet faktorer er redusert fra 4 til 3 i samvalgsekvensene, må derfor ta hensyn til dette.

Tabell 4.11: Fordeling (prosent) av hovedformål ved reisen etter designvariant. Andelen summerer ikke til 100 % fordi dette bare er et utvalg av reiseformålene

Design	Til/fra arbeid/skole	Innkjøp	Privat ærend/service	Egen fritidsaktivitet
1	49	6	29	9
2	48	16	19	6
3	68	8	8	8
4	79	8	2	4
5	59	7	20	5
6	56	4	5	15
13	79	5	5	0
14	82	0	11	4
15	78	0	11	4
16	60	10	20	10

©TØI 2001

Hva slags type billett de reisende benyttet vil påvirke designet på de adaptive SP-undersøkelsene (SC, CV/TP og FM). Samtidig er det sannsynlig at respondentenes betalingsvilje for ulike kvalitetsfaktorer i noen grad varierer med hva slags korttype de benytter. Tabell 4.12 viser fordelingen av de viktigste billetttypene mellom designvariantene. Design 1 og 2, som hadde en lavere andel yrkes-/skolareiser, har også en lavere andel 3-månederskort enn de fleste andre designene. Den høyeste andelen med enkeltbillett 23 %

i design 2, mens ingen i design 13 og 16 reiste på enkeltbillett. Det kan se ut som om de selvrekrutterte (design 13-16) er mer aktive brukere av kollektivtransport, ved at de har lave andeler reisende på enkeltbillett og høye andeler periodekort. Dette stemmer godt med at de også har høyere andeler arbeids- og skolereiser, og at de aktivt har oppsøkt Trafikanten og Oslo Sporveiers hjemmesider.

Vi kan derfor slå fast at det er en del variasjon i bakgrunnvariablene som følge av tilfeldige utvalgskjevheter i de små utvalgene. Dette er det viktig å ta hensyn til i betraktningene omkring hvordan designet påvirker resultatet.

Imidlertid er de mest påfallende forskjellene de som går mellom rekrutterings- og datainnsamlingsmetodene. Mange av de ”ekstreme” observasjonene av bak-

grunnvariabler er i design 13-16, som ikke er noe tilfeldig utvalg av befolkningen, men et selvrekruttert utvalg. Videre er det en tendens til at utvalgene for hjemmeintervjuer har større overvekt av kvinner, er eldre, reiser i mindre grad med lokaltog, har færre arbeidsreiser og benytter periodekort sjeldnere enn utvalget som gjennomførte egenadministrerte intervjuer. Det selvrekrutterte utvalget er yngst av utvalgene. Dette er skjevheter som ikke skyldes tilfeldig variasjon i små utvalg, men som sannsynligvis er resultat av rekrutterings- og datainnsamlingsmetodene. Vi merker oss spesielt at et opplegg med egenadministrasjon av intervjuene må suppleres med hjemmeintervju for å sikre representasjon fra de høyere aldersgruppene.

Tabell 4.12: *Andeler (prosent) som benytter de ulike korttypene, fordelt på designvariant. Ungdomskort og "andre billettyper" er utelatt*

Design	7-dagerskort	Dagskort	Enkeltbillett	Flexikort	Månedskort	3-månederskort
1	9	3	6	57	14	9
2		3	23	27	30	10
3		2	11	40	25	13
4	8	6	15	31	17	13
5	7		16	39	25	11
6	5	2	22	42	11	13
13	5	5		26	37	16
14	7	4	7	18	36	18
15	4	4	4	22	33	19
16	15	15		30	20	5

©TØI 2001

## 4.7 Oppsummering

Det er blitt utarbeidet 10 ulike design for undersøkelsen, som omfatter følgende variasjoner i *design*:

- **Utvalgs- og rekrutteringsmetode.**
  - 1) telefonrekruttering (gjennomført av Gallup)
  - 2) rekruttering via Trafikantens og Oslo Sporveiers internettsider
- **Metode for datainnsamling.**
  - 1) "Tradisjonelle" hjemmeintervjuer
  - 2) egenadministrerte, PC-baserte intervjuer på tilsendt diskett
- **SP-metoder.** I undersøkelsen har følgende metoder vært benyttet:
  1. Parvise samvalg (SC)
  2. Betinget verdsetting (CV) og likeverdspris (TP)
  3. Frisch-metoden (FM)
  4. Fleksibel pakkesammensetning (POG)
- **Struktur på spørreskjema.** Rekkefølgen på SP-sekvensene er variert
- **Utforming av de enkelte SP-metodene.** Her har vi konsentrert oss om samvalgsekvensene. Følgende seks varianter har vært brukt:
  1. Standarddesign: Fire kvalitetsfaktorer med tre nivåer hver og opptil ni valg. Preferanse-usikkerhet (kan velge "Absolutt alternativ X" eller "Trolig alternativ X").
  2. Balansering av nivåene ved hjelp av CV/TP.
  3. CV/TP-balansering, samt svaralternativ "Vet ikke" og "Like gode".
  4. Standarddesign, og i tillegg et eget blokkdesign med grupper av faktorer på flere nivåer.



5. CV/TP-balansering. Ta med opptil 15 valg for å studere tretthetseffekter.
6. Redusere antall kvalitetsfaktorer fra fire til tre.

Tabell 4.13 beskriver de 10 ulike designene for spørreundersøkelsen. De fordeler seg på seks ulike designvarianter for samvalgene, to rekrutteringsmetoder og to datainnsamlingsmetoder. I tillegg er rekkefølgen på de ulike SP-sekvensene variert.

Våre tester omfatter:

- Størrelse, signifikans og fortegn på verdsettingene
- Vurdering av vanskelighetsgrad ved metodene
- Leksikografi
- 0-svar og overbud
- Preferanseusikkerhet
- Avvikende svar
- Tretthetstegn
- Intern konsistens

- Frafall
- Beskrivelser av utvalget og frafall

Vi har påvist skjevheter i utvalgene, knyttet til både rekrutteringsmetode og datainnsamlingsmetode. Utvalget som er trukket av Gallup er i stor grad representativt for befolkningen, men det er en ikke ubetydelig forskjell mellom utvalget som godtok hjemmeintervju og utvalget som fikk tilsendt disketter for egenadministrerte intervjuer.

Selv om enkelte av våre delutvalg har vært svært små, har det i seg selv ikke forårsaket store tilfeldige variasjoner i egenskaper ved respondentene og reisene de foretok. Imidlertid er det påfallende forskjeller på grunn av rekrutteringsmetode.

Frafallet blant dem som fikk tilsendt diskett for egenadministrerte intervjuer ligger jevnt på ca 50 %. Denne andelen kan reduseres ved å utvikle mer brukervennlige opplegg og løsninger som ikke avhenger av operativsystem, som internettbaserte løsninger.

Tabell 4.13: Oversikt over de ulike designnumrene. "Designvariant" viser til varianten av samvalgsekvensen

Designvariant for samvalg	Rekkefølge på SP-sekvensene	a) diskett pr post telefonrekruttert	b) diskett pr post selvrekruttert (internett)	c) hjemmeintervju forhåndsrekruttert
I	SC, POG, FM			1
II	CV/TP, SC, POG			2
III	CV/TP, SC, FM	3	13	
IV	SC, FM, CV/TP	4	14	
V	CV/TP, SC, FM	5	15	
VI	FM, SC, CV/TP	6	16	

©TØI 2001

## 5 Samvalganalyser (SC)

Dette kapitlet presenterer resultatene fra samvalg-analysene i alle de ulike designvariantene. Vi fokuserer ikke primært på selve verdsettingen av de forskjellige kvalitetsfaktorene som inngår i undersøkelsen, men ser på mulige årsaker til forskjeller i resultatene. I tillegg presenteres en serie med tester for hvor godt undersøkelsen har fungert: Leksikografi, avvikende svar og lærings-effekter og tretthetstegn.

Som nevnt tidligere blir samvalganalysen brukt til å estimere parametrene i nyttefunksjonene i en logitmodell ved hjelp av *maximum likelihood*-estimering. Våre beregninger er gjort med programpakken ALOGIT som er utviklet av HCG/Rand. Modellkjøringene produserer indikatorer for hvor god forklaringskraft modellen har:

- Log likelihood
- $\rho^2$  (w.r.t. konstant)
- t-tester for parameterestimaterne

Det vil ha lite for seg å legge stor vekt på nivået til disse indikatorene, siden vi opererer med relativt små utvalg i hver av designvariantene og dermed ikke kan vente at modellene frembringer robuste resultater. Dette gjelder spesielt de "myke" kvalitetsfaktorene informasjon, sikkerhet og leskur.

I analysene av samvalgdatabene fra de ulike valgsekvensene er det benyttet enkle logit-modeller. Logit-analysene forutsetter homogene preferanser blant respondentene og homoskedastiske variable, dvs. variable med konstant varians. Dette er strenge forutsetninger som kan påvirke verdsettingene dersom de ikke holder (jf. Sælensminde 2000). Forutsetningene kan løses på ved å estimere modeller som tar hensyn til at både preferanser og varians kan variere blant respondentene. Til tross for at mer avanserte modeller kan gi riktigere anslag på verdsettingene fra samvalgdatabene er det likevel grunn til å tro at enkle logit-modellene kan gjøre nytten for å undersøke om og i hvilken grad designet påvirker verdsettingene.

### 5.1 Verdsetting av kvaliteten på reisen

Den relative verdsettingen av enkeltegenskapene beregnes som forholdstallet mellom parameteren for egenskapen og parameteren for taksten i nyttefunksjonen. Det

kan beregnes standardavvik og t-verdier også for disse relative verdsettingene.

Standardavviket for verdsettingene beregnes som (Sjöstrand, 1999)

$$\sqrt{\left(\frac{\beta_t}{\beta_p}\right)^2 \left( \frac{\text{var}(\beta_p)}{\beta_p^2} + \frac{\text{var}(\beta_t)}{\beta_t^2} - 2 \frac{\text{cov}(\beta_p, \beta_t)}{\beta_p \beta_t} \right)}$$

der  $\beta_p$  er prisparameteren og  $\beta_t$  er parameteren for egenskapen som verdsettes.

Vi har antatt at konfidensintervallet er lik  $\pm 2$ \*standardavviket, og at dette med 95 % sannsynlighet dekker den sanne verdsettingen<sup>4</sup>.

Like viktig som de statistiske testene er en generell vurdering av størrelsen og fortegnet på de beregnede verdiene ut fra de forventningene vi har til disse på forhånd.

Tabell 5.1 gjengir de beregnede verdsettingene av egenskapene ved kollektivtransporten i Oslo fordelt på de ulike designene, innsamlings- og rekrutteringsmetodene.

Fra tabellen ser vi at det er til dels store variasjoner i verdsettingene, avhengig av designvariant, rekruttering og gjennomføring av intervjuet. Blant de "tunge" kvalitetsfaktorene reisetid, reisetid med ståplass og ventetid er det ingen beregninger med galt fortegn. Derimot er det flere eksempler på verdsettinger av "myke" kvalitetsfaktorer som får galt fortegn.

Tabell 5.2 gjengir de beregnede tidsverdiene fra 1992-undersøkelsen blant Oslos kollektivtrafikanter (Norheim 1996). I den grad vi kan si at disse verdiene representerer et "riktig" nivå på verdsettingene, kan vi gjøre oss følgende betraktninger:

- De telefonrekrutterte, egenadministrerte intervjuene (i tillegg til design 13) stemmer best overens med 1992-undersøkelsen når det gjelder tidsverdier.
- Rekruttering på internett har gitt store avvik i forhold til 1992-undersøkelsen.

<sup>4</sup> Dette er en ikke helt uproblematisk antakelse fordi de gjentatte valgene gjort av én person ikke er uavhengige. For eksempel har Lindquist-Dillen og Algers (1999) argumentert at standardavviket kan være så mye som 50 % underestimert på grunn av dette problemet.

Tabell 5.1: Beregnede verdsetninger (kroner pr time eller pr reise) av egenskapene fordelt på ulike design, rekrutteringsmetoder og datainnsamlingsmetoder<sup>+</sup>

	Design 1	Design 2	Design 3	Design 4	Design 5	Design 6	Design 13	Design 14	Design 15	Design 16
Kvalitetsfaktor	1	2	3	4	5	6	13	14	15	16
Reisetid, kr pr time	7,1	8,8	20,4	16,9	10,3	24,5	18,1	35,0	68,5	49,5
Ståplass, tillegg/time	15,0	17,2	16,8	14,7	11,3	*	21,6	12,1	10,0	*
Ventetid, kr pr time	77,0	87,6	69,4	61,1	47,7	44,4	76,8	82,7	117,6	99,8
Vektore, kr pr reise	1,5	1,1	-0,6	-1,0	-0,0	1,5	-2,5	1,0	-2,1	0,6
Nødtelefon, kr pr reise	1,1	1,2	0,3	1,0	-2,5	-0,2	0,9	1,1	-1,6	0,5
Overbygg, kr pr reise	0,8	0,2	-0,3	0,4	-4,7	*	-1,7	0,5	-7,7	*
Leskur, kr pr reise	1,9	1,6	1,2	1,3	0,5	*	2,6	1,6	0,9	*
Realtid, kr pr reise	0,9	-0,0	1,4	0,8	2,8	2,0	1,1	1,7	2,6	2,8
Rutekart, kr pr reise	1,1	1,0	-1,0	-0,3	-4,1	-1,2	-0,1	1,2	-7,0	-0,2
$\rho^2$ sekvens 1/2	0,2/0,2	0,2/0,2	0,3/0,3	0,3/0,2	0,2/0,1	0,1/0,2	0,2/0,2	0,3/0,3	0,2/0,1	0,1/0,3
Ant. respondenter	35	31	53	52	44	55	19	28	27	20

Tabell 5.1 forts.

	Design 3&13	Design 4&14	Design 5&15	Design 6&16	Rekr. Gallup	Rekruttert Internett	Hjemme-intervju	Egenadm.	Alle
Kvalitetsfaktor	3&13	4&14	5&15	6&16	Gallup	Internett	intervju	Egenadm.	Alle
Reisetid, pr time	19,8	24,2	26,9	29,1	15,3	50,1	8,0	25,6	23,3
Ståplass, tillegg/time	18,0	13,5	10,7	*	13,3	12,4	16,0	12,1	12,6
Ventetid, pr time	70,9	67,6	67,1	54,2	59,6	104,6	81,9	66,0	68,3
Vektore, pr reise	-0,9	-0,2	-0,7	1,3	0,1	-0,5	1,4	-0,2	-0,1
Nødtelefon, pr reise	0,4	1,1	-2,3	-0,0	0,3	0,5	1,2	0,1	0,3
Overbygg, pr reise	-0,4	0,4	-5,7	*	-0,4	-1,6	0,6	-0,9	-0,7
Leskur, pr reise	1,4	1,5	0,6	*	1,4	1,7	1,7	1,4	1,4
Realtid, pr reise	1,45	1,1	2,8	2,2	1,3	2,0	0,6	1,6	1,4
Rutekart, pr reise	-0,9	0,3	-5,1	-1,0	-0,7	-0,7	1,1	-1,0	-0,7
$\rho^2$ sekvens 1/2	0,3/0,3	0,3/0,2	0,2/0,1	0,1/0,2	0,2/0,2	0,2/0,2	0,2/0,2	0,2/0,2	0,2/0,2
Ant. respondenter	72	80	71	75	270	94	66	298	364

\* inngår ikke i analysen.

<sup>+</sup> Forklaring av designnumrene:

- Design 1 og 2 er hjemmeintervjuer rekruttert av Gallup
- Design 3-6 er egenadministrerte intervjuer og rekruttert av Gallup
- Design 13-15 er egenadministrerte og rekruttert på internett (selvrekuttert)
- Design 3 og 13, design 4 og 14 osv. har samme design men forskjellig rekruttering

©TØI 2001

Tabell 5.2: Verdsetninger av kvalitetsfaktorer fra 1992-undersøkelsen (beregnet fra Norheim, 1996) oppjustert til 2001-priser ved hjelp av konsumprisindeksen

	SP-Oslo 2001-priser
Reisetid	18,8
Ståplass, tillegg pr time	20,6
Ventetid	58,4
Rutekart	0,5
RTI	0,0
Leskur	0,7
Vektore	1,8
Nødtelefon	1,0

©TØI 2001

Med unntak av hjemmeintervjuene er det generelt dårlig samsvar i verdsettingen av de mykere kvalitetsfaktorene.

Vi skal ikke gå inn på en dyptgående diskusjon av hvilke beregninger som er "riktigst," men heller konsentrere oss om å identifisere egenskaper ved designet og gjennomføringen av undersøkelsen som kan ha påvirket resultatene. I drøftingene som kommer vil vi likevel i noen sammenhenger vurdere nivåene på verdsettingene i forhold til resultater fra andre undersøkelser.

### 5.1.1 Rekrutteringen påvirker resultatene

Oversikten gir klare indikasjoner på at rekrutteringsmetoden påvirker verdsettingene. De selvrekuttede

respondentene har generelt høyere, eller mer ekstrem, verdsetting av kvalitetsfaktorene enn dem som er rekruttert av Gallup. Dette kan skyldes at det selvrekrutterte utvalget ikke er representativt mht alder, kjønn, inntekt osv, men også at de selvrekrutterte respondentene sannsynligvis har et mer aktivt forhold til kollektivtransporten fordi de besøkte hjemmesidene til Sporveien eller Trafikanten. Blant annet har vi påvist at det selvrekrutterte utvalget reiser langt sjeldnere på enkeltbillett enn utvalget som Gallup rekrutterte. Dette viser igjen at omhyggelige rutiner for rekruttering er viktig, og at forsøk på å spare inn på denne biten kan medføre betydelige skjevheter i beregninger av befolkningens preferanser.

### 5.1.2 Datainnsamling

Tabell 5.3 gjengir verdsettinger og signifikanstester for henholdsvis hjemmeintervjuene og de egenadministrerte intervjuene. Blant hjemmeintervjuene har alle verdsettingene riktig (positivt) fortegn, mens de egenadministrerte intervjuene har tre verdsettinger med "feil" fortegn. Dette tyder på at det har vært problemer i forbindelse med de egenadministrerte intervjuene med samvalgsekvens 1, der de mykere kvalitetsfaktorene inngikk. Når det gjelder de mykere kvalitetsfaktorene, er parameterestimaterne relativt gode for hjemmeintervjuene, tatt i betraktning at det generelt er svake estimater i samvalgsekvens 1 for alle designvariantene. De egenadministrerte intervjuene har lavere beregnede verdsettinger, noe som kan forklare den hyppige forekomsten av insignifikante beregninger.

Tabell 5.3: Verdsettinger og signifikanstester for hjemmeintervjuer og egenadministrerte intervjuer

		Hjemme- intervju	Egen- administrert	Signifikant forskjell?
Sekvens 2	Reisetid, pr time	8*	26	Ja
	Ståplass, tillegg/time	16	12	Nei
	Ventetid, pr time	82	66	Nei
Sekvens 1	Vektere, pr reise	1	-0*	Ja
	Nødtelefon, pr reise	1	0*	Ja
	Overbygg, pr reise	1*	-1	Ja
	Leskur, pr reise	2	1	Nei
	Realtid, pr reise	1	2	Nei
	Rutekart, pr reise	1*	-1	Ja

\* Ikke signifikant forskjellig fra 0

©TØI 2001

Samtidig noterer vi at de egenadministrerte intervjuene har gitt verdsettinger av de tyngre kvalitetsfaktorene (reisetid, ventetid og ståplass) som er mer i tråd med bl.a.

anbefalingene i Minken mfl.(2000)<sup>5</sup> og beregningene fra Oslo-undersøkelsen fra 1992 (Norheim 1996). I Tabell 5.3 gjenga vi verdsettingene fra Osloundersøkelsen. Ved å sammenligne 1992-tallene med våre beregninger, finner vi at hjemmeintervjuene stemmer best overens med hensyn til de mykere kvalitetsfaktorene, mens de egenadministrerte intervjuene fungerte best for samvalgsekvens 2.

Til sammenligning er forholdet mellom verdsettingene reisetid, ventetid og tid med ståplass ganske unormale, for hjemmeintervjuene. For eksempel er ventetid beregnet til over 10 ganger så tungtveiende som reisetid. I tillegg er ikke beregningen av reisetid signifikant forskjellig fra 0, noe som indikerer at hjemmeintervju-estimaterne for samvalgsekvens 2 er svært svake. Selv når vi tar hensyn til at utvalget i hjemmeintervjuene hadde marginalt kortere gjennomsnittlig reisetid enn det øvrige utvalget, kan ikke dette forklare hele forskjellen i signifikansnivået. Verdsetting av reisetidsfaktorer er vanligvis blant de mer robuste beregningene i samvalganalyser av kollektivtrafikantenes preferanser (jf. Norheim 1996).

Dermed er det ingen entydige tegn på at de egenadministrerte intervjuene har fungert dårligere enn hjemmeintervjuene med hensyn til de beregnede verdsettingene og signifikanstestene. På noen områder har de fungert dårligere og på andre områder har de fungert bedre. Verdsettingene av de tyngre kvalitetsfaktorene har på mange måter fungert bedre i de egenadministrerte intervjuene, mens hjemmeintervjuene har fungert best for verdsettingen av de mykere kvalitetsfaktorene. Vi anbefaler derfor at det utarbeides mer brukervennlige løsninger for egenadministrerte intervjuer, f.eks. internett- eller windows-baserte løsninger, i videre forsøk med egenadministrerte intervjuer.

### 5.1.3 Antallet respondenter

Oversikten viser også at verdsettingene i noen grad konvergerer når antallet respondenter øker. I designvariantene med under 50 respondenter er det til dels store sprik i verdsettingene, men de mer ekstreme verdiene forsvinner etter hvert som antallet respondenter øker.

Tradisjonelt er kravet til en pilotundersøkelse at det er minst 30 gjennomførte intervjuer. Sammenligningen vår viser at dette kan være for lavt, hvis man ønsker å balansere nivåene i størst mulig grad i forhold til respondentenes preferanser. Vi har derfor fått et problem i design 13-16 på grunn av den lave svarprosenten og bare 19-28 respondenter i hver designvariant. Problemet er størst for

<sup>5</sup> Minken mfl. (2000, Tabell 8.1) anbefaler følgende verdier (1998-priser): Reisetid 34 kr/time, ventetid 0-7,5 minutter 61,2 kr/time. Det oppgis ingen anbefalinger for reisetid med ståplass.

de mykere kvalitetsfaktorene, slik også tidligere studier (f.eks. Norheim 1996) har vist.

Oversikten vår indikerer at minimumsantallet for en pilotundersøkelse med hell kan økes til 50. Hvor mange man velger, blir en avveining mellom kostnader og kvalitet på pilotundersøkelsen, og avhenger bl.a. av hvilke kvalitetsfaktorer som inngår i analysen og av graden av mulighet for "gjenbruk" av tidligere design for studien.

#### 5.1.4 Antallet kvalitetsfaktorer i sekvensen: Pakkeeffekt?

I design 6 og 16 tok vi ut en egenskap i hver av de to samvalgsekvensene, slik at bare tre egenskaper inngikk i hvert spill<sup>6</sup>. Det er to forhold som bør tilsi at verdsettingen av de resterende egenskapene blir høyere i disse spillene:

- *Fokuseringseffekt*. Respondentene har en bestemt mengde penger til kollektivforbedringer og fokuserer kun på det som blir tilbudt i samvalgene.
- *Avtakende grensenytte*: Generelt er det slik at jo flere egenskaper som forbedres ved en reise, jo mindre nytte har man av en ytterligere – marginal – forbedring. Et eksempel kan være reisetid og sitteplass. Hvis reisetiden reduseres fra for eksempel 15 til 5 minutter, blir nytten av økt sjanse for sitteplass redusert, fordi tiden med eventuell ståplass er redusert betraktelig.

Samtidig kan det også hende at prisvariabelen får større vekt i disse spillene fordi den blir mer synlig. Dette vil ha motsatt effekt ved at verdsettingene blir lavere.

I våre analyser har vi funnet noen signifikant forskjellige verdsettinger av reise- og ventetid mellom design 6/16 og de øvrige designene, se Tabell 5.4. Men vi har ikke identifisert noen entydig retning på effekten av antallet faktorer som inngår i samvalgsekvensen. I design 6/16 er beregningen av verdien av reisetid signifikant høyere, mens ventetid er signifikant lavere enn i de øvrige designene. Verdsettingen av realtidsinformasjon er ikke signifikant forskjellig. Vi har derfor ikke grunnlag ut fra våre data til å si noe om hvordan antallet egenskaper som inngår i samvalgsekvensen påvirker verdsettingen. Signifikansnivåene og  $\rho^2$  er omtrent like i de to beregningene av tidsverdier, mens modellen med realtidsinformasjon har noe bedre forklaringskraft for de øvrige designene.

Tabell 5.4: Verdsetting av egenskaper i design 6/16 og de øvrige designene

	Design 6/16 (N=75)	Øvrige design (N=289)
Realtidsinformasjon, pr reise	1,22	0,6
Reisetid, pr time	29,1*	22,5
Ventetid, pr time	54,2*	71,0

\* Signifikant forskjellig fra beregningen av de øvrige designene  
©TØI 2001

## 5.2 Leksikografi

Det kalles leksikografi når en respondent sorterer alle valgene i SP-sekvensen etter den samme egenskapen (f.eks. prisen). Leksikografi er et alvorlig problem i samvalganalyser, fordi vi ikke kan vite om det bidrar til å øke eller minske verdsettingen av den egenskapen det sorteres leksikografisk etter. Hvis leksikografiske valg skyldes manglende balansering av egenskapsnivåene i designet – og at respondentene derfor prioriterer denne faktoren høyere enn de andre – vil vi risikere at verdsettingen av den egenskapen det sorteres etter blir undervurdert i de senere analysene. Hvis derimot sorteringen skyldes at respondenten forenkler beslutningsprosessen ved å fokusere bare på en av faktorene og ignorere de andre, vil det medføre at vi beregner for høy relativ verdsetting av denne faktoren.

Det er derfor en vesentlig utfordring å balansere samvalgdesignet slik at andelen leksikografiske svar blir lavest mulig.

Norheim (1996) rapporterer om mellom 20 og 71 % leksikografiske svar i 1992-undersøkelsen for Oslo Sporveier og Forskningsrådet, mens Staffan Widlert (1994) fikk mellom 7 og 61 % leksikografiske svar i sin omfattende sammenligning av SP-design. I Tabell 5.5 og 5.6 har vi sammenstilt andelen leksikografiske svar og de relative verdsettingene i samvalgsekvens 1 og 2.

I samvalgsekvens 1 er det generelt svært lave andeler leksikografiske svar for de andre egenskapene enn prisen, noe som skyldes at det er relativt 'myke' kvalitetsfaktorer som inngår i spørsmålene. Det er derfor sannsynlig at for mange er det prisen som veier tyngst i dette spillet. Den høye andelen respondenter som sorterer svarene etter prisen indikerer likevel at egenskapsnivåene ikke er optimalt balanserte. Dette er en vesentlig årsak til at verdsettingene er svært usikre og ofte har negative fortegn. På denne måten gir andelen leksikografiske svar en god indikasjon på hvor vellykket undersøkelsen har vært.

<sup>6</sup> Trygghet, informasjon og pris i samvalgsekvens 1, og reisetid, frekvens og pris i samvalgsekvens 2

Tabell 5.5: *Andeler leksikografiske svar, med verdsetting av egenskaper (kr/reise av egenskap i parentes) i samvalgsekvens 1*

Design	Informasjon (Realtidsinfo)	Holdeplass (Overbygg / tak)	Trygghet (Vektore)	Pris	Totalt
Hele datasettet	1% (1,4)	0% (0,7)	1% (-0,1)	42%	45%
Design 1	0% (0,9)	0% (0,8)	0% (1,5)	40%	40%
Design 2	3% (-0,0)	3% (0,2)	3% (1,1)	35%	45%
Design 03	2% (1,4)	0% (-0,3)	0% (-0,6)	47%	49%
Design 04	2% (0,8)	0% (0,4)	0% (-1,0)	52%	54%
Design 05	0% (2,8)	0% (-4,7)	0% (-0,0)	43%	43%
Design 06	0% (2,0)	*	2% (1,5)	48%	50%
Design 13	0% (1,1)	0% (-1,7)	5% (-2,5)	32%	37%
Design 14	4% (1,7)	0% (0,5)	4% (1,0)	44%	52%
Design 15	0% (2,6)	0% (-7,7)	0% (-2,1)	26%	26%
Design 16	0% (2,8)	*	0% (0,6)	27%	27%
Design 3 og 13	1% (1,5)	0% (-0,4)	1% (-0,9)	43%	46%
Design 4 og 14	3% (1,1)	0% (0,4)	1% (-0,2)	49%	53%
Design 5 og 15	0% (2,8)	0% (-5,7)	0% (-0,7)	37%	37%
Design 6 og 16	0% (2,2)	*	2% (1,3)	42%	44%
Galluprekruttert	1% (1,3)	0% (-0,4)	1% (0,1)	45%	48%
www-rekruttert	1% (2,0)	0% (-1,6)	2% (-0,5)	33%	36%
Hjemmeintervju	2% (0,6)	2% (0,6)	2% (1,4)	38%	42%
Egenadministrert	1% (1,6)	0% (-0,9)	1% (-0,2)	43%	45%

\* Inngår ikke i analysen

©TØI 2001

Tabell 5.6: *Andeler leksikografiske svar, med verdsetting av egenskaper (kr/time i parentes) i samvalgsekvens 2*

Design	Frekvens (skult ventetid)	Reisetid	Sitteplass	Pris	Totalt
Hele datasettet	8% (68)	3% (23)	3% (13)	15%	28%
Design 1	9% (77)	0% (7)	3% (15)	26%	37%
Design 2	23% (88)	3% (9)	10% (17)	10%	47%
Design 03	2% (69)	2% (20)	0% (17)	11%	15%
Design 04	2% (61)	2% (17)	4% (15)	27%	35%
Design 05	0% (48)	0% (10)	2% (11)	11%	14%
Design 06	17% (44)	0% (25)	*	19%	35%
Design 13	5% (77)	0% (18)	11% (22)	11%	26%
Design 14	7% (83)	15% (35)	4% (12)	15%	41%
Design 15	7% (118)	4% (69)	0% (10)	0%	11%
Design 16	18% (100)	6% (49)	*	0%	24%
Design 3 og 13	3% (71)	1% (20)	3% (18)	11%	18%
Design 4 og 14	4% (68)	6% (24)	4% (13)	23%	37%
Design 5 og 15	3% (67)	1% (27)	1% (11)	7%	13%
Design 6 og 16	17% (54)	2% (29)	*	14%	32%
Galluprekruttert	8% (60)	1% (15)	3% (13)	18%	29%
www-rekruttert	9% (105)	7% (50)	3% (12)	7%	26%
Hjemmeintervju	15% (82)	2% (8)	6% (16)	18%	42%
Egenadministrert	6% (66)	3% (26)	2% (12)	14%	25%

\* Inngår ikke i analysen

©TØI 2001

Vi har ønsket å undersøke om det er noen systematikk i hvem som svarer leksikografisk. Totalt var det 35 personer, eller ca 10 % av respondentene, som sorterte etter prisen både i samvalgsekvens 1 og 2.

Tabell 5.7 sammenligner fordelingen av leksikografiske svar i henholdsvis samvalgsekvens 1 og samvalgsekvens 2. Blant dem som sorterte leksikografisk etter pris i sekvens 1 var det 24 % som også sorterte leksikografisk i spill 2, og 69 % som ikke gjorde det. Til sammenligning var det totalt 15 % som sorterte etter prisen i spill 2 (Tabell 5.6). De som svarte leksikografisk etter prisen i sekvens 1 er altså overrepresentert blant dem som sorterte etter pris i sekvens 2. Dette ser vi også nederst på tabellen: Hele 70 % av dem som sorterte etter pris i sekvens 2 gjorde det også i sekvens 1.

Samtidig, blant dem som ikke har sortert svarene i sekvens 1, er det 75 % som heller ikke svarer leksikografisk i det andre spillet, mens 25 % altså har sortert leksikografisk i spill 2. Dette er ikke ulikt de 28 prosentene, totalt sett, som har svart leksikografisk i spill 2. Vi ser også at det er en ganske stor andel – 41 % – av dem som ikke sorterte etter prisen i sekvens 2 som likevel gjorde det i sekvens 1. Dette er to eksempler på at det også er en del variasjon i hvem som sorterer leksikografisk.

Totalt sett er det med andre ord tegn til systematikk i hvem som svarer leksikografisk i de to samvalgsekvensene. Det vil være en viktig og interessant oppgave å undersøke hva som kjennetegner disse respondentene som alltid sorterer leksikografisk. Trolig er dette personer som har problemer med å mestre valgsituasjonen i samvalgsekvensene (se Norheim 1996, s. 24).

### 5.2.1 Leksikografi og designvariant

I samvalgsekvens 1 er forskjellene mellom hjemmeintervjuene og de egenadministrerte intervjuene

i andeler leksikografiske svar liten. I sekvens 2 er det en signifikant høyere andel som sorterer etter frekvens i hjemmeintervjuene, men ellers ingen vesentlige forskjeller. Det er derfor vanskelig å konkludere med at metoden for datainnsamling påvirker andelen leksikografi.

Det er flere mulige måter antallet kvalitetsfaktorer som inngår i samvalgsekvensen kan påvirke andelen leksikografiske svar. For det første gjør dette valgsituasjonen enklere ved at det er mindre informasjon å forholde seg til. For det andre er det større sjanse for at svarene blir leksikografiske ved tilfeldighet, siden det er færre egenskaper som inngår. For det tredje kan færre kvalitetsfaktorer føre til at alternativene blir dårligere balanserte, og at en enkelt faktor kan få større betydning for respondenten. Det første punktet vil medføre en lavere andel leksikografiske svar, mens de to sistnevnte fører til en høyere andel. Design 6 og 16 har bare tre kvalitetsfaktorer i samvalgsekvensene. I sekvens 1 er det ingen signifikant forskjell mellom design 6/16 og resten av datasettet. Men i sekvens 2 er det en høyere andel svar sortert etter frekvens i design 6/16 enn for hele datasettet. Og i sum er det ingen påviselig forskjell i andelene leksikografiske svar. Vi har dermed ikke identifisert noen sammenheng mellom antallet faktorer som inngår i samvalgsekvensen og andelen leksikografiske svar.

Det er helt som forventet at samvalgsekvensene med 15 valgoganger (design 5 og 15) har lavere andeler leksikografiske svar. Dette er fordi sjansen for at alle svarene skal bli sortert etter én egenskap faller fort når antallet valgoganger øker (se f.eks. Sælensminde 2000, s. 60-61).

Tabell 5.7: *Andeler respondenter som ikke sorterer /sorterer etter pris i spill 2 etter å ha gjort det samme i spill 1*

	Ikke sortert etter prisen i sekvens 1	Sortert etter pris i sekvens 1
Ikke sortert etter prisen i sekvens 2	75 %	69 %
Sortert etter pris i sekvens 2	8 %	24 %
	Ikke sortert etter prisen i sekvens 2	Sortert etter pris i sekvens 2
Ikke sortert etter prisen i sekvens 1	57 %	30 %
Sortert etter pris i sekvens 1	41 %	70 %

### 5.3 9 versus 15 valgoganger

Fem av de seks designvariantene inneholder samvalgsekvenser med 9 valgrunder, mens design 5/15 inneholder samvalgsekvenser med 15 valgoganger. I spørsmålet om hvor vanskelig respondentene syntes det var å svare på samvalgsspørsmålene er det en klar forskjell mellom de som fikk 9 og de som fikk 15 spørsmål. Tabell 5.8 viser fordelingen av svarene.

Tabell 5.8: Fordelingen av svarene på kontrollspørsmålene om hvor vanskelig det var å svare på samvalgsekvensene. Prosent

Kontrollspørsmål SC	9 spørsmål	15 spørsmål
Svært vanskelig	5,8	14,1*
Litt vanskelig	46,4	54,9
Nokså lett	43,0	25,4*
Svært lett	4,8	5,6

\* Signifikant forskjellig fra 9 spørsmål

©TØI 2001

Det er en langt større andel respondenter som opplever problemer med de lange samvalgsekvensene. Dette er ikke overraskende. Blant andre har Wardman (1998) rapportert en jevn nedgang i antallet valgoganger i samvalganalyser de siste 20 årene i Storbritannia og begrunnet det med problemene knyttet til respondentenes motstand mot, og problemer med, lange samvalgsekvenser. Derfor er antallet valgoganger stadig blitt redusert for å forbedre kvaliteten på dataene.

På TØI er det aldri benyttet mer enn 9 valgoganger for samvalganalyser ut fra den samme begrunnelsen. Denne testen med 15 valgoganger var primært for å undersøke om 9 valgoganger var en naturlig grense eller om det var mulig å øke informasjonen fra hver enkelt respondent.

Når vi ser nærmere på analyseberegningene av samvalgsekvensene med 15 valgoganger (altså design 5 og 15 i Tabell 5.1) har vi funnet at disse modellene har fungert relativt dårlig når det gjelder de vanligste indikatorene for forklaringskraft. En rekke av de mykere kvalitetsfaktorene fått galt (altså negativt) fortegn. Videre, i samvalgsekvens 2 har modellen med 15 valg den laveste  $\rho^2$ -verdien av alle modellene og generelt lave t-verdier for parameterestimaten.

Dette er klare indikasjoner på at antallet valgoganger bør ligge under 15. I stedet for å gi oss mer informasjon, risikerer vi at en økning i antallet valgoganger utover 9 snarere vil skape mer usikkerhet rundt beregningene i. Akkurat hvor mange valgoganger som er optimalt vil avhenge bl.a. av omfanget av bakgrunns-

spørsmål, andre verdsettingsspørsmål osv. I pkt. 5.7 (Tretthet) vil vi se nærmere på tretthet hos respondentene.

### 5.4 Små og store prisvariasjoner

I designene har vi ikke eksplisitt lagt inn små og store variasjoner i prisene som er presentert i samvalgsekvensene. I samvalgsekvens 1 har prisene variert med  $\pm 3$  kroner, og i sekvens 2 med  $\pm 5$  kroner i forhold til prisen respondenten betalte for reisen. For dem som reiser med 3-månederskort, ungdomskort eller er honnørreisende, er prisvariasjonene, målt i prosent, større enn for dem som reiser med enkeltbillett.

Vi har beregnet gjennomsnittspriser pr reise ut fra hvilken korttype respondenten oppga å ha brukt. Nivåene er beskrevet i Tabell 5.9. Det er verdt å merke seg at prisene pr reise for periodekort er beregnet ut fra gjennomsnittlig bruk, og at de faktiske prisene kan variere en god del. Dette innebærer at prisene, og dermed balanseringen av designet i undersøkelsen, blir korrelert med reisefrekvens (periodekortbrukere reiser oftere) og alder.

Tabell 5.9: Gjennomsnittspris pr reise for de ulike billettypene og den prosentvise endringen 3 og 5 kroner utgjør i henholdsvis sekvens 1 og 2\*

Billettype	Gjennomsnittspris	Prisendring SC-sekvens 1 ( $\pm 3$ kr)	Prisendring SC-sekvens 2 ( $\pm 5$ kr)
Enkeltbillett	20 kroner	$\pm 15\%$	$\pm 25\%$
Flexikort	16 kroner	$\pm 19\%$	$\pm 31\%$
Dagskort	14 kroner	$\pm 21\%$	$\pm 36\%$
7-dagerskort	15 kroner	$\pm 20\%$	$\pm 33\%$
Månedskort	13 kroner	$\pm 23\%$	$\pm 39\%$
3-månederskort	11 kroner	$\pm 27\%$	$\pm 45\%$
Ungdomskort (1 md.)	7 kroner	$\pm 43\%$	$\pm 71\%$
Ungdomskort (3 md.)	6 kroner	$\pm 50\%$	$\pm 83\%$
Annet	15 kroner	$\pm 20\%$	$\pm 33\%$
Honnørreisende	50% av pris		

\* Her beskrives ikke de designene som har balansert prisenivået ved hjelp av betalingsvilje i CV/TP

©TØI 2001

De som betalte lavere pris pr reise har altså fått større prosentvise endringer i prisen. Prisvariasjonen rundt enkeltbilletten er således  $\pm 25\%$ , mens den er hele  $83\%$  for ungdommer med 3-månederskort i samvalgsekvens 2. Det er derfor ikke usannsynlig at denne sistnevnte gruppen legger større vekt på prisen, og dermed relativt mindre vekt på de andre kvalitetsfaktorene. I Tabell 5.10 har vi beregnet separate verdsettinger av de tunge kvali-



tetsfaktorene for dem som betalte henholdsvis mindre enn og mer enn 14,50 kroner pr reise. Forskjellene i verdsetting er små, og bare for verdsetting av reisetid er forskjellen signifikant. Dette innebærer at vi ikke har påvist noen klar sammenheng mellom verdsetting og utgangsnivå for prisen i samvalgsekvensen.

Vi merker oss i Tabell 5.9 at for noen av periodekortene, og for ungdoms- og honnørreisende spesielt, er det en del priser som kan virke urimelig lave. En reisende med 3-månedskort vil for eksempel få oppgitt prisalternativene 8, 11 og 14 kroner pr reise i samvalgsekvens 1, mens den populære oppfatningen ofte er at enkeltbillettprisen (i vårt tilfelle kr 20) er den faktiske prisen pr reise<sup>7</sup>. Med så lave nivåer kan det være naturlig å fokusere på de andre kvalitetsfaktorene. Dette er sannsynligvis årsaken til at de som reiser billigst i mindre grad har sortert svarene etter prisen, jf. Tabell 5.10. Dette strider altså mot våre antakelser om at prisen skulle få større betydning for dem som betalte lite for reisen og som dermed fikk presentert større prosentvise forskjeller i prisen. Samtidig har Carlquist og Fearnley (2001) funnet en tilsvarende tendens: Jo høyere prisnivået i utgangspunktet er, jo mer sensitivt er markedet til én prosents endring i prisen, sannsynligvis fordi den absolutte størrelsen på den ene prosenten blir større. Der er trolig denne effekten som har veid tyngst og gitt som effekt at de som hadde størst prosentvise forskjeller i prisnivået likevel fokuserte minst på prisen.

Tabell 5.10: Verdsetting blant dem som betalte under kr 14,50 og dem som betalte over kr 14,50 for reisen sin (andel som sorterte etter egenskapen i parentes)

	Ståplass	Reisetid	Ventetid	Pris	N
Betalte under 14,50	12,9 (5%*)	21,0*(4%)	67,1(7%)	(10%*)	180
Betalte over 14,50	12,1(1%*)	26,0*(2%)	69,5(9%)	(19%*)	184

\* Signifikant forskjell mellom dem som betalte over og dem som betalte under kr 14,50 pr reise

©TØI 2001

En annen forklaring kan bygge på at for reisende med periodekort er marginalkostnaden for en ekstra reise lik 0. I dette tilfelle kan "pris pr. reise" oppfattes som noe som angår andre, eller det kan være vanskelig å forestille seg akkurat hvor mye dyrere det blir å kjøpe et periodekort.

Det er to viktige betraktninger vi kan gjøre på bakgrunn av dette:

- I designet av samvalgstudien er det viktig å vurdere om man skal benytte absolutte eller prosentvise endringer i forhold til respondentens faktiske pris pr reise. Absolutte endringer kan være bedre egnet når respondenten i utgangspunktet betalte mye for reisen. Ellers vil respondenten kunne få presentert altfor store variasjoner i prisen. Men rundt lavere prisnivåer er prosentvise variasjoner å anbefale.
- Det er viktig å vurdere nøye hvordan prisvariabelen presenteres. Trafikanter som reiser med månedskort kan ha vanskelig for å forholde seg til "pris pr reise," og med å overskue hva det innebærer at denne økes med f.eks. 2 kroner. I senere studier bør det derfor vurderes om periodekortbrukere skal få tilrettelagt forklaring.

Til tross for små utvalg i de enkelte designene, er det forholdsvis liten variasjon i gjennomsnittspris mellom de ulike designene. De effektene vi har sett her har derfor i liten grad påvirket forskjeller mellom designene.

## 5.5 Balansering ved hjelp av CV/TP

Vi har gjort forsøk med å balansere prisnivåene som presenteres i samvalgsekvensen ut fra betalingsviljen i CV/TP-sekvensen. I stedet for å definere prisvariasjonen som  $\pm 3$  og  $\pm 5$  kroner i forhold til den faktiske prisen respondenten betalte for reisen i henholdsvis samvalgsekvens 1 og 2, har vi brukt betalingsviljen for de to pakkene i CV/TP-sekvensen (se pkt. 4.2.1) som grunnlag for variasjonen rundt billettprisen. Avhengig av respondentens betalingsvilje for pakkene i CV/TP, kan dette innebære at variasjonen rundt basisprisen kan bli enten større eller mindre enn i standarddesignet.

Målet med denne tilnærmingen er å bruke informasjon fra de andre betalingsvillighetsspørsmålene til å "skreddersy" designet til den enkelte respondent. Når nivåene på kvalitetsfaktorene som inngår i samvalgsekvensen er bedre balansert i forhold til respondentens preferanser, håper vi på å gi respondentene mer tilpassede valg der deres faktiske preferanser blir mer synlige. Dette vil kunne redusere andelen leksikografiske svar. Tabell 5.11 viser andelen leksikografiske svar for henholdsvis ikke-balanserte og balanserte samvalgsekvenser. Generelt ser det ut til at balanseringen ved hjelp av CV/TP har gitt en lavere andel leksikografiske svar, selv om det bare er én av observasjonene hvor forskjellen er signifikant.

Samtidig har balanseringen ved hjelp av CV/TP ført til at en del ekstreme prisnivåer er blitt presentert for respondentene i samvalgsekvensene. Hvis en respondents betalingsvilje for pakke 1 eller pakke 2 i CV/TP var

<sup>7</sup> Vi mottok kommentarer om usannsynlige prisnivåer fra flere respondenter.

veldig høy, vil også prisvariasjonen i samvalgsekvensene bli tilsvarende stor. Som beskrevet i pkt. 4.2.1 kan det være flere årsaker til at noen har gitt ekstreme svar i CV/TP-sekvensen, f.eks. overbud. Når disse ekstreme verdiene blir brukt i samvalgsekvenser vil det medføre *dårligere* balansering av nivåene.

Tabell 5.11: Andelen leksikografiske svar fordelt på ikke-balanserte og balanserte samvalgsekvenser. Prosent

Sortert etter:	Ikke balansert	Balansert
Trygghet	1,2	1,1
Leskur	0,0	0,6
Informasjon på holdeplassen	1,2	1,1
Pris SC-sekvens 1	45,0	39,1
Sitteplass	2,2	3,5
Reisetid	3,4	1,7
Frekvens	9,6	6,4
Pris SC-sekvens 2	20,2	9,2*

\* Signifikant lavere andel enn de ikke-balanserte.

©TØI 2001

Med bakgrunn i at det nærmest ikke er noen signifikante forskjeller mellom standarddesignet og balansering av nivåer ved hjelp av CV/TP, og at metoden i noen tilfeller kan forverre balanseringen, kan vi ikke anbefale metoden i dens nåværende form. Det er likevel en klar *tendens* til lavere andeler leksikografi i designene med balansering av prisnivået. En forbedring vi vil anbefale forsøkt, er å benytte verdsettinger fra f.eks. Frisch-sekvensene istedenfor CV/TP-sekvensene til å balansere prisnivåene i samvalgsekvensen.

## 5.6 Avvikende svar i SC-sekvensene

I vårt tilfelle med to valgalternativer er sannsynligheten for å velge alternativ 1 eller 2 gitt ved:

$$P_1 = \frac{e^{U_1}}{\sum_i e^{U_i}} = \frac{1}{1 + e^{U_2 - U_1}}, \text{ og}$$

$$P_2 = 1 - P_1.$$

der

$P_i$  er sannsynligheten for valg  $i = 1, 2$ , og

$U_i$  er nyttefunksjonen for alternativ  $i$ , som vi har estimert i ALOGIT.

I denne analysen har vi lagt nyttefunksjonsberegningene fra hele datasettet til grunn, men vi har bare konsentrert

oss om SC-sekvens 2 der estimatene er relativt gode<sup>8</sup>. SC-sekvens 1 ga svært usikre estimater og vi har derfor valgt å ikke analysere sekvens 1-dataene i denne sammenhengen.

Hvis en respondent velger et alternativ som det ut fra denne sannsynlighetsberegningen er svært liten sjanse for at han/hun skal velge, kaller vi det avvikende svar. Andelen avvikende svar kan således gi en indikasjon på hvordan SC-sekvensen har fungert. Det er verdt å merke seg at det ikke nødvendigvis er snakk om *gale* svar, men at det også kan være uttrykk for avvikende preferanser. Vi har definert det som sterkt avvikende og avvikende svar hvis en respondent velger et alternativ som det er mindre enn henholdsvis 5 % og 10 % sannsynlighet for at skal velges ut fra beregningen over.

Av de 364 personene som gjennomførte undersøkelsen var det 8 personer som hadde ett eller flere sterkt avvikende svar<sup>9</sup>, eller altså ca 2,2 %. Når vi tar hensyn til at hver respondent har avgitt 9 eller 15 svar i hver SC-sekvens, har vi 16 sterkt avvikende svar blant totalt 3683 individuelle SC-valg. Til sammen er altså ca 0,4 % av svarene i vår analyse sterkt avvikende. Dette svarer godt til resultatene presentert i Norheim (1996). Norheim rapporterte mellom 0,1 og 0,7 % respondenter med sterkt avvikende svar ut fra samme definisjon.

Når vi ser på andelen avvikende svar ( $P < 10\%$ ), er det også god overensstemmelse med Norheim (1996). I vårt datasett er andelen 1,1 % (40 svar), mot 0,7–2,1 % rapportert i Norheim (1996).

Tabell 5.12 fordeler avvikende svar på de ulike designene. Design 5/15 er omtrent det eneste som har forekomster av sterkt avvikende svar ( $P < 5\%$ ) og har også den høyeste andelen avvikende svar ( $P < 10\%$ ). Det som skiller design 5/15 fra de andre designene, er at det er 15 valg i hver sekvens. Det kan tyde på at 15 valg blir for mye å forholde seg til, og at det blir vanskelig å være fokusert over 15 runder med kompliserte avveieringer.

I alle spillene er det åpnet for preferanseusikkerhet ved at respondentene har valget mellom å oppgi "trolig alternativ x" eller "absolutt alternativ x". Vi kan ta hensyn til at noen av de avvikende svarene var "trolig"-svar. Mens 16 av de avvikende svarene var "absolutt" var 24 oppgitt som "trolig". Disse 24 "trolig"-svarene er uttrykk for at respondentene har hatt problemer i valg-situasjonen. Ved å tillate preferanseusikkerhet kan vi altså se at en stor andel av de avvikende svarene skyldes at respondentene har funnet valgsituasjonen vanskelig.

<sup>8</sup> I SC-sekvens 2 inngår ståplass, reisetid, frekvens og pris.

<sup>9</sup> 1 av disse hadde 2 avvikende svar, 2 hadde 3 avvikende svar og 1 person hadde 4 avvikende svar.

Tabell 5.12: Andel avvikende svar definert som henholdsvis  $P < 5\%$  og  $P < 10\%$ , fordelt på de ulike designvariantene. Prosent

Design	Andel sterkt avvikende svar ( $P < 5\%$ )	Andel avvikende svar ( $P < 10\%$ )
1	0,0	0,6
2	0,4	1,1
3 og 13	0,0	0,8
4 og 14	0,0	0,1
5 og 15	1,4	2,6
6 og 16	0,0	0,2
Alle	0,4	1,1

©TØI 2001

Blant de øvrige designene er andelen avvikende svar generell svært liten. Utover det faktum at et høyt antall valg i hver samvalgsekvens har påvirket andelen avvikende svar, er det ingenting i som tyder på at de andre designvariantene har hatt noen innvirkning på andelen avvikende svar.

## 5.7 Læringseffekt og tretthet

Det er tidligere gjort undersøkelser av om det første valget i SC-sekvensene har større andel avvikende svar enn de resterende valgene (f.eks. Norheim 1996). Dette er i så fall et uttrykk for at respondenten trenger "innlæring" til den spesielle valgsituasjonen.

I testene har vi måttet konsentrere oss om samvalgsekvens 2, som fungerte best med hensyn til parameterestimater og verdsetting av kvalitetsfaktorer. Vi har gjort forsøk med å ekskludere den første valgongangen i sekvens 1, men dette har ikke forbedret modellen i forhold til å beholde alle valgongangene. Det har derfor ingenting for seg å vurdere forekomsten innlæringseffekter på grunnlag av dette datamaterialet.

I Tabell 5.13 sammenligner vi resultatene av tre estimeringer, henholdsvis første, andre og tredje tredjedel av samvalgsspørsmålene (spørsmål 1-3, 4-6 og 7-9) i design 1, 2, 3/13, 4/14 og 6/16<sup>10</sup>. Tabellen kan indikere om det er ulike deler av spørsmålssekvensen som har hatt større grad av forstyrrelser enn de andre.

Det er ved første øyesyn svært liten forskjell i parameterestimatenes signifikansnivå mellom de ulike segmentene, noe som tyder på at resultatene påvirkes i liten grad av læring og tretthet hos respondentene. I de siste tre spørsmålene (7-9) har vi likevel beregnet en signifikant lavere verdsetting av de tre kvalitetsfaktorene enn i de to andre segmentene. De lavere verdsettingene som fremkommer i den siste tredjedelen av spillet kan muligens

skyldes at respondentene har fokusert mer på prisen i disse valgene. Dette er i så fall et tegn på at tretthet hos respondenten kan tilkjennegis ved at de forenkler beslutningsrutinen og sorterer svarene etter prisen. Innenfor dette prosjektet har vi ikke anledning til å teste en slik hypotese<sup>11</sup>, men det kan gi viktig informasjon å få undersøkt dette nærmere.

Design 5/15 har den høyeste andelen avvikende svar, og er samtidig den SC-sekvensen med flest valgonganger (15). I Tabell 5.14 har vi beskrevet fordelingen av de avvikende svarene i design 5/15 på valgnummer. Det er interessant å merke seg at hyppigheten avvikende svar er høyest i den midterste tredjedelen av valgene (sekvens 6-10 med 9 prosent i gjennomsnitt). Samtidig er det omtrent like mange avvikende svar i den første og siste tredjedelen (henholdsvis 5 og 6 prosent i gjennomsnitt). Ut fra en slik betraktning er det ingen indikasjoner på at det er utpregede effekter av tretthet i svarene, selv i spillet med 15 valgonganger.

I de 9 første valgsekvensene i design 5/15 har vi funnet en høy andel avvikende svar, ca 1,7 % (ikke vist i tabellen her). Til sammenligning varierer andelen mellom 0,1 og 1,1 % i de andre designene. Den høyere andelen avvikende svar også innenfor de 9 første valgene kan skyldes tretthet fra den forutgående samvalgsekvens 1, som også hadde 15 valgsekvenser i design 5/15. Det kan i så fall tyde på at en SC-sekvens med mer enn 9 valg også vil kunne påvirke kvaliteten på senere SC-sekvenser innenfor samme intervju. Man skal derfor være forsiktig med å bruke lange SC-sekvenser (flere enn 9 valg) selv i undersøkelser som hovedsakelig har kortere spill (9 eller færre valg).

### 5.7.1 Logit scaling for å identifisere lærings- og tretthetseffekter

For å teste for om variansen endres utover i samvalgsekvensen (heteroskedastisitet, som følge av læring/tretthet), har vi gjennomført en *logit scaling*-analyse som både identifiserer forekomsten av denne typen heteroskedastisitet og korrigerer for dette. Analysemetoden er identisk med Sælensminde (2001f), og bygger på Bradley og Daly (1994) (se disse for en mer utførlig beskrivelse av metoden.). Slik vi har brukt skaleringsmetoden er skaleringsfaktoren fastsatt til 1 for første valg i valgsekvensen og variansen til senere valg inngår i de andre skaleringsfaktorene inverst i forhold til variansen i første valg. Det vil si at dersom senere valg har større varians blir skaleringsfaktoren mindre enn 1 og dersom senere valg har mindre varians blir skaleringsfaktoren større enn

<sup>10</sup> Siden designe 5 og 15 har 15 valg i samvalgsekvensen, er disse utelatt.

<sup>11</sup> En slik test forutsetter programmering av en helt ny type tester for leksikografi.

1. T-verdiene er beregnet slik at de viser om de estimerte skaleringsfaktorene er signifikant forskjellig fra 1. En T-verdi større enn 2 viser dermed om vi har en signifikant endring av variansen utover i valgsekvensen.

Tabell 5.15 viser beregningen av en modell som inneholder en skaleringsfaktor for hvert av de 9 valgene

(designvariant 3, 5 og 6 er ekskludert). I forhold til en modell uten skaleringsparametere gir denne modellen ikke noen signifikant forbedring i forklaringskraft, men den hjelper oss med å påvise forekomsten av heteroskedastisitet.

Tabell 5.13: Sammenligning av første, andre og tredje tredjedel av SC-valgene. (Design 5 og 15, som har 15 valg, er utelatt.)

		Tid Ståplass	Reisetid	Frekvens	Pris	Konstant	$\rho^2$
Spm 1,2,3	Estimate	-0,0441	-0,0755	-0,1080	-0,1780	-0,0010	0,19
	Std. Error	0,0076	0,0114	0,0095	0,0156	0,0792	
	"T" Ratio	-5,8	-6,6	-11,5	-11,4	0,0	
	Verdsetting	14,85 kr/t	25,44 kr/t	73,08 kr/t ventetid			
Spm 4,5,6	Estimate	-0,0503	-0,0662	-0,1120	-0,1900	-0,0017	0,23
	Std. Error	0,0073	0,0100	0,0096	0,0164	0,0809	
	"T"Ratio	-6,9	-6,6	-11,7	-11,5	0	
	Verdsetting	15,91 kr/t	20,91 kr/t	71,06 kr/t ventetid			
Spm 7,8,9	Estimate	-0,0384	-0,0510	-0,0972	-0,2060	0,0026	0,20
	Std. Error	0,0064	0,0105	0,0099	0,0176	0,0791	
	"T" Ratio	-6	-4,9	-9,9	-11,7	0	
	Verdsetting	11,21 kr/t	14,88 kr/t	56,72 kr/t ventetid			

©TØI 2001

Tabell 5.14: Fordeling av de avvikende svar i design 5/15 på de forskjellige valgsekvensnumrene. Prosent

Valgnummer	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Sum
Andel avvikende	0	7	4	11	4	18	7	7	7	7	4	7	4	11	4	102*

\* Er ikke lik 100% pga avrunding i tabellen.

©TØI 2001

Tabell 5.15: Skaleringsprosedyre for å identifisere og korrigere for heteroskedastisitet forårsaket av lærings- og tretthetseffekter

	Estimat	T-verdi i forhold til 1
Valg 1	1,000	
Valg 2	0,739	1,121
Valg 3	0,640	1,791
Valg 4	0,745	1,238
Valg 5	1,128	0,417
Valg 6	0,608	2,242
Valg 7	0,596	2,233
Valg 8	0,699	1,506
Valg 9	0,490	2,929

©TØI 2001

I forhold til valg nr. 1 går det frem av tabellen at det er en signifikant økning i variansen i valgene 6, 7 og 9, noe

som indikerer tretthetstegn hos respondentene mot slutten av samvalgsekvensen. Diss forholdsvis tidlige tegnene til tretthet hos respondentene skyldes sannsynligvis at vi ser på samvalgsekvens nummer 2, og vi dermed egentlig snakker om valgomgang nummer 15-18 i samvalgsekvensen.

Bare i valg nr. 5 er det beregnet lavere varians (dog ikke signifikant) enn i den første valgomgangen. Dette kan tolkes som om det ikke er tegn til læringseffekter i datamaterialet fordi det i valg nummer 2-4 er høyere varians enn i den første valgomgangen.

## 5.8 Preferanseusikkerhet i samvalg

Tradisjonelt har respondentene i samvalg (parvise valg) fått valget mellom to reiser og har kun fått to svaralternativ; dvs. de er bedt om å fortelle om de foretrekker reise A eller reise B. De har altså ikke fått anledning til å for-

telle hvor sikre de er på at de foretrekker reise A eller B. (Eller hvor mye bedre f.eks. A er enn B. Dette kan for øvrig anses som to sidere av samme sak.)

Med den tradisjonelle varianten, må altså respondenten velge enten reise A eller B selv i tilfeller der han/hun ikke er sikker på hvilket valg som er ”det beste”. I slike tilfeller der respondenten ikke er sikker og velger mer eller mindre ”tilfeldig”, vil valget ikke nødvendigvis gjenspeile respondentens preferanser. Dette kan dermed være årsak til at noen valg er motstridende (inkonsistente) med hovedtyngden av respondentens valg.

Ved å gi respondentene mulighet til å tilkjenne hvor sikre de er på sine valg kan dette kanskje gjøre det lettere for respondentene i de tilfeller der de er i tvil om hva de skal velge. Dette kan dermed kanskje redusere problemene med inkonsistente valg (jf. Sælensminde 2000), redusere ”støyen” i datamaterialet og gi sikrere estimeringsresultater.

I datamaterialet som inkluderer preferanseusikkerhet ga vi respondentene enten fire eller seks svaralternativ da de skulle velge mellom reise A og reise B. I tilfellet med fire svaralternativ var disse:

- 1) ”absolutt A”,
- 2) ”trolig A”,
- 3) ”trolig B”, og
- 4) ”absolutt B”.

I tilfellet med seks svaralternativ (design 3 og 13) var disse:

- 1) ”absolutt A”,
- 2) ”trolig A”,
- 3) ”A og B like gode”,
- 4) ”trolig B”,
- 5) ”absolutt B”, og
- 6) ”vet ikke”.

Tabell 5.16 viser hvordan respondentenes valg fordelte seg på de fire eller seks svaralternativene i valgsekvens-

sene. I tabellen er ”absolutt A” og ”absolutt B” slått sammen i en kolonne og ”trolig A” og ”trolig B” slått sammen i en kolonne.

Tabell 5.16 viser at mellom 47 og 67 % av de kollektivreisendes valg har vært sikre valg. De kollektivreisende synes å være sikrere på sine valg i valgsekvens 1 (faktorer relatert til holdeplassen) enn i valgsekvens 2 (faktorer relatert til selve reisen). Dette kan skyldes at faktorene i valgsekvens 2 er mer vanlige å vurdere i forbindelse med en kollektivreise og bedre balansert enn faktorene i valgsekvens 1. Dersom de kollektivreisende også får muligheten til å tilkjenne ”like gode” og ”vet ikke” (design 3) synker andelen sikre valg. Andelen sikre valg er mindre når de kollektivreisende får valgsekvenser med tre faktorer (design 6) enn når de får valgsekvenser med fire faktorer.

### 5.8.1 Segmentering av datamaterialet i sikre og mindre sikre valg

Vi vil her presentere resultater fra modeller der datamaterialet er segmentert i sikre og mindre sikre valg ut fra om respondentene har valgt at de ”absolutt” ville velge reise A eller B eller ”trolig” ville velge reise A eller B. Resultatene fra en slik segmentering kan imidlertid være avhengig av både:

- antall svaralternativ
- hvilke faktorer som er med i sekvensene
- antall faktorer i valgene

Vi har data som gjør at vi kan undersøke preferanseusikkerhet i alle disse dimensjonene og resultatene presenteres i de neste seksjonene. I analysene av samvalgdatabaene fra de ulike valgsekvensene er det benyttet enkle logit-modeller.

Tabell 5.16: Andel av respondentenes valg fordelt på de fire eller seks svaralternativene som ble gitt i ulike valgsekvenser (spill)

Valgsekvens	Antall faktorer	”Absolutt”	”Trolig”	”Like gode”	”Vet ikke”
Design 1, 2, 4, 5, spill 1 (n=2377)	4	67,7	32,3		
Design 1, 2, 4, 5, spill 2 (n=2375)	4	56,2	43,8		
Design 6, spill 1 (n=644)	3	57,9	42,1		
Design 6, spill 2 (n=660)	3	47,0	53,0		
Design 3, spill 1 (n=648)	4	60,3	32,9	3,9	2,9
Design 3, spill 2 (n=648)	4	50,0	41,2	4,0	4,8

### 5.8.2 "Sikre" valg gir bedre estimater

Resultatene fra valgsekvens 2 i Tabell 5.18 viser at uansett om datamaterialet brukes samlet eller segmenteres i sikre og mindre sikre valg blir alle parameterestimaterne signifikante og med riktig fortegn. Det er imidlertid klart at modellen med bare sikre valg har betydelig større forklaringskraft (høyere  $\rho^2$ ) enn modellen med mer usikre valg, men ikke så mye mindre varians i parameterestimaterne. Det må her påpekes at dette i seg selv ikke gir grunnlag for å ekskludere de mer usikre valgene.

Tabell 5.19 viser imidlertid at verdsettingene blir nokså forskjellige i de tre modellene. Det er interessant å observere at verdsetting av kvalitetsfaktorene i

modellen med bare sikre valg blir lavere enn i de modellene som inkluderer også de usikre valgene. Dersom det er slik at sikre valg i større grad er konsistente enn usikre valg, er det god overensstemmelse mellom disse resultatene og Sælensminde (2000) der resultater med verdsetting fra respondenter med konsistente valg ble sammenlignet med verdsetting fra respondenter med ett eller flere inkonsistente valg. I Sælensminde (2000) diskuteres det også om det å gi større forskjell på valgalternativene, og på den måten få flere valg respondenten er sikker på, kan redusere problemene med inkonsistente valg.

Tabell 5.17: Resultater fra valgsekvens 1. Design 1, 2, 4 og 5 med 4 svaralternativ og 4 faktorer

Parameter	Hele datamaterialet (N=2377)		Bare "absolutt-valg" (N=1609)		Bare "trolig-valg" (N=768)	
	Estimat	T-verdi	Estimat	T-verdi	Estimat	T-verdi
Takst	-0,2660	-17,4	-0,4563	-18,5	-0,0388	-2,5
Trygghet nivå 1: Vektore	0,0117	0,1	0,3429	2,3	-0,3972	-2,6
Trygghet nivå 2: Nødtelefon	0,1205	1,6	0,2829	2,8	-0,1232	-1,1
Værbeskyttelse nivå 1: Overbygget holdeplass	-0,2497	-2,7	-0,1952	-1,5	-0,3082	-2,3
Værbeskyttelse nivå 2: Leskur	0,3469	4,0	0,5398	4,5	0,1334	1,0
Info nivå 1: Realtids-informasjon om ankomsttid	0,1383	1,4	0,3133	2,3	-0,0867	-0,6
Info nivå 2: Rutekart	-0,1966	-2,3	0,0112	0,1	-0,3182	-2,3
Konstant	0,0373	0,8	0,0823	1,2	-0,0337	-0,5
Rho-squared ( $\rho^2$ )	0,1914		0,3587		0,036	
Log-likelihood	-1332,2051		-714,7965		-512,6695	

©TØI 2001

Tabell 5.18: Resultater fra valgsekvens 2. Design 1, 2, 4 og 5 med 4 svaralternativ og 4 faktorer

Parameter	Hele datamaterialet (N=2375)		Bare "absolutt-valg" (N=1335)		Bare "trolig-valg" (N=1040)	
	Estimat	T-verdi	Estimat	T-verdi	Estimat	T-verdi
Takst	-0,1213	-15,6	-0,1758	-15,0	-0,0652	-6,5
Sitteplass halve veien	0,3307	4,8	0,3820	3,9	0,2755	2,8
Sitteplass hele veien	0,5311	5,7	0,6023	4,5	0,4112	3,1
Reisetid	-0,0561	-9,6	-0,0533	-6,2	-0,0498	-6,3
Tid mellom avgangene	-0,0682	-12,9	-0,0899	-11,6	-0,0433	-6,0
Konstant	-0,0203	-0,5	0,0611	1,0	-0,0975	-1,5
Rho-squared ( $\rho^2$ )	0,1247		0,2164		0,0556	
Log-likelihood	-1440,6233		-725,0545		-679,3449	

©TØI 2001

### 5.8.3 Tester av om sikre valg er forskjellige fra mer usikre valg

For å teste om det er signifikante forskjeller mellom sikre valg og mindre sikre valg har vi brukt en "likelihood ratio test" (LR-test). Dette er en test for å undersøke om de estimerte parameterne i de alternative modellene ("absolutt-valg" og "trolig-valg") er statistisk forskjellige eller om de er like og om vi dermed kan estimere en felles ("pooled") modell for hele datamaterialet. Teststatistikken i LR-testen er  $\chi^2$  fordelt med antall frihetsgrader

lik antall restriksjoner (dvs. summen av antall parametre som estimeres i de segmenterte modellene minus antall parametre som antas å være like og estimeres i den felles modellen).

LR-testen i Tabell 5.20 viser i de fleste tilfeller at sikre valg er statistisk forskjellige fra mindre sikre valg (test statistikk større enn  $\chi^2$ ,  $P=0,05$ ). Spesielt sikre resultater får vi i de tilfellene der datamaterialet er størst. Bare design 3, sekvens 2 har ikke signifikante forskjeller mellom de sikre og de mindre sikre valgene.

Tabell 5.19: Verdsetting av kvalitetsfaktorer fra valgsekvens 2. Design 1, 2, 4 og 5 med 4 svaralternativ og 4 faktorer

	Hele datamaterialet (N=2375)	Bare "absolutt-valg" (N=1335)	Bare "trolig-valg" (N=1040)
Sitteplass halve veien	2,73	2,17	4,22
Sitteplass hele veien	4,38	3,54	6,30
Reisetid	27,77	18,20	45,83
Tid mellom avgangene	33,74	30,67	39,80

©TØI 2001

Tabell 5.20: LR-tester av om sikre valg er statistisk forskjellige fra mindre sikre valg for ulike modeller/design

Valgsekvens	Antall faktorer	Modell	Test-statistikk	Antall frihetsgrader	$\chi^2$ , $P=0,05$	Signifikant forskjell?
Design 1, 2, 4, 5, spill 1 (n=2377)	4	Tabell 3.1	209,48	8	15,51	✓
Design 1, 2, 4, 5, spill 2 (n=2375)	4	Tabell 3.2	72,45	6	12,59	✓
Design 6, spill 2 (n=660)	3	Tabell 3.5	28,27	4	9,49	✓
Design 3, spill 2 (n=648)	4	Tabell 3.8	11,85	6	12,59	-

©TØI 2001

## 5.9 Oppsummering samvalganalyser

Det er helt vesentlig med omhyggelige rutiner for rekruttering av respondenter. Selvrekruttering (i vårt tilfelle på internett) gir store skjevheter i verdsettingene, og utvalget representerer i liten grad befolkningen.

Når det gjelder innsamlingen av data har vi ikke funnet noen tydelige indikasjoner på at hjemmeintervjuer fungerer bedre eller dårligere enn egenadministrerte intervjuer. Verdsettingene av de "tyngre" kvalitetsfaktorene (reisetid, frekvens og sitteplass) har på mange måter fungert bedre i de egenadministrerte intervjuene, mens hjemmeintervjuene har fungert best for verdsettingen av de "mykere" kvalitetsfaktorene. Hjemmeintervjuer er svært ressurskrevende, og vi mener det er et stort potensial for sparing uten kvalitetsmessig forverring av dataene ved å gjennomføre egenadministrerte intervjuer, f.eks. over internett.

Vi har ikke identifisert virkningen av antallet kvalitetsfaktorer som inngår i samvalgsekvensen på verdsettingene. Når det gjelder andelen leksikografiske svar, har antallet faktorer hatt en nøytral virkning. Vi har derfor konkludert med at 3 eller 4 faktorer har hatt lite å si i forhold til verdsettingsresultatene i våre data.

Leksikografi er et vesentlig problem i samvalganalyser. I samvalgsekvens 1, som omfattet de "mykere" kvalitetsfaktorene, er det mellom 26 og hele 54 % av respondentene som har sortert svarene leksikografisk. Samtidig har denne sekvensen fungert dårlig med hensyn til estimering av kvalitetsfaktorene. En bedre balansering av nivåene på kvalitetsfaktorene vil kunne forbedre analysene betraktelig. Andelen leksikografiske svar er derfor en viktig indikator for hvordan undersøkelsen har fungert, og bør inngå som standard rapportering av SP-undersøkelser.

Vi har funnet indikasjoner på en viss systematikk i de leksikografiske svarene. For eksempel, blant dem som sorterte etter prisen i samvalgsekvens 2, var det 70 % som også gjorde det i sekvens 1. Dette kan bety at det er en del av respondentene som ikke takler valgsituasjonen i samvalgene. Vi anbefaler at det gjøres ytterligere analyser av disse respondentene for å se om de har felles karakteristika, slik at det kan gjøres mulig å fange dem opp og tilrettelegge undersøkelsen til deres situasjon.

Det er en utfordring å beskrive prisvariabelen for respondentene. Det kan f.eks. være vanskelig å forholde seg til "pris pr reise" når de bruker månedskort. Dette har ført til at noen respondenter har rapportert lite troverdige prisalternativer.

Videre er det en utfordring å definere et riktig nivå på variasjonen rundt prisen pr reise. Prosentvis variasjon kan gi store utslag for dem som reiser på dyre billetter, mens absolutte endringer gir store utslag for dem som reiser på billigere billettyper. Våre analyser tyder på en større sensitivitet for prisendringer som er store i absolutte tall, enn for store prosentvise endringer.

Det er nærmest ikke noen signifikante forskjeller i verdsettingene mellom standarddesignet og designene der prisvariabelen er balansert ut fra svarene i den forutgående CV/TP-sekvensen. I tillegg kan CV/TP-balanseringen i noen tilfeller forverre designet, og vi kan derfor ikke anbefale metoden i dens nåværende form. Det er likevel en klar tendens til lavere andeler leksikografi i designene med balansering av prisnivået ved hjelp av CV/TP. En forbedring vi vil anbefale forsøkt, er å benytte verdsettinger fra f.eks. Frisch-sekvensene istedenfor CV/TP-sekvensene til å balansere prisnivåene i samvalgsekvensen.

Utover lengden på samvalgsekvensene, målt i antallet valgrunder, er det ingen tegn til at de ulike designene som inngår i analysene har påvirket forekomsten av avvikende, eller sterkt avvikende, svar. I samvalgsekvenser med 15 valgrunder er det en signifikant større andel avvikende svar enn i sekvenser med 9 valgrunder.

I samvalgsekvensene med 15 valg fant vi ingen tegn til at de avvikende svarene var konsentrert i de siste valg-omgangene. Derfor er det trolig at avvikene er et resultat av respondentenes tretthet fra samvalgsekvens 1, som respondentene svarte på i forkant av samvalgsekvensen vi har analysert på. Tretthet påvirker i så fall ikke bare den enkelte samvalgsekvensen, men også senere samvalgsekvenser i samme intervju. Man skal derfor være forsiktig med å bruke lange SC-sekvenser (flere enn 9 valg) selv i undersøkelser som hovedsakelig har kortere spill (9 eller færre valg).

Ved å åpne for preferanseusikkerhet har vi påvist at en stor del av de avvikende svarene skyldes at respondentene har hatt problemer med å takle den vanskelige valgsituasjonen, og ikke at de nødvendigvis har hatt sterkt avvikende preferanser.

I analysene av tretthetstegn i samvalgsekvenser med 9 valgomganger har vi identifisert en større fokus på pris og dermed lavere verdsetting av kvalitetsfaktorene i de siste tre spørsmålrundene enn i de 6 forutgående spørsmålene. Sortering etter pris (eller en annen egenskap) kan derfor være et tretthetstegn. Vi vil anbefale at dette i fremtiden undersøkes nærmere ved sammenligning med andre datasett.

Tretthetstegn er også identifisert ved hjelp av *logit-scaling*. Denne metoden viser at variansen øker i de siste 3-4 valgomgangene i samvalgsekvensene med 9 valg. Når tretthetstegn viser seg allerede etter 5-6 valg-omganger, skyldes dette tretthet fra foregående samvalgsekvenser.

Segmentering av datamaterialet i sikre og mindre sikre valg tyder på at dette er signifikant forskjellige data og at de enten ikke bør inngå i samme modell eller ikke inngå på samme måte i modellen. Basert på de modellene der vi har "et tilstrekkelig stort datamateriale" kan vi konkludere at sammenligner vi modellene med sikre og mindre sikre valg, gir modellene med respondentenes sikre valg:

- Bedre forklaringskraft
- Mer signifikante parameterestimater
- Riktig fortegn på signifikante parameterestimater (der dette tidligere var et problem)
- Lavere verdsettinger av kvalitetsfaktorene

I det videre arbeidet relatert til preferanseusikkerhet i samvalgdataba bør følgende undersøkes:

- Kan skaleringsmodeller brukes til å modellere preferanseusikkerheten på en bedre måte enn segmentering der usikre valg forkastes?
- Vil en type ad hoc-vektning, avhengig av hvor sikker respondenten har vært på sine valg, kunne brukes i analysene?
- Vil bruk av kontrollspørsmål kunne gi bedre innsikt i hvordan respondentene har brukt muligheten til å tilkjenne preferanseusikkerhet?
- Er valgalternativet "like gode" et for krevende alternativ? Har respondentene virkelig så presise preferanser at "like gode" er utrygg for noe annet enn "vet ikke"?



## 6 Samvalg og kontekst: Blokkdesign

Vi vil se på hva som er riktig kontekst for verdsettingen i hypotetiske valgstudier dersom vi ønsker å finne verdien av å gjøre forbedringer i flere goder samtidig. For eksempel dersom et kollektivselskap ønsker å finne trafikantenes nytte (målt som betalingsvillighet) av å forbedre både:

- Reisetid
- Tid mellom avgangene
- Tilgjengelighet til sitteplass
- Sikkerhet på holdeplassen
- Leskur/overbygg på holdeplassen
- Informasjon om ankomsttider på holdeplassen

Spørsmålet er altså hvordan en skal måle trafikantenes nytte av en generell bedring av kollektivtilbudet, men der en i størst mulig grad ønsker å forbedre de forhold som gir størst nytteøkning pr krone. En trenger altså også en avveining mellom kvalitetsfaktorene. Kan en f.eks. forvente å få samme relative verdsettelser dersom:

1. de tre første faktorene i listen over verdsettes som enkeltfaktorer i én valgsekvens, og de tre siste verdsettes i en annen sekvens?
2. alle seks faktorene var verdsatt (veid mot hverandre) i samme sekvens?
3. faktorene var gruppert i blokker (tre og tre) og blokkene var veid mot hverandre i samme sekvens?

Svaret på disse spørsmålene er nei. Fordi dette er verdsettelser i ulike kontekster, kan en ikke forvente at betalingsvilligheten for forbedring av tre enkeltfaktorer innenfor en sekvens kan summeres og gi tilsvarende verdsetting som dersom disse tre faktorene var veid mot andre faktorer enkeltvis eller som en blokk. Dette har vi tidligere omtalt som *pakkeeffekter*, jf. Wardman (1999b).

Konteksten for verdsettingen er altså helt vesentlig og legger begrensninger med hensyn til hvordan resultatene kan brukes, dvs. hvilket gyldighetsområde verdsettingene har. For å si noe om resultatenes gyldighetsområde må en se på hvilke *endringer* som er presentert for respondentene og hvilke *forutsetninger* som ligger til grunn for de valgene respondentene har gjort.

Endringene (nivåene for hver enkelt faktor) som er presentert for respondentene kan være forholdsvis store (f.eks.  $\pm 25\%$  i reisetiden). Den marginale betalingsvilligheten for endring i enkeltfaktorene som beregnes ut

fra respondentenes valg kan altså ha et forholdsvis stort gyldighetsområde dersom de endringer som er presentert for respondentene er store. Men følgende to forhold legger likevel viktige begrensninger på gyldighetsområdet<sup>12</sup>:

1. Det forutsettes at faktorer som ikke er med i valgsekvensen ikke endres (respondentene er forelagt partielle endringer, dvs. ”alt annet likt”),
2. Innen valgsekvensen er respondentene forelagt endringer som går i ulike retninger, dvs. Ikke forbedringer eller forverring i alle faktorene samtidig.

Forutsetningen om ”alt annet likt”, at vi altså ser på partielle endringer, og at endring av alle faktorene i samme retning vanligvis ikke har vært inkludert i valgsekvensen, har betydning for hvordan resultatene kan brukes. Endring i en faktor som er substitutt (erstatte) eller komplementær (forsterker) kan altså påvirke både den relative og absolutte verdsettingen av faktorene i en sekvens. Og i tillegg kan altså endring av mange faktorer i samme retning også påvirke den absolutte verdsettingen.

I Sælensminde (2000) er det vist at ved bruk av metodene betinget verdsetting eller likeverdpris (CV/TP) får vi betydelig lavere verdsetting av ikke-markedsgoder når disse verdsettes i pakker (blokker) enn når de verdsettes separat.

Spørsmålet om hvordan ikke-markedsgoder skal verdsettes er altså avhengig av hvordan verdsettingsresultatene skal brukes. Det er dette som avgjør om det er mest riktig å verdsette de(n) forbedring(er) som en har tenkt å iverksette som en del av en samlet blokk og/eller i valgsekvenser der faktorene veies mot hverandre enkeltvis? Ved hjelp av *Stated Choice* (SC) data der faktorene avveies i både separate valgsekvenser og i en sekvens der de inngår som del av en større blokk vil vi her undersøke om og hvor mye verdsetting av ikke-markedsgoder ved SC påvirkes av konteksten for verdsettingen..

<sup>12</sup> Kottenhoff og Schmidt (2000) lister dessuten opp en rekke forhold relatert til hvordan respondentene gjennomfører valgoppgaven som også kan være grunner for at verdsetting i en pakke gir forskjellig resultat fra separat verdsetting.

## 6.1 Test av kontekst og blokkdesign ved hjelp av tre valgsekvenser

I denne undersøkelsen inngikk tre valgsekvenser for å kunne teste i hvilken grad konteksten påvirker verdsettingsresultatene fra SC. De to første sekvensene har vi beskrevet tidligere. I valgsekvens 3 inngikk de samme faktorene som i sekvens 1 og 2, men denne gang ble faktorene grupper i blokker slik at endring i faktorene relatert til holdeplassen kunne veies opp mot endring i faktorene relatert til reisen. Hvordan nivåene for faktorene var sammensatt i de ulike blokkene er vist i Tabell 6.1. I alle valgsekvensene inngikk også endringer i taksten. Blokkene ble altså dannet slik at enten inngikk alle faktorene på holdeplassen med beste nivå (nivå 1), nest beste nivå (nivå 2) eller dårligste nivå (nivå 3), og tilsvarende for faktorene forbundet med reisen. Nivå 1 for holdeplassen og reisen (beste nivå for alle faktorene fra A til F) kan da f.eks. i ett av valgene respondentene får veies opp mot dårligste nivå mht taksten. Dvs. om en er villig til å betale 5 kroner mer for en reise som samtidig gir forbedringer i alle de seks faktorene vi har med i de to blokkene.

## 6.2 Resultater fra de ulike valgsekvensene

I undersøkelsen av kontekstens betydning for verdsettingen fra samvalgdataba inngikk 80 respondenter. Disse ble gitt maksimalt 9 valg i hver valgsekvens. Dersom alle

respondentene hadde svart på alle 9 valgene, hadde vi hatt 720 valg som kunne inngå i analysene. Av ulike grunner har ikke alle gjort 9 valg, så antall valg i analysene er noe mindre enn dette.

### 6.2.1 Valgsekvens 1 og 2: Holdeplassen og reisen

Resultatene fra valgsekvens 1, som er presentert i Tabell 6.2, viser at ikke alle parameterne er signifikant forskjellig fra 0. Det gjelder vektere, som altså ikke oppleves som en signifikant forbedring når respondentene allerede har nødtelefoner. Det gjelder overbygget holdeplass og det gjelder rutekart på holdeplassen. (At overbygget holdeplass ikke er blitt signifikant kan skyldes at dette også er gitt til bussreisende og ikke bare reisende med T-banen.)

Faktorene som inngikk i valgsekvens 1 ser ut til å ha vært ubalansert i den forstand at prisforskjellene ( $\pm 3$  kr) trolig har vært for store sammenlignet med endringer i de andre faktorene. Mindre prisforskjeller kunne trolig også gitt signifikant parameterestimat for vektere.

Når vi beregner verdsettingen av de kvalitetsfaktorene som er signifikante fra valgsekvens 1, finner vi nødtelefoner kr 1,05 pr reise, leskur kr 1,47 pr reise og realtidsinformasjon kr 1,45 pr reise.

Tabell 6.1: Valgsekvens 3 – Holdeplassen og reisen

Kvalitetsfaktor	Nivå 1	Nivå 2	Nivå 3
Takst	- 5 kr	Basis*	+5 kr
Holdeplassen	A. Nødtelefon + vektere som patruljerer	A. Nødtelefon	A.
	B. Større overbygg/tak	B. Leskur	B.
	C. Rutetabell + Rutekart + Info om faktisk ankomsttid	C. Rutetabell + Rutekart	C. Rutetabell
Reisen	D. Sitteplass hele veien	D. Sitteplass halve veien	D. Ståplass hele veien
	E. Reisetid: -25 %	E. Reisetid: Basis*	E. Reisetid: +25 %
	F. Tid mellom avgangene: -50 %	F. Tid mellom avgangene: Basis*	F. Tid mellom avgangene: + 10 min

\*Basis er det respondenten oppgir i tidligere spørsmål

Tabell 6.2: Resultater fra valgsekvens 1. (N=718)

Parameter	Estimat	T-verdi
Takst	-0,3919	-12,4
Trygghet nivå 1: Vektore	0,3343	1,4
Trygghet nivå 2: Nødtelefon	0,4103	3,1
Værbeskyttelse nivå 1: Overbygget holdeplass	0,1602	0,8
Værbeskyttelse nivå 2: Leskur	0,5775	3,5
Info nivå 1: Realtidsinformasjon om ankomsttid	0,5454	2,6
Info nivå 2: Rutekart	0,1330	0,8
Konstant	0,0049	0,1

©TØI 2001

Resultatene fra valgsekvens 2 (Tabell 6.3) viser at nivåene for kvalitetsfaktorene i denne sekvensen har vært forholdsvis godt balansert. Alle parameterne er signifikante og følgende verdsetninger kan beregnes. Sitteplass hele veien 6,99 kr/reise, sitteplass halve veien 3,08 kr/reise, (reduisert) reisetid 0,58 kr/min og (reduisert) tid mellom avgangene 0,58 kr/min.

Tabell 6.3: Resultater fra valgsekvens 2 (N=717)

Parameter	Estimat	T-verdi
Takst	-0,1729	-11,0
Sitteplass hele veien	1,2090	5,2
Sitteplass halve veien	0,5327	4,2
Reisetid	-0,1009	-8,0
Tid mellom avgangene	-0,1011	-8,8
Konstant	-0,0866	-1,0

©TØI 2001

### 6.2.2 Valgsekvens 3: Blokkvis avveining av kvalitetsfaktorer relatert til både holdeplassen og reisen

Tabell 6.4 viser resultatene fra valgsekvens 3 der to ulike blokker/pakker av kvalitetsfaktorer ble veid opp mot takstendringer. I den første pakken inngikk faktorene fra valgsekvens 1 som har med standarden på holdeplassen å gjøre. I den andre pakken inngikk faktorene fra valgsekvens 2 som har med standarden på reisen/transportmiddelet å gjøre.

Valgsekvens 3 synes å ha vært dårlig balansert. Pakken som inneholdt faktorene sitteplass, reisetid og tid mellom avgangene synes å ha vært så dominerende at den helt klart har overskygget pakken med holdeplassfaktorene. Men i og med at designen var slik at endringene i enkeltfaktorene i den enkelte pakke ble alle forbedret eller alle forverret samtidig, har pakken med sitteplass, reisetid og tid mellom avgangene til dels også overskygget takstendringene. Dette kunne kanskje vært unngått

dersom endringene i reisetid og tid mellom avgangene hadde blitt gjort mindre i valgsekvens 3 enn det som ble brukt i valgsekvens 2.

Til tross for at valgsekvens 3 har vært forholdsvis dårlig balansert, kan vi beregne verdsetninger av de ulike nivåene for pakkene.<sup>13</sup> Standard på reisen nivå 2 (nest beste nivå) verdsettes til 12,77 kr/reise sammenlignet med nivå 3 (dårligste nivå). Standard på reisen nivå 1 (beste nivå) verdsettes til 24,66 kr/reise sammenlignet med nivå 3 (dårligste nivå). Dersom vi beregner en verdsetting for standard på holdeplass nivå 2 til tross for at det ikke gir signifikant parameterestimat, finner vi at denne er 1,11 kr/reise. Dette er altså betydelig lavere enn for pakkene med faktorer av betydning for standarden på reisen.

Tabell 6.4: Resultater fra valgsekvens 3. (N=716)

Parameter	Estimat	T-verdi
Takst	-0,1740	-6,4
Standard på holdeplass nivå 1	0,1915	0,8
Standard på holdeplass nivå 2	0,1933	1,3
Standard på reisen nivå 1	2,7660	10,2
Standard på reisen nivå 2	1,9540	10,2
Konstant	0,0332	0,4

©TØI 2001

## 6.3 Sammenligning av verdsetting funnet i ulike kontekster

Tabell 6.5 viser en sammenligning av verdsetting av kvalitetsfaktorer relatert til holdeplassen gjort separat (som i sekvens 1) og i en pakke (som i sekvens 3). Til tross for at en god del av verdsettingene ikke er signifikante, har vi valgt å presentere resultatene fra både sekvens 1 og sekvens 3 sammen med en beregnet sum av verdsettingene fra sekvens 1. I summen av enkeltfaktorene fra sekvens 1 har vi bare tatt med de faktorene som var signifikante. Dette gir altså et nedre anslag på verdsettingen av alle tre faktorene fordi de faktorene som ikke ble signifikante i sekvens 1 trolig også har en verdi som kunne vært estimert dersom faktorene hadde vært bedre balansert. Til tross for at resultatene i Tabell 6.5 ikke er signifikante for alle faktorene gir de en pekepinn på hvordan konteksten påvirker verdsettingen. Resultatene der verdsetting i en blokk/pakke som avveies mot andre blokk(er)/pakke(r) er lavere enn verdsetting av den separate verdsettingen/avveiningen av enkeltfaktorene er

<sup>13</sup> På grunn av at designen av pakkene i valgsekvens 3 har enten beste, nest beste eller dårligste nivå for alle faktorene, er det ikke nok variasjon til å kunne estimere en intern avveining av enkeltfaktorene innen den enkelte pakke.

dessuten i overensstemmelse med hva en a priori kunne forvente. Resultatene er dessuten i samsvar med tilsvarende tester av kontekstens betydning ved verdsettingen ved hjelp av CV/TP-metoden (Sælensminde 2000).

Tabell 6.6 viser en sammenligning av verdsetting av kvalitetsfaktorer relatert til reisen gjort separat (som i sekvens 2) og i en pakke (som i sekvens 3). Disse verdsettingene er alle signifikante og gir derfor god anledning til å vurdere i hvilken grad konteksten for verdsettingen påvirker resultatene. Her må vi imidlertid ta forbehold om at pakken med kvalitetsfaktorer relatert til holdeplassen som ikke ble signifikant verdsatt i sekvens 3 kan ha påvirket verdsettingen av pakken med kvalitetsfaktorer relatert til reisen. I hvilken retning verdsettingen er påvirket er usikkert, men det er ikke urimelig å tenke seg at mer likeverdige pakker kan gi lavere verdsetting av pakken med kvalitetsfaktorer relatert til reisen. Det er derfor mer trolig at verdsettingen av pakken med kvalitets-

faktorer relatert til reisen er et øvre anslag enn at det er et nedre anslag.

Som Tabell 6.6 viser gir summen av de separate avveiningene fra sekvens 2 en verdsetting på 12,77 og 24,66 kr/reise for hhv. nivå 2 (nest beste nivå) og nivå 1 (beste nivå). Tilsvarende verdsetting av en simultan forbedring i alle de tre faktorene fra sekvens 3 gir 11,26 kr/reise for nivå 2 og 15,90 kr/reise for nivå 1. Verdsetting i kvalitetsfaktorer relatert til reisen i en blokk (som avveies mot en blokk av kvalitetsfaktorer relatert til holdeplassen) gir en verdsetting som er hhv. 12,2 og 35,5 % lavere enn om faktorene i blokken verdsettes separat. En tilsvarende CV/TP studie av hvordan redusert reisetid ved en bilreise påvirkes av om den verdsettes separat eller i en pakke med trafiksikkerhet og miljøforhold viser en reduksjon i verdsettingen på 32,1 % (Sælensminde 2000). Resultatene ser også ut til å samsvare relativt godt med resultater fra en rekke tidligere studier som er presentert i Sælensminde (2000).

Tabell 6.5: Sammenligning av verdsetting av kvalitetsfaktorer relatert til holdeplassen gjort separat (som i sekvens 1) og i en pakke (som i sekvens 3). Enhet: kr/reise

Kvalitetsfaktorer	Verdsatt/ avveid separat	Sum av separate verdsettinger	Verdsatt/ avveid i blokk/ pakke
Trygghet nivå 1: Vektore	Ikke signifikant	1,39 <sup>1</sup>	Ikke signifikant
Værbeskyttelse nivå 1: Overbygget holdeplass	Ikke signifikant		
Info nivå 1: Realtidsinformasjon om ankomsttid	1,39	2,52 <sup>1</sup>	1,11 <sup>2</sup>
Trygghet nivå 2: Nødtelefon	1,05		
Værbeskyttelse nivå 2: Leskur	1,47		
Info nivå 2: Rutekart	Ikke signifikant		

<sup>1</sup> Har bare summert verdsettinger som er signifikante

<sup>2</sup> Ikke signifikant, men tatt med for å illustrere størrelsesorden

©TØI 2001

Tabell 6.6: Sammenligning av verdsetting av kvalitetsfaktorer relatert til reisen gjort separat (som i sekvens 2) og i en pakke (som i sekvens 3). Enhet: kr/reise

Kvalitetsfaktorer	Verdsatt/ avveid separat	Sum av separate verdsettinger	Verdsatt/ avveid i blokk/pakke
Sitteplass hele veien	6,99	24,66	15,90
Reisetid nivå 1 (Gj.snitt 19,85 min)	7,60 <sup>1)</sup>		
Tid mellom avgangene nivå 1 (Gj.snitt 8,05 min)	10,07 <sup>2)</sup>		
Sitteplass halve veien	3,08	12,77	11,26
Reisetid nivå 2(Gj.snitt 26,26 min)	3,89 <sup>3)</sup>		
Tid mellom avgangene nivå 2(Gj.snitt 15,43 min)	5,80 <sup>3)</sup>		

<sup>1)</sup> Beregnet som  $((32,96 \text{ min} - 19,85 \text{ min}) * 0,58 \text{ kr/min}) = 7,60 \text{ kr}$  pr gjennomsnittsreise, der 32,96 min er gjennomsnittlig reisetid presentert i nivå 3.

<sup>2)</sup> Beregnet som  $((25,43 \text{ min} - 8,05 \text{ min}) * 0,58 \text{ kr/min}) = 10,07 \text{ kr}$  pr gjennomsnittsreise, der 25,43 min er gjennomsnittlig tid mellom avgangene presentert i nivå 3.

<sup>3)</sup> Beregnet som for nivå 1.

©TØI 2001

## 6.4 Konklusjoner basert på tester av blokkdesign

Resultatene viser at det kan være minst like vanskelig å balansere blokker av kvalitetsfaktorer som å balansere enkeltfaktorer i en valgsekvens og vi har ikke klart å balansere faktorene godt nok i denne studien. Til tross for dette kan vi konkludere med at:

- Faktorene relatert til reisen (sitteplass, reisetid og tid mellom avgangene) verdsettes opptil 35 % lavere når de forbedres samtidig og veies mot andre faktorer (i en pakke/blokk) enn når de bare veies mot hverandre enkeltvis.
- Faktorene relatert til holdeplassen (trygghet, informasjon og overbygg) ser også ut til å verdsettes lavere når de forbedres samtidig og veies mot andre faktorer (i en blokk) enn når de bare veies mot hverandre enkeltvis, men på grunn av dårlig balanserte faktorer er dette ikke signifikante resultater.

Hvor stor del av reduksjonen i verdsettingen som skyldes substitusjonseffekter<sup>14</sup> på grunn av at alle faktorene forbedres samtidig i den enkelte pakken og hvor stor del som skyldes at pakkene veies mot hverandre bør undersøkes nærmere. Det viktige resultatet fra denne studien er imidlertid at også ”tunge” faktorer som sitteplass, reisetid og frekvens ”rammes” av pakkeeffekter. Dette er i samsvar med tidligere studier der faktorene er verdsatt med CV/TP (f.eks. Sælensminde 2000), men oppsummering av tidligere SC-studier i Kottenhoff og Schmidt (2000) tyder på at ”pakkeeffekter” først og fremst har vært mistenkt å ramme ”lettere” faktorer relatert til f.eks. komfort og service på transportmiddelet.

At også verdsettingen av de tunge faktorene som sitteplass, reisetid og frekvens i SC-studier påvirkes av konteksten de verdsettes i er viktig fordi disse er sentrale i mange transportanalyser. Dersom en f.eks. ønsker å iverksette en generell forbedring av kollektivtilbudet og tror at en kan øke takstene like mye som summen av betalingsvilligheten for forbedringer funnet ved separate avveininger, vil en kunne få en ”uventet” reduksjon i antall reisende. En må derfor klarlegge hvordan kollektivselskapet har tenkt å bruke verdsettingsresultatene før verdsettingsstudien designes.

### 6.4.1 Forslag til videre testing av kontekst og blokkdesign

Til tross for at denne pilotstudien kunne basere seg på resultater fra en tidligere verdsettingsstudie for Oslo Sporveier, har vi innført noen nye faktorer og nivåene for en del av faktorene har vist seg å ikke være balansert godt nok. For å kunne trekke sikrere konklusjoner med hensyn til valg av riktig kontekst og hvordan blokkdesign påvirker verdsettingene er det viktig å forsøke å balansere nivåene for hver enkelt faktor bedre. I tillegg vil det å balansere blokkene/pakkene bedre være viktig. Her kan verdsettingen av enkeltfaktorene være nyttig input. En trenger ikke bruke samme endringer for enkeltfaktorene når de skal veies mot hverandre separat som når de skal inngå i en pakke som skal veies mot andre pakker.

Et alternativ til å dele inn de ulike kvalitetsfaktorene i blokker/pakker er å bruke valgsekvenser med flere faktorer enn 4, der alle enkeltfaktorene veies mot hverandre. Spørsmålet er om dette er en farbar vei eller om det vil medføre større grad av leksikografiske og inkonsistente valg; dvs. dårligere kvalitet på dataene. Som et utgangspunkt for slike vurderinger kan en f.eks. oppsummere erfaringer andre har gjort med hensyn til å inkludere flere enn 4 faktorer i valgsekvensene.

I analyser av dataene kan det være en bedre løsning å estimere hierarkiske modeller der data fra alle sekvensene analyseres simultant.

<sup>14</sup> For eksempel kan en tenke seg at sitteplass kan være et substitutt for redusert reisetid. Det å få redusert reisetiden verdsettes derfor lavere dersom en har sitteplass enn dersom en har ståplass.

## 7 Betinget verdsetting og likeverdpris (CV/TP)

For å vurdere hvordan CV/TP-sekvensene har fungert, har vi sett på gjennomsnittlige verdsettinger av de egenskapene vi undersøker, fordelingen av betalingsvilje for reisene, på andelen respondenter med 0 i verdsetting og på andelen respondenter som gir ekstremt høye verdsettinger. De to siste kategoriene vil kunne avdekke ulike former for protestvalg, og dermed tegn på i hvilken grad CV/TP-metoden er robust mot slike valg.

Tabell 7.1 beskriver verdsettinger av de ulike forandringene i reisen som inngikk i CV/TP-sekvensen, fordelt på designvariantene 1-6, rekrutteringsmetode samt hele datasettet.

Til tross for at det ikke er noen forskjell på hvordan CV/TP-spørsmålene er definert og presentert i de ulike designene, er det er til dels store sprik mellom dem i verdsettingene. Betalingsvilje pr times forbedring i reisetiden varierer f.eks. mellom 3 og 25 kroner pr time (830%), mens betalingsviljen for ventetid varierer mellom 35 og 79 (226%) kroner pr time. For kvalitetsfaktorene som verdsettes pr reise (i motsetning til pr time) er det stort sett mindre grad av variasjon. Unntaket er forbedring fra sitteplass halve veien til sitteplass hele veien.

Disse store variasjonene tolker vi som at CV/TP-metoden er en lite stabil beregningsmetode for verdsetting av kvalitetsfaktorer, og spesielt for tidsfaktorene. Som vi skal se senere i dette kapitlet, skyldes dette blant annet en stor andel "0-svar".

Det som skiller designene fra hverandre, i tillegg til rekrutteringen og datainnsamlingen, er hvor i spørreskjemaet CV/TP-sekvensen kommer. I design 2, 3 og 5 kommer sekvensen først, mens den kommer sist blant verdsettingssekvensene i design 4 og 6.

Vi kan ikke ut fra Tabell 7.1 trekke noen konklusjoner med hensyn til hvordan rekkefølgen av verdsettingsspørsmålene påvirker verdsettingen idet det ikke er noen kare tendenser til at den blir høyere eller lavere. Imidlertid kan det se ut som om verdsettingene er mer stabile, det vil si mindre grad av ekstremverdier og variasjon, i design 4 og 6 der CV/TP-sekvensen kom lenger ut i intervjuet.

### 7.1 Forverring vektlegges mer enn forbedring

Det er en slående og til dels ekstrem assymetri i verdsettingen av forverringer og forbedringer i tilbudet. For eksempel er betalingsviljen for endret reisetid langt høyere når det er snakk om forverring av reisetiden i forhold til forbedring av reisetiden.

Vi ser at den samme tendensen viser seg for de andre egenskapene vi har undersøkt. Tabell 7.2 viser gjennomsnittlig betalingsvilje for sitte-/ståplass i CV/TP, fordelt på de 5 designene som hadde CV/TP-sekvenser. Mens om lag halvparten av beregningene for forbedringer ikke er signifikant forskjellig fra 0, er det ingen av beregningene for forverringer som er det.

Tabell 7.1: Beregnet verdsetting av de ulike egenskapene i CV/TP-sekvensen. Kroner pr time og kroner pr reise

Designvariant	2	3	4	5	6	Alle	Gallup	Internett
Forbedret reisetid kr/time	2,87*	14,65	24,18	16,10	25,19	18,67	17,15	22,44
Forverret reisetid, kt/time	25,29	50,62	48,73	41,52	39,09	43,29	44,00	41,54
Forbedring fra sitte halve reisen til sitte hele reisen, kr/reise	0,33*	7,50*	1,67	0,35*	1,09*	2,28*	3,17*	0,45
Forbedring stå hele til sitte hele veien, kr/reise	2,60*	1,50*	1,10	2,22*	1,20*	1,61	1,62	1,60
Forverring fra sitthalv til stå hele veien, kr/reise	7,00*	4,00	4,17	3,29	3,91	3,98	3,22	5,55
Forverring fra sitthel til stå hele veien, kr/reise	4,91	4,96	5,69	5,56	6,46	5,61	5,63	5,55
Redusert ventetid, kr/times ventetidreduksjon	73,70*	78,14*	34,94	79,28	36,51	58,04	59,26	55,01
Pakke 1: RTI, overbygg/tak, nødtel. og vektere	1,52*	2,85	1,89	3,46	2,17	2,47	2,51	2,37
Pakke 2: Kort reisetid, sitteplass, høy frekvens	2,10	3,74	3,01	3,54	2,88	3,17	2,96	3,69

\* Ikke signifikant forskjellig fra 0 på 5% nivå. Dette skyldes hovedsakelig at det er svært få observasjoner i disse gruppene, og mange 0-verdsettinger

Tabell 7.2: Gjennomsnittlig betalingsvillighet for forbedring til sitteplass hele veien og verdsetting av forverring til ståplass hele veien. Kr pr reise

Design	FORVERRING		FORBEDRING	
	Fra sitteplass hele veien til ståplass hele veien	Fra ståplass halve veien til ståplass hele veien	Fra ståplass hele veien til sitteplass hele veien	Fra sitteplass halve veien til sitteplass hele veien
2	4,91	7,00*	2,60*	0,33*
3	4,96	4,00	1,50*	7,50*
4	5,69	4,17	1,10	1,67
5	5,56	3,29	2,22*	0,35
6	6,46	3,91	1,20	1,09
Alle	5,61	3,98	1,61	2,28*

\* Gjennomsnittet er ikke forskjellig fra 0 på 10% signifikansnivå

©TØI 2001

Dette er funn som bekrefter at forverringer i egenskaper ved reisen tillegges større vekt enn forbedringer. Dette er et kontroversielt tema som har vært gjenstand for en rekke studier (se f.eks. Bates mfl. 2001 eller Gunn og Sheldon 2001). Diskusjonen ligger ikke så mye i hvorvidt forverringer veier tyngre enn forbedringer, som i hvordan dette faktum skal operasjonaliseres i bl.a. nytte-kostnadsanalyser. Ved å ta hensyn til denne assymetrien kan det oppstå fornuftstridige situasjoner der summen av en forverring etterfulgt av en tilsvarende forbedring blir negativ, og ikke 0.

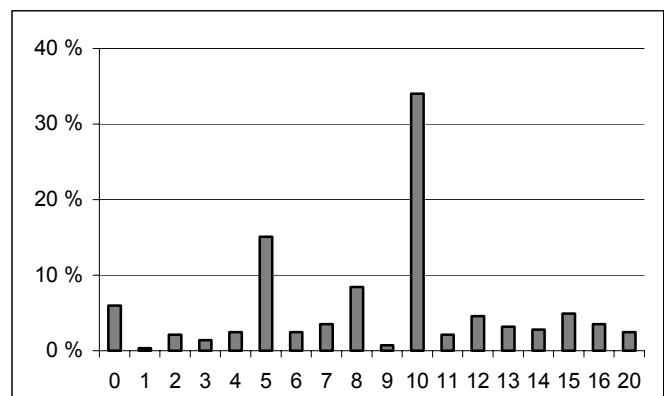
Det er klart at spørsmålsstillingen kan virke inn på resultatene og være blant årsakene til den til dels ekstreme assymetrien i verdsetting av forverringer og forbedringer av tilbudet. I spørsmålene om betalingsvilje for *forbedringer* kan respondentene oppgi svaret i intervallet mellom prisen pr reise de betalte og 99 kroner. For *forverringene* er intervallet fra 0 kroner til det de betalte pr reise (for eksempel 0–15 kroner). Dette siste gir respondenten mulighet til å velge en gratis reise (ved å svare 0) i spørsmål om betalingsvilje ved forverringer. Det kan dermed oppstå situasjoner der respondenter, som av andre årsaker mener at kollektivtransport bør være gratis (eller billig), oppgir 0 (eller lite) i betalingsvilje når de får muligheten til det. Sammenhengen med den egenskapen vi fokuserer på (f.eks. en forverring til ståplass hele veien) kan dermed bli svak. En respondent som reiser med enkeltbillett til 20 kroner og oppgir betalingsvilje på 0 kroner vil dermed tilkjenne en verdsetting av forverringen (kompensasjonskrav) på 20 kroner, som er ekstremt høyt, og sannsynligvis ikke i tråd med vedkommendes preferanser.

Det er ikke opplagt hvordan dette problemet skal løses innenfor CV/TP-metoden. Ved å spesifisere et laveste prisalternativ som er forskjellig fra 0, legger man bindinger på respondentens mulighet til å oppgi sin faktiske preferanse, i tillegg til å risikere å påvirke svarene i

retning av det laveste prisalternativet. Ingen av delene er heldige, og resultatet kan bli en skjevhet som er vel så problematisk som de skjevhetene som skyldes betalingsvilje på 0 kroner for en forverret reise.

## 7.2 Respondentene avrunder svarene

Fra Figur 7.1 kan vi identifisere en viktig svakhet ved CV/TP-metoden, som har med respondenters tilbøyelighet til å avrunde svarene. Figuren viser fordelingen av svarene på et tilfeldig valgt CV/TP-spørsmål. Det ikke tilfeldig at betalingsvillighetene 0, 5, 10 og 15 kroner er overrepresentert. Dette er nemlig en vanlig type skjevhet i direkte betalingsvillighetsspørsmål. Respondentenes preferanser er sannsynligvis jevnere fordelt, og tendensen til å avrunde svarene kan gi gale signaler om deres faktiske preferanser.



©TØI 2001

Figur 7.1: Eksempel på avrundingskjevheter. Fordelingen av respondentenes maksimale betalingsvilje (kr) for en reise med ståplass hele veien

### 7.3 Høy andel 0-verdsettinger i CV/TP

I spørsmålene om betalingsvillighet (eller kompensasjonskrav) har respondentene fått oppgitt prisen på reisen de foretok som grenseverdi for hva de kan være villige til å betale. Med andre ord har de f.eks. ikke mulighet til å velge å betale en lavere pris for en reise med kortere reisetid. Med 0 i verdsetting ("0-svar") mener vi at respondentens betalingsvilje for en endring er identisk med prisen vedkommende betalte for reisen.

Andel 0-svar er generelt svært høy, og langt høyere enn både andelen 0-svar i Frisch-sekvensen og leksikografiske svar i samvalgsekvensene.

Det grunnleggende problemet med CV/TP-metoden er den direkte spørsmålstillingen om betalingsvillighet. Spørsmålsmetoden tvinger respondentene til å fokusere på prisen, i motsetning til avveininger mellom pris og andre kvalitetsfaktorer i de andre typene verdsettings-spørsmål. Som nevnt tidligere, kan 0-svar være et uttrykk for protestvalg. Ved å bruke en såpass "gjennomskubar" spørsmålsformulering har vi sett at mange respondenter reagerer med å "protestere" og oppgi 0 i verdsetting.

I designvariantene 4 og 6 er lavere andeler 0-svar enn i de andre designene. Siden dette er designvarianter hvor CV/TP-sekvensen kommer langt bak i intervjuet (og etter de andre sekvensene med verdsettingsspørsmål), tyder dette på at rekkefølgen på de ulike sekvensene påvirker svarene i noen grad.

Vi merker oss at de laveste andelene 0-svar er for verdsettingen av en forverring fra sitteplass hele veien til ståplass hele veien. Her er andelen helt nede i 4 % i design 4 og 6, og for hele datasettet er andelen under 10 %. Vi kan tolke dette som et klart signal om at trafikantene oppfatter en forverring til å måtte stå hele veien som en vesentlig forverring av tilbudet. Kompensasjonskravet fra trafikantene var da også kr 5,63 pr. reise med en slik forverring.

Samtidig går det fram av Tabell 7.3 at det er svært mange som ikke er villige til å verdsette en forbedring til sitteplass hele veien. Opptil 82 % har 0 i betalingsvilje for det. Til tross for at vi ikke er overrasket over en svak assymetri mellom verdsettinger av forverringer og forbedringer, tyder andelene 0-svar for forbedring til sitteplass hele veien og for forverring til ståplass hele veien på at CV/TP-metoden gir et skjevt bilde av trafikantenes preferanser og betalingvilje.

### 7.4 Overbud i CV/TP

Et annet problem i CV/TP er at noen respondenter oppgir svært høy betalingsvillighet (overbud) for godet som verdsettes. Dette kan være svar som ligger urealistisk høyt i forhold til respondentenes betalingsevne. Årsaken til slike urealistiske overbud kan f.eks. være:

- at respondenten ønsker å bruke den hypotetiske verdsettingsstudien til å signalisere at dette er et gode som er svært viktig,
- et bevisst forsøk på å påvirke resultatene fra verdsettingsstudien gjennom å svare strategisk,
- at respondenten har misforstått konteksten, godets omfang, betalingsmåten eller andre forhold ved verdsettingsoppgaven, eller
- betalingsvillighet for reisen på kr 0 ved kvalitetsforverring.

Dersom slike overbud inkluderes i analysene fra datamaterialet, vil det kunne påvirke den gjennomsnittlige verdsettingen i betydelig grad selv om omfanget av overbud er forholdsvis lite. Imidlertid er det vanskelig å avgjøre hva som er et overbud, dvs. hvor en skal sette grensen for hvor høy betalingsvillighet en skal akseptere som realistisk og hvilke respondenter som skal ekskluderes. I tillegg vil ekskludering av respondenter gjøre datamaterialet mindre representativt.

Tabell 7.3: Andel (prosent) respondenter som oppgir 0 i verdsetting i CV/TP-spørsmålene

Designvariant	2	3	4	5	6	Alle	Gallup	Internett
Forbedret reisetid	86	65	36	68	44	56	59	47
Forverret reisetid	59	38	13	23	28	28	31	20
Forbedring fra sitthelv til sitthel	67	67	44	82	64	64	59	75
Forbedring stå hele til sitte hele veien	60	70	60	67	60	64	71	40
Forverring fra sitthelv til stå hele veien	33	33	22	41	36	31	37	20
Forverring fra sitthel til stå hele veien	22	16	4	11	4	10	11	6
Redusert ventetid	68	43	43	58	41	48	50	44
Pakke 1: RTI, overbygg/tak, nødtel. og vektere	74	46	38	56	41	48	52	38
Pakke 2: Kort reisetid, sitteplass, høy frekvens	65	36	26	44	29	37	40	28



Tabell 7.4 viser andelen respondenter som har oppgitt at de er villige til å øke billettprisen med (eventuelt krever kompensasjon på) 10 kroner pr reise eller mer. Det er sannsynlig at flesteparten av disse observasjonene er overbud, selv om det for noen kan være et uttrykk for ekstreme preferanser. Vi ser igjen at forverringer vektlegges mer, noe som tyder på at også overbudene gir signaler om den relative betydningen av forverringer versus forbedringer.

Med unntak for verdsettingen av forverring til å måtte stå hele veien, er andelen ikke store.

Tabell 7.4: Andeler respondenter som verdsetter endringene til mer enn 10 kroner. Prosent

	Andel med verdsetting over 10 kr
Forbedret reisetid kr/time	2,4
Forverret reisetid, kt/time	6,9
Forbedring til sitteplass hele veien	3,0
Forverring til ståplass hele veien	16,9
Redusert ventetid, kr/times ventetidreduksjon	3,6
Pakke 1: RTI, overbygg/tak, nødtel. og vektore	2,7
Pakke 2: Kort reisetid, sitteplass, høy frekvens	5,4

©TØI 2001

## 7.5 Oppsummering

Verdsetting av kvalitetsegenskaper ved hjelp av likeverpris og betinget verdsetting (CV/TP) gir til dels store sprik mellom ulike segmenter av utvalget. Derfor gir ikke CV/TP-metoden tilstrekkelig pålitelige estimater av trafikantenes verdsetting av kvaliteten på kollektivtransporten.

Vi har påvist følgende sammenhenger mellom egenskaper ved designet og CV/TP-verdsettinger av kvalitetsfaktorer:

- Effekten av rekrutterings- og datainnsamlingsmetoden ser ut til å være nøytral.
- I intervjuer hvor CV/TP-sekvensen kommer etter de andre typene verdsettingsspørsmål, er det en tendens til at CV/TP-beregningene er mer stabile og i mindre grad ekstreme. Det er også lavere andeler 0-svar i CV/TP-sekvensene i disse intervjuene.
- Respondenters tendens til å avrunde svarene sine (i vårt tilfelle til 0, 5, 10, 15 osv. kroner) påvirker resultatene og skaper skjevheter i deres verdsetting av kvalitet.
- I spørsmål om betalingsvilje for en *forverret* reise kan verdsettingen av den forverrede egenskapen bli overdreven stor, bl.a. fordi det ikke er spesifisert noen nedre grense for hva respondenten kan være villig til å betale for en forverret reise. Dette kan være en viktig årsak til at forverringer tillegges uforholdsmessig mye større vekt enn forbedringer i CV/TP-beregningene.
- Spørsmålsstillingen i CV/TP er direkte og har ingen elementer av avveining. Derfor er metoden utsatt for protestvalg og strategiske svar, som kan gi seg utslag i 0 betalingsvilje (gir høy verdsetting av en egenskap som forverres), 0 i verdsetting, eller ekstremt høye verdsettinger på grunn av overbud.
- Gjennom CV/TP-metoden har respondentene klart å signalisere at én endring i tilbudet er klart viktigere enn alle de andre: Spørsmålet om betalingsvilje for en forverring til å måtte stå hele veien gir høy betalingsvilje, en svært lav andel 0-svar og en høy andel ekstremt høye verdsettinger.

## 8 Frisch-metoden (FM)

I betinget verdsetting og likeverdspris (CV/TP) er det ofte en stor andel av respondentene som oppgir 0 i betalingsvillighet. Dersom en følger opp disse med kontrollspørsmål viser det seg at mange oppgir 0 i betalingsvillighet fordi de protesterer mot spørsmålsstillingen, konteksten, betalingsmåten eller av strategiske grunner og ikke fordi de ikke har råd til å betale eller ikke har nytte av forbedringer i de goder som verdsettes (jf. Killi, Samstad og Sælensminde 2001 for et eksempel relatert til transportforbedringer).

Tilsvarende er det ikke uvanlig med ekstremt høye verdsettinger, eller overbud, i CV/TP-spørsmål.

I Frisch-metoden (FM) blir prisforskjellen gradvis mindre, og respondenter som gjentatte ganger velger alternativet med lavest pris presenteres til slutt i sekvensen for svært små prisforskjeller. En slik valgsekvens kan antas å gi mindre andel 0- og ekstreme svar enn i CV/TP. Årsakene til dette kan være at:

1. FM gjør det mulig (og legitimt) for respondenter med liten, men positiv betalingsvillighet å tilkjenne dette. I CV/TP kan slike respondenter i større grad antas å svare 0.
2. Respondenter som i utgangspunktet antar at de har 0 i betalingsvillighet og svarer dette på et enkeltstående CV/TP spørsmål, vil i FM måtte gå gjennom mange valg (i vårt tilfelle 7) for å velge alternativet med lavest pris hele veien. Disse får dermed i større grad enn ved CV/TP tenkt over om deres betalingsvillighet virkelig er 0.
3. Respondenter som i utgangspunktet vil angi 0 eller en ekstrem pris i betalingsvillighet som et protestsvar eller strategisk svar på verdsettingsspørsmålene, vil også måtte gå gjennom mange valg for å velge alternativet med lavest/høyest pris hele veien. Disse får dermed i større grad enn ved et enkeltstående CV/TP-spørsmål tenkt over om de skal gi fra seg muligheten til å tilkjenne sine preferanser i et samfunns-spørsmål som egentlig er viktig for dem.

### 8.1 FM kan korrigere leksikografiske svar

Som vi har sett, er det mange problemer som kan unngås dersom en unngår leksikografiske valg i samvalgstudier. Leksikografiske valg vil imidlertid også skape problemer

i FM dersom respondentene av en eller annen grunn skulle vise seg å ikke veie faktorene mot hverandre.

Dersom FM gjør det lettere for respondentene å velge, kan en imidlertid anta at respondenter som i SC ville valgt leksikografisk for å forenkle valgoppgaven i FM i større grad vil veie faktorene mot hverandre. Og likeledes at respondenter som i SC velger leksikografisk på grunn av sine preferanser og dårlig balanserte nivåer på faktorene gis en bedre mulighet til å foreta avveininger i FM fordi det i utgangspunktet er svært stor forskjell på alternativene og fordi denne forskjellen kan gjøres svært liten. Det er altså grunn til å tro at FM vil ha mindre problemer med leksikografiske valg enn SC enten disse er forårsaket av et ønske om forenkling av valgoppgaven eller for dårlig balansering av faktorenes nivåer. Et annet viktig poeng er at resultatene fra FM kan brukes som input til balansering av nivåene i SC og dermed også redusere problemene med leksikografiske valg i SC. Med hensyn til å redusere leksikografiske valg, kan metodene altså med fordel supplere hverandre (jf. pkt. 5.5 – Balansering ved hjelp av CV/TP).

### 8.2 Inkonsistente valg er vanlig i SC-sekvenser, men umulig innen den enkelte FM-sekvens

Sælensminde (2000) viser at mellom 60 og 75 % av respondentene velger inkonsistent i en samvalgsekvens der bare tre faktorer varierer og at inkonsistente valg påvirker verdsettingen av faktorene som inngår i valgsekvensen<sup>15</sup>. I SC-studier med flere enn tre faktorer er det dessuten grunn til å tro at andelen inkonsistente valg er større og følgelig at verdsettingene i større grad påvirkes.

FM gir ikke respondentene mulighet til å velge inkonsistent innenfor den enkelte FM-sekvens, men dersom en vil verdsette flere faktorer i ulike FM-sekvenser, er det ikke gitt at det er konsistens mellom disse FM-sekvensene.

<sup>15</sup> Med inkonsistens i en SC-sekvens menes at ett eller flere valg foretatt av en respondent strider mot andre valg foretatt av den samme respondenten i samme valgsekvens.

### 8.3 Resultater ved bruk av Frisch-metoden

Respondentene fikk fire ulike FM-sekvenser der endring i følgende forhold ved kollektivreisen skulle avveies:

1. Redusert reisetid mot lengre tid mellom avgangene (dvs. Redusert frekvens)
2. Redusert reisetid mot høyere takst
3. Mindre tid mellom avgangene (dvs. Økt frekvens) mot høyere takst
4. Mindre tid mellom avgangene (dvs. Økt frekvens) mot lengre reisetid

Sekvens 1 og 4 gir altså avveining mellom reisetid og frekvens på en slik måte at en direkte kan teste om det er symmetri mellom det respondentene vil avgi i form av lengre tid mellom avgangene for å få redusert reisetid (sekvens 1) og det respondentene vil avgi i form av lengre reisetid for å få mindre tid mellom avgangene (sekvens 4). Likeledes er det mulig å teste om avveininger gjort i en kontekst kan overføres til en annen. For eksempel om avveining mellom hhv. reisetid og takst (sekvens 2) og tid mellom avganger og takst (sekvens 3) kan brukes til å anslå avveiningen mellom reisetid og tid mellom avgangene.

Sekvens 2 og 3 er mer å betrakte som ”vanlige” betalingsvillighetsspørsmål der betalingen er i form av høyere kostnad. Verdsettingene fra disse sekvensene kan derfor direkte sammenlignes med verdsettingen fra CV/TP og delvis samvalg. Når vi skriver delvis samvalg, er det fordi i samvalg inngår det også forverringer i tilbudet, og det å unngå forverringer kan verdsettes høyere enn å få forbedringer. Sammenligning av verdsettingen fra FM, CV/TP og samvalg er gjort i pkt. 11.4 (Tester av Frisch-metoden).

### 8.4 Respondentenes valg i FM-sekvensene

Her gis det først, i Tabell 8.1, en oversikt over hvordan de 333 respondentene har valgt i FM-sekvensene. Inndelingen i Tabell 8.1 viser i første linje andelen av respondenter som valgte reise 2 eller ”like gode” i første valg. Dette er altså respondenter som godtar en stor forverring for å få den skisserte forbedringen. I andre linje er andelen respondenter som valgte reise 1 i alle valgene. Dette er respondenter som ikke godtar selv en liten forverring for å få den skisserte forbedringen. Tredje linje er andelen respondenter som i løpet av sekvensen valgte både reise 1 og reise 2, men som ikke kom frem til nivåer som var slik at reisene ble oppfattet som ”like gode”. Fjerde linje viser andelen respondenter som i løpet av sekvensen nådde sitt indifferensnivå og tilkjennega at reise 1 og reise 2 var ”like gode”. Resultatene i Tabell 8.1 er kommentert og vurdert i de neste seksjonene av pkt. 2.1.

#### 8.4.1 Forståelsen for valgoppgavene synes god

Dersom en respondent i første valget i FM-sekvensene valgte reise 2 eller at alternativene var ”like gode”, dvs. alternativene som tilkjennega stor betalingsvillighet for den forbedringen som var skissert, fikk han/hun et kontrollspørsmål for å få bekreftet om betalingsvilligheten virkelig var så høy eller om det var et feilvalg. Dataene viser at bare 2 (0,6 %) av respondentene oppgir at de på denne måten ”valgte feil” i første FM-sekvens. (Som det fremgår av Tabell 8.1 var det 2,4 % av respondentene, dvs. 8 av 333 personer, som valgte reise 2 eller ”like gode” i første valg.) I de tre påfølgende FM-sekvensene er det ingen som oppgir at de ”valgte feil”. De som valgte feil fikk anledning til å rette opp sitt valg.

Tabell 8.1: Fordeling av respondentenes valg i de fire ulike sekvenser av ”Frisch-metoden”. Enhet: Prosent. (N=333.)

Respondenter som i sekvensen valgte:	FM-sekvens 1, reisetid vs frekv.	FM-sekvens 2, reisetid vs takst	FM-sekvens 3, frekv. vs takst	FM-sekvens 4, frekv. vs reisetid
Reise 2 eller ”like gode” i første valg	2,4	2,4	0,9	3,3
Reise 1 i alle valgene	4,8	18,3	13,8	10,8
Både reise 1 og 2	41,1	37,5	41,7	33,6
”Like gode” i andre eller senere valg	51,7	41,7	43,5	52,3

#### 8.4.2 Mellom 1 og 3 % av respondentene valgte alternativet med høyeste betalingsvillighet i første valg

De 1-3 % av respondentene som velger å tilkjenne høyeste betalingsvillighet i første valg oppgir i kontrollspørsmål (se over) at deres betalingsvillighet virkelig er så høy. Hvorvidt dette gjenspeiler reell betalingsvillighet, strategiske valg eller misforståelser kan imidlertid diskuteres. Det kan følgelig også diskuteres hvorvidt slike respondenter skal tas ut av datamaterialet fordi en må/bør sette en grense for hvor høye verdsetninger som skal aksepteres som realistiske for at ikke gjennomsnittlige verdsetninger skal påvirkes i vesentlig grad. Øvre grense for forskjellene mellom de to reisene i FM var tre ganger så høy som ved dagens reise.

#### 8.4.3 Mellom 14 og 18 % av respondentene valgte laveste pris i alle valgene

14 og 18 % av respondentene valgte laveste pris i alle FM-valgene, dvs. ville ikke godta høyere pris for å få hhv. redusert reisetid eller høyere frekvens. Å anta at alle disse har 0 i betalingsvillighet er imidlertid ikke riktig fordi bare hhv. 7 og 2 % av disse ble presentert for alternativer der prisforskjellen var 0. Disse til sammen 7 respondentene har altså valgt et alternativ (dagens reise) som var dårligere enn den nye reisen med redusert reisetid til samme takst. Disse har enten misforstått, eller så foretrekker de den reisetiden de har i dag fremfor redusert reisetid. De øvrige, hhv. 93 og 98 % av dem som valgte laveste pris hele veien, fikk en prisforskjell på 1 krone.

Hvordan skal så respondenter som tilkjenner null i betalingsvillighet behandles? Et konservativt anslag (for ikke å få for høy betalingsvillighet) er å anta at alle de som har valgt laveste pris hele veien har null i betalingsvillighet. Slik er det gjort i de videre analysene fra dette datamaterialet. For de metodiske vurderingene som gjøres her gjør det liten forskjell om deres betalingsvillighet settes lik 0 eller settes svært lavt.<sup>16</sup> En vil imidlertid få mer innsikt i motivene for å tilkjenne null betalingsvillighet og dermed hvordan en skal behandle slike, ved hjelp av kontrollspørsmål. Inntil en vet med sikkerhet hvilke respondenter som

<sup>16</sup> En kan f.eks. sette betalingsvilligheten til respondentene som har valgt laveste pris hele veien, men som fortsatt har en prisforskjell på 1 krone mellom alternativene, til kr 0,25 (som er mer sannsynlig enn kr 0,5 gitt at en del av disse trolig har kr 0 i betalingsvillighet). Ved en gjennomsnittlig reduksjon i reisetiden på 5 minutter, vil en få 3 kr/t i betalingsvillighet for redusert reisetid for denne gruppen. Dette vil øke gjennomsnittlig verdsetting fra ca 32 kr/t til ca 32,5 kr/t for redusert reisetid.

har svart null i protest, kan en ikke ekskludere respondenter fra denne gruppen.

#### 8.4.4 Bare 5 % av respondentene ville ikke godta lengre reisetid for å få redusert tid mellom avgangene

I de FM-sekvensene der respondentene skulle veie endringer i reisetid mot endringer i frekvens var det 5 % av respondentene som ikke ville godta lengre reisetid for å få redusert tid mellom avgangene og 10 % som ikke ville godta lengre tid i mellom avgangene for å få redusert reisetid. Det var altså en betydelig mindre andel som valgte laveste alternativ i alle valgene i disse FM-sekvensene der kostnad ikke inngikk enn i de tilfellene der kostnad inngikk i FM-sekvensene (jf. over).

Men også i FM-sekvensene uten kostnad er det en stor andel som har 1-2 enheters forskjell på alternativene i det siste valget de fikk i FM-sekvensen. Vurderingene over om hvordan denne andelen kan reduseres og hvordan respondenter med null i betalingsvillighet skal inkluderes gjelder også her. I de videre analysene er også disse respondentenes betalingsvillighet satt lik null.

#### 8.4.5 Mellom 82 og 95 % av respondentene har veid godene mot hverandre i Frisch-sekvensene

Mellom 82 og 95 % av respondentene har en eller flere ganger valgt alternativet der de måtte avgi noe for å få en forbedring. Disse har altså veid endringer i godene mot hverandre. Av dem som har foretatt avveininger mellom godene var det:

1. Bare 1-3 % som sa ja til å avgi i første FM-valg og som dermed falt utenfor øvre grense satt i FM-sekvensene.
2. 42-52 % som oppga at FM-sekvensen hadde gitt alternativer som var "like gode".
3. 37-44 % som gikk gjennom hele valgsekvensen og som endte opp med svært liten (eller ingen) forskjell mellom den minste betalingen/avgivelsen de sa nei til og den største betalingen/avgivelsen de sa ja til, men som ikke sa at alternativene var "like gode".

For denne siste gruppen var forskjellen mellom "nei-alternativet" og "ja-alternativet" i 99 % av tilfellene mindre eller lik en enhet.

Årsaken til at ca 40 % av respondentene havner i gruppe 3. kan skyldes at vi bare brukte hele minutter og hele kroner som enhet i valgene. Det vil si at en respondent som f.eks. hadde svart ja til å betale 17

kroner for en reise og deretter nei til å betale 18 kroner for denne reisen, i sitt neste valg igjen ville kunne få en kostnad på 18 eller 17 kroner (avhengig av avrunding) å ta stilling til. Om respondenten også denne gang ville si nei til 18 kroner og ja til 17 kroner varierer i vårt datamateriale. Det er imidlertid ikke gitt hvilket svar som er det ”riktige”, og det vil derfor være tilfeldig om en ender opp med 0 eller 1 enhets forskjell mellom ”nei-alternativet” og ”ja-alternativet”.

Hvordan skal vi så behandle respondenter i gruppe 3? På grunn av at forskjellen mellom ”nei- og ja-alternativet” er så liten, har vi valgt å behandle denne gruppen på lik linje med de som svarer at alternativene er ”like gode” ved å si at respondentenes betalingsvillighet ligger midt mellom ”nei-alternativet” og ”ja-alternativet”. Denne konklusjonen er basert på en vurdering av at årsaken til at respondentene i gruppe 3. ikke har svart ”like gode”, til tross for små forskjeller, trolig skyldes utformingen av FM-sekvensen i denne studien.

Ved å opplyse respondentene før de ulike FM-sekvensene om hva som er minste enhet og/eller å gi f.eks. halve enheter kan en trolig redusere andelen som havner i gruppe 3. Men i og med at også de fleste respondentene i gruppe 3. kommer maksimalt en enhet fra sitt indifferensnivå gjør vi trolig ingen stor feil ved å behandle denne gruppen på lik linje med dem som svarer at alternativene er ”like gode”.

## 8.5 Avveininger mellom reisetid, tid mellom avgangene og takst

Generelt kan vi si at avveiningene i FM-sekvensen avhenger i betydelig grad av om vi inkluderer eller ekskluderer respondenter som valgte reise 2 eller ”like gode” i første valg. Dette er ikke en stor gruppe (jf. Tabell 8.1), men som ved bruk av metodene CV og TP er gjennomsnittlige verdsetninger fra FM også påvirket av ekstremverdier. Riktignok vil størrelsen på ekstremverdiene i FM kunne begrenses ved at man ikke aksepterer høyere verdsetninger enn et visst nivå, men som resultatene i tabell 2.2 – 2.5 viser, skal det ikke mange slike til før verdsettingene trekkes opp med 10-20 %.

I pkt. 8.4.1 (Forståelsen av valgoppgavene i FM synes god) diskuteres det hvordan en skal få bedre innsikt i hvordan slike skal behandles, men inntil

videre har vi valgt å presentere både avveininger der respondenter som valgte reise 2 eller ”like gode” i første valg er inkludert og der disse er ekskludert. En viktig grunn til at disse kan forsvares inkludert er at i alle de fire FM-sekvensene viser det seg at denne gruppen både har fått skissert større reduksjoner i reisetiden og i tiden mellom avgangene enn de andre respondentene. Og størrelse på reduksjonen er ikke tilfeldig fordi disse er beregnet ut fra det tilbudet de ulike respondentene har i forbindelse med sin aktuelle reise. Respondentene som valgte reise 2 eller ”like gode” i første valg, og dermed tilkjennega en høy vilje til å avgi/betale noe for å oppnå forbedringer, har altså gjennomsnittlig lengre reisetider og lengre tid mellom avgangene enn gjennomsnittet for de øvrige respondentene. Til tross for at disse respondentene får skissert større absolutte forbedringer i kollektivtilbudet enn andre, kan det vanskelig forsvares/forklares at de f.eks. skal ha verdsetninger av redusert reisetid og redusert tid mellom avgangene på hhv. 297 kr/t og 205 kr/t (hhv. 4,95 og 3,42 kr/min fra Tabell 8.3 og 8.4). Det er altså grunn til å tro at en del av (alle?) disse respondentene har brukt verdsettingsspørsmålene til å tilkjennegi sin misnøye med kollektivtilbudet og/eller har forsøkt å svare strategisk for å påvirke resultatene fra undersøkelsen. Men ettersom dette er respondenter som har større nytte av forbedringer i kollektivtilbudet enn andre, er det også trolig at de har større betalingsvillighet for disse forbedringene enn andre. Med andre ord vil det ikke være riktig å ekskludere dem fra analysen. Et slikt standpunkt kan også forsvares ut fra det faktum at det er deres *nytte* av de skisserte forbedringene som er den riktige størrelsen å inkludere i en samfunnsøkonomisk analyse. Problemet er å få tak i deres betalingsvillighet og dermed et mest mulig riktig anslag på deres nytte av forbedringer, gjennom en hypotetisk verdsettingsstudie. Inntil vi kommer nærmere en løsning på dette problemet, velger vi altså å presentere både anslag på betalingsvilligheten der respondenter som valgte reise 2 eller ”like gode” i første valg er inkludert og der disse er ekskludert. Et slikt intervall gir trolig et godt bilde på den ”riktige” betalingsvilligheten og altså respondentenes ”nytte” av de skisserte forbedringer.

Tabell 8.2: Resultater fra FM-sekvens 1: Redusert tid mellom avgangene mot lengre reisetid. Gjennomsnittlig reduksjon i tid mellom avgangene (min), akseptert økning i reisetid (min) og forholdet mellom økt reisetid og redusert tid mellom avgangene for ulike segmenter av respondenter. (N totalt = 333)

Respondenter som i sekvensen valgte:	Reduksjon i tid mellom avgangene (min)	Akseptert økning i reisetid (min)	Økt reisetid/ redusert tid mellom avgangene
Reise 2 eller "like gode" i første valg	13,38	50,25	5,06
Reise 1 i alle valgene	6,81	Satt lik 0	Satt lik 0
Både reise 1 og 2, reise 1 i siste valg	7,99	8,06	1,13
Både reise 1 og 2, reise 2 i siste valg	7,25	7,85	1,18
"Like gode" i andre eller senere valg	7,44	7,14	1,13
Alle	7,66	8,18	1,18
Alle, unntatt de som valgte reise 2 eller "like gode" i 1. valg	7,52	7,14	1,08

©TØI 2001

Tabell 8.3: Resultater fra FM-sekvens 2: Redusert reisetid mot høyere takst. Gjennomsnittlig reisetidsreduksjon (min), akseptert takstøkning (kr) og betalingsvillighet for redusert reisetid (kr/min) for ulike segmenter av respondenter. (N totalt = 333)

Respondenter som i sekvensen valgte:	Reisetidsreduksjon (min)	Akseptert takstøkning (kr)	Økt takst/Redusert reisetid (kr/min)
Reise 2 eller "like gode" i første valg	10,88	27,25	4,95
Reise 1 i alle valgene	4,23	Satt lik 0	Satt lik 0
Både reise 1 og 2, reise 1 i siste valg	6,78	4,45	0,75
Både reise 1 og 2, reise 2 i siste valg	7,00	3,28	0,51
"Like gode" i andre eller senere valg	6,17	3,56	0,65
Alle	6,19	3,66	0,64
Alle, unntatt de som valgte reise 2 eller "like gode" i 1. valg	6,07	3,08	0,54

©TØI 2001

Tabell 8.4: Resultater fra FM-sekvens 3: Redusert tid mellom avgangene mot høyere takst. Gjennomsnittlig reduksjon i tid mellom avgangene (min), akseptert takstøkning (kr) og betalingsvillighet for redusert tid mellom avgangene (kr/min) for ulike segmenter av respondenter. (N totalt = 333)

Respondenter som i sekvensen valgte:	Reduksjon i tid mellom avgangene (min)	Akseptert takstøkning (kr)	Økt takst/Redusert tid mellom avgangene (kr/min)
Reise 2 eller "like gode" i første valg	9,67	29,33	3,42
Reise 1 i alle valgene	6,37	Satt lik 0	Satt lik 0
Både reise 1 og 2, reise 1 i siste valg	7,92	4,12	0,59
Både reise 1 og 2, reise 2 i siste valg	9,04	3,52	0,46
"Like gode" i andre eller senere valg	7,41	4,04	0,68
Alle	7,66	3,65	0,73
Alle, unntatt de som valgte reise 2 eller "like gode" i 1. valg	7,64	3,42	0,67

©TØI 2001

Tabell 8.5: Resultater fra FM-sekvens 4: Redusert reisetid mot lengre tid mellom avgangene. Gjennomsnittlig reisetidsreduksjon (min), akseptert økt tid mellom avgangene (min) og forholdet mellom økt tid mellom avgangene og redusert reisetid for ulike segmenter av respondenter. (N totalt = 333)

Respondenter som i sekvensen valgte:	Reduksjon i reisetid (min)	Akseptert økt tid mellom avgangene (min)	Økt tid mellom avgangene reisetid
Reise 2 eller "like gode" i første valg	14,36	49,81	3,24
Reise 1 i alle valgene	4,16	Satt lik 0	Satt lik 0
Både reise 1 og 2, reise 1 i siste valg	6,63	4,14	0,72
Både reise 1 og 2, reise 2 i siste valg	6,70	5,48	0,80
"Like gode" i andre eller senere valg	5,75	4,99	1,04
Alle	6,19	5,87	0,90
Alle, unntatt de som valgte reise 2 eller "like gode" i 1. valg	5,91	4,37	0,82

©TØI 2001

## 8.6 Tester av konsistens mellom ulike FM-sekvenser

For diskusjoner omkring konsistensbegrepet og tester av om det er konsistens mellom ulike FM-sekvenser viser vi til Sælensminde (2001c). Her oppsummeres kun noen hovedresultater fra disse testene:

- En "tradisjonell test" av konsistens tilsier at 24,3 % av respondenter har valgt konsistent. Denne testen ville da ha forutsatt at redusert tid mellom avgangene skulle verdsettes høyere enn redusert reisetid (f.eks. basert på a priori-informasjon fra tidligere studier) og at respondentene i tillegg skulle velge i henhold til disse preferansene selv når tid mellom avgangene ikke blir veiet eksplisitt mot reisetid.
- En "utvidet tradisjonell test" av konsistens ville kunne si at 43,5 % av respondentene til sammen har valgt konsistent.<sup>17</sup> Denne testen har da akseptert at enkelte respondenter kan mene at redusert reisetid er viktigere enn redusert tid mellom avgangene, men krever fortsatt at respondentene i tillegg skal velge i henhold til disse preferansene også når tid mellom avgangene ikke blir veiet eksplisitt mot reisetid.

For tester av om design påvirker avveiningene i FM henvises det også til Sælensminde (2001c). Resultatene fra disse testene gir imidlertid ingen entydige konklusjoner med hensyn til at egenadministrert intervju eller egenrekrutterte respondenter gir data med dårligere kon-

<sup>17</sup> Andelen som valgte konsistent i samvalgsspørsmålene på korte kollektiv-reiser i den norske tidsverdistudien var på 39,7 % (Sælensminde 2000). Den "utvidede" testen av konsistens fra FM gir altså resultater i samme størrelsesorden. Disse konsistenstestene av FM og SC er trolig nokså sammenlignbare fordi det i begge tilfeller inngikk tre faktorer (takst, reisetid og tid mellom avgangene).

sistens enn hhv. hjemmeintervju og respondenter som ble rekruttert av Norsk Gallup institutt.

## 8.7 Forslag til videre testing av FM

### 8.7.1 Kontrollspørsmål

For å få mer kunnskap om hva som faktisk ligger til grunn for de svarene som avgis i SP-studier kan det være en fordel å bruke flere kontrollspørsmål. Aktuelle kontrollspørsmål er:

- Spørsmål om vanskelighetsgrad og eventuelle andre vanskeligheter
- Spørsmål til respondenter med null (og ev. svært lav) betalingsvillighet
- Spørsmål til respondenter med svært høy betalingsvillighet

Hvordan slike kontrollspørsmål kan formuleres og bygges opp, er beskrevet i Sælensminde (2001c).

### 8.7.2 Bruk av lengre FM-sekvenser

FM-dataene i denne studien viser at omtrent 40 % av respondentene har skiftet på hvilket av de skisserte alternativene de har valgt og at de derfor trolig har avveid godene mot hverandre, men at de i løpet av 7 valg ikke har kommet helt frem til sitt indifferensnivå og derfor ikke har endt opp med å velge at alternativene er "like gode". En FM-sekvens med noen flere valg kan derfor trolig få flere av disse respondentene til å ende opp med alternativer som de anser som "like gode".

For respondenter som f.eks. har lav betalingsvillighet og derfor i hele FM-sekvensen velger alternativet med lavest pris, men uten å komme frem til liten nok prisforskjell til å kunne velge at alternativene er ”like gode”, vil en forlengelse av FM-sekvensen kunne gi bedre mulighet for å nå sitt indifferensnivå. Også for denne gruppen respondenter vil en lengre FM-sekvens kunne få flere til å avslutte sekvensen med å velge at alternativene er ”like gode”.

Et sikrere anslag på andelen respondenter som har 0 i betalingsvillighet kan en også få ved hjelp av en lengre FM-sekvens der en til slutt i sekvensen faktisk presenterer et valg som er slik at respondenten ikke behøver å gi noe for å få den skisserte forbedringen. Under forutsetning av at respondenten ikke har negativ nytte av den forbedringen vi har skissert, vil han/hun dermed ved å velge at alternativene er ”like gode” kunne tilkjenne at betalingsvilligheten er null for den skisserte forbedringen.

Nytten av lengre FM-sekvenser må i alle tilfeller veies opp mot problemer dette kan skape for respondentene, som tretthet.

### 8.7.3 Verdsetting av pakker av goder og interaksjonseffekter mellom goder

I de testene av FM som her er gjort er ett og ett gode veid mot hverandre. Det er grunn til å tro at FM også vil fungere tilfredsstillende til verdsetting av forbedringer i flere goder samtidig. For eksempel kan en ønske å verdsette en pakke av forbedringer i forbindelse med en mer omfattende opprustning av kollektivtilbudet. Slike pakker av goder har vært verdsatt ved hjelp av CV/TP og verdsettingen ser i de fleste tilfeller ut til å bli mindre for pakken som helhet enn for summen av de godene som inngår når disse er verdsatt separat (Sælensminde 2000). Det ser altså ut til at respondentene tar hensyn til sin betalingsvillighet når de tilkjenner sin betalingsvillighet og/eller at godene som inngikk i disse pakkene var substitutter. Uansett er det slik at den riktige konteksten for avveiningene vi presenterer respondentene for er den konteksten verdsettingene senere er tenkt brukt i. En kan ikke forvente å få riktige relative (og absolutte) verdsettinger av ikke-markedsgoder (og markedsgoder!) dersom viktige substitutter ikke inngår i avveiningene verdsettingene er basert på.



## 9 Fleksibelt pakkevalg (POG)

Respondentene har hatt muligheten til å velge mellom tre nivåer for reisetid, frekvens, sitteplass, holdeplass-informasjon, leskur og trygghetsforhold. Alle har innvirkning på taksten, som er blitt oppdatert fortløpende etter hvert som respondenten foretar valgene.

Tabell 9.1 beskriver hvordan respondentene valgte. Nivå 1 er alltid det alternativet som gir billigere billettpris mot dårligere kvalitet, mens nivå 3 gir høyere pris og bedre kvalitet. Vi ser at nivå 2 er det oftest foretrukne valget for alle faktorer unntatt informasjon på holdeplassen. Samtidig ser vi at det i gjennomsnitt er negativ betalingsvilje for trygghetstiltak. For alle faktorene har respondentene i gjennomsnitt valgt kombinasjoner som gir en takstøkning på 81 øre.

For egenskapene reisetid, frekvens og sitteplass, kan nivå 2 forstås som 0, eller svært lav, betalingsvilje for endringer. Nivå 1 kan dermed betraktes som en positiv vektlegging av lavere pris.

Vår erfaring med POG er at det er et svært lite fleksibelt verktøy av flere årsaker. Bl a er nivåene på de ulike egenskapene r vanskelig å relatere til respondentens reise. Derfor har vi representert endringene som prosentvise i forhold til den faktiske reisen respondenten gjennomførte, men dette kan igjen skape problemer fordi prosentvise endringer kan være vanskelig å forholde seg til. Videre gir metoden heller ingen lettvinde beregninger av betalingsvilje pr minutt endring i reise- og ventetid og reisetid med ståplass fordi respondentene i realiteten bare gjør én avveiing for hver kvalitetsfaktor: Er verdsetningen av den enkelte kvalitetsfaktoren høyere, lavere eller lik det oppgitte beløpet?

POG kan derfor være en grei måte å få oversikt over de reisendes prioriteringer, på lik linje med rangering, men gir i sin nåværende form ingen god indikasjon på den faktiske verdsetningen av kvaliteten på kollektivtransporten.

Tabell 9.1: Fordelingen av POG valgene og gjennomsnittlig takstendring i øre pr reise (betalingsvilje) som følge av valgene.

	Reisetid	Frekvens	Sitteplass	Info	Leskur	Trygghet	Alle endringer
Nivå 1	18,2%	7,6%	21,2%	47,0%	21,2%	34,8%	25,0%
Nivå 2	57,6%	75,8%	48,5%	31,8%	71,2%	47,0%	55,3%
Nivå 3	24,2%	16,7%	30,3%	21,2%	7,6%	18,2%	19,7%
Gj.snittlig takstendring	12,12	27,27	18,18	7,58	17,58	-1,82	80,91

©TØI 2001

# 10 Tester av konsistens mellom de ulike SP-metodene

En indikasjon på i hvilken grad respondentene har taklet intervjusituasjonen er om det er konsistens mellom svarene, og måten de svarer, i de ulike undersøkelsesmetodene. Vi vil her se på sammenhenger mellom hvordan respondentene har svart i Frisch-, POG-, CV/TP- og samvalgsekvensene. Hvis det er stor grad av samsvar mellom verdsettingen av de ulike egenskapene, kan vi si at respondentene har taklet intervjusituasjonen, og at de klarer å tilkjenne sine preferanser og verdsettinger.

Vi kan også forvente at de som har sortert valgene sine i samvalgsekvensen etter en bestemt faktor også tilkjenner høyere relativ verdsetting av denne faktoren i CV/TP-, POG- og Frisch-sekvensene, gitt at de har tilkjenner sine faktiske preferanser i samvalgsekvensene. På samme måte vil vi vente at en del av de som sorterer leksikografisk etter prisen vil tilkjenne lavere betalingsvilje for de øvrige faktorene i de andre spillene.

## 10.1 CV/TP versus samvalganalyser

Vi har sett at flere respondenter har oppgitt 0 i betalingsvilje i CV/TP-sekvensene. Det er interessant å se på hvordan dette samsvarer med leksikografi i samvalgsundersøkelsene.

Dersom leksikografisk sortering av svarene skyldes faktiske avveininger og ikke at respondentene benyttet forenklede beslutningsrutiner, vil vi forvente at respondenter som sorterer leksikografisk etter f.eks. frekvens også tilkjenner en høyere betalingsvillighet for frekvens i CV/TP-sekvensen. Tabell 10.1 sammenligner gjennomsnittlig betalingsvillighet i CV/TP-sekvensene mellom dem som har sortert leksikografisk og dem som ikke har sortert leksikografisk for de aktuelle egenskapene og pakkene. Vi ser at det er en god sammenheng mellom de to verdsettingsmetodene ved at de som sorterer etter egenskapen i samvalgene også i stor grad legger større vekt på egenskapene i CV/TP. Dette indikerer at respondentene i stor grad har klart å takle valgsituasjonen i samvalgsekvensen, og at leksikografiske svar ikke hovedsakelig skyldes forenklinger av beslutningsrutinene.

Vi kan ikke fastslå om nivåene på verdsettingene i Tabell 10.1 er riktige uttrykk for trafikantenes verdsetting, men det er helt tydelig at det er stor grad av konsistens i svarene mellom CV/TP og samvalg.

Tabell 10.2 viser andelen 0-svar i CV/TP i forhold til hva de sorterte leksikografisk etter i samvalgene. For eksempel ser vi at blant dem som sorterte leksikografisk etter reisetid i samvalgsekvensen, var det 44,4 % som hadde 0 i betalingsvilje for kortere reisetid i CV/TP, mot 55,7 % for hele datasettet.

Hvis respondentene har svart konsistent i henhold til sine preferanser, forventer vi at:

- De som sorterer etter reisetid i SC ikke har 0-verdsettinger av reisetid i CV/TP
- De som sorterer etter frekvens i SC ikke har 0-verdsettinger av kortere ventetid i CV/TP
- De som sorterer etter pris i SC skal ha høye andeler 0-verdsettinger i CV/TP

Selv om andelen 0-svar er noe lavere blant dem som sortere etter reisetid enn for hele datasettet, virker derfor bildet ulogisk. En person som sorterer etter reisetid i samvalgene gir nemlig klare signaler om at reisetid er viktig. Samtidig er det altså hele 44,4 % av disse personene som ikke har tilkjenner betalingsvilje for kortere reisetid i CV/TP-sekvensen. Vi konstaterer derfor at mange respondenter ikke klarer å verdsette kvalitetsfaktorene på en konsistent måte mellom samvalgsekvensen og CV/TP-sekvensen.

Det er sannsynlig at det er den direkte spørsmålsstillingen i CV/TP som har forårsaket inkonsistensen, jf. diskusjonen omkring protestvalg og takstisk svargivning. CV/TP gjør det derfor på mange måter "lett" for respondentene å velge 0 i verdsetting.

Samtidig er det til dels god korrespondanse mellom sortering etter prisen og 0 i verdsetting av pakke 2 (kort reisetid, sitteplass, høy frekvens). De som sorterte etter prisen tilla per definisjon ingen vekt til de øvrige kvalitetsfaktorene i samvalgene. Denne gruppen har en stor overvekt 0-svar i CV/TP-spørsmålene, slik vi skulle forvente.

Tabell 10.1: Gjennomsnittlig (median i parentes) betalingsvilje i CV/TP fordelt på de som har sortert og de som ikke har sortert etter egenskapen i samvalgsekvensene. (Respondenter som sorterer etter pris i samvalg forventes å ha lavere betalingsvilje for de andre egenskapene ved reisen)

Egenskap	Betalingsvilje blant dem som har sortert etter egenskapen i samvalgsekvensen	Betalingsvilje blant dem som ikke har sortert etter egenskapen i samvalgsekvensen
Sitteplass (forverring fra sitthel til ståhel)	10,4 (10,0)	5,4 (5,0)
Sitteplass (forverring fra sitthel til ståhel)	5,5 (5,5)	4,0 (3,0)
Sitteplass (forbedring fra sitthel til sitthel)	3,5 (3,5)	2,3 (0)
Reisetid (forverret reisetid)*	26,7 (20,0) kr/time	44,0 (27,7) kr/time
Reisetid (forbedret reisetid)**	12,4 (10,0) kr/time	17,5 (0,0) kr/time
Frekvens**	37,8 (34,3) kr/time	55,5 (0,0) kr/time
Pris (Betalingsvilje for Pakke 1)	1,6 (0,0)	2,2 (2,0)
Pris (Betalingsvilje for Pakke 2)	1,0 (0,0)	2,5 (2,0)

\* Ikke samsvar mellom leksikografi i SP og gjennomsnittlig betalingsvilje i CV/TP for denne egenskapen

\*\* Som over, men medianen samsvarer. Høyt gjennomsnitt skyldes noen få som har tilkjennegitt høye verdsettinger i CV/TP

©TØI 2001

Tabell 10.2: Andel (prosent) respondenter med 0 i verdsetting i CV/TP-sekvensen

	Pakke 2 = 0	Kortere reisetid = 0	Redusert ventetid = 0
Ikke leksikografisk samvalgsekvens 2	34	52	46
Sortert etter reisetid	22	44	33
Sortert etter frekvens	20	63	28
Sortert etter pris	65	79	74
Hele datasettet	37	56	48

©TØI 2001

## 10.2 FM versus samvalganalyser

Dersom det er slik at respondentene som velger leksikografisk etter pris i samvalgene reelt sett har 0 i betalingsvillighet<sup>18</sup>, så skulle de også i Frisch-sekvensen ha valgt laveste pris hele veien. Av de respondentene som valgte leksikografisk etter pris i samvalgene er det imidlertid bare 42,8 og 24,5 % som i FM hele veien velger laveste pris for hhv. redusert reisetid og redusert tid mellom avgangene.<sup>19</sup> Manglende samsvar mellom FM og SC tyder altså på at en betydelig andel av respondentene som velger leksikografisk mht pris i SC

gjør dette av andre grunner enn at de har 0 i betalingsvillighet for forbedringer i godene som inngår. En slik konklusjon støttes av resultatene i Sælensminde (2000) som også påpeker at respondenter med lav utdanning i større grad sorterer leksikografisk i SC og disse derfor har større problemer med å tilkjenne sine preferanser i slike kompliserte valgoppgaver.

Dersom FM gir et reelt anslag på andelen som virkelig har 0 i betalingsvillighet (som antydnet i konklusjonen over), synes FM å få frem en betalingsvillighet fra respondenter som ellers ville valgt leksikografisk for å forenkle valgoppgaven eller fordi prisforskjellene var for store. Dette kan bety at respondentene får bedre mulighet til å tilkjenne sine preferanser i FM enn i SC.

<sup>18</sup> Leksikografiske svar kan også skyldes at samvalgdesignet er ubalansert eller at respondentene forenkler beslutningsrutinene.

<sup>19</sup> Det er ikke noe entydig forhold motsatt vei (fra FM til samvalg) fordi vi også har et tredje gode (sitteplass) med i samvalgene. Respondenter som konsekvent har valgt laveste pris i FM for redusert reisetid og bedret frekvens kan altså ha tilkjennegitt betalingsvillighet for sitteplass i SC og derfor ikke ha valgt leksikografisk etter pris i samvalgene. Dessuten er det slik at respondenter som i FM konsekvent velger laveste pris og dermed ikke tilkjenner betalingsvillighet for forbedring, kan ha betalingsvillighet for å unngå forverring og derfor velger ikke-leksikografisk i samvalgene.

### 10.3 POG versus samvalganalyser

Vi testet om de som har valgt det dyreste alternativet i POG også har høyere verdsetting i SC-spillene. Tabell 10.3 viser resultatene.

Tabell 10.3: Verdsetting av ulike faktorer i samvalganalysene sortert etter tilkjennegitt betalingsvilje for den samme faktoren i POG-sekvensen. (Antall respondenter i parentes).

	Betalingsvilje i POG	Ikke betalingsvilje i POG
Realtidsinformasjon (RTI), kr/reise	4,05 (14)	1,08* (52)
Reisetid, kr/time	21,93 (16)	3,40* (50)
Reisetid m/ståplass, ekstra kostn./time	38,59 (20)	9,12 (46)
Frekvens, kr/time	86,38 (11)	34,74 (55)

\* Ikke signifikant forskjellig fra 0

©TØI 2001

Vi ser at verdsettingen blant dem som har valgt de dyreste alternativene i POG-sekvensen er klart høyere enn for dem som ikke har valgt det dyreste alternativet i POG. Blant dem som ikke tilkjennega betalingsvilje for reisetid og realtidsinformasjon i POG, er ikke samvalgberegningene signifikant forskjellige fra null.

Dette viser at det er samsvar mellom respondentenes svar i POG- og SC-sekvensen, men med ett vesentlig forbehold: Den beregnede verdsettingen av RTI på 1,08 kroner pr. reise for dem som ikke har betalingsvilje i POG ligger langt over POG-alternativet som kostet 50 øre. I samvalgene tilkjennegir de en verdsetting som altså er langt høyere enn det de *ikke* var villige til å betale i POG-sekvensen. Et slikt utslag er ulogisk og viser at det likevel er dårlig sammenheng mellom verdsettingene i de to metodene. Som vi har beskrevet tidligere skyldes dette i første rekke svakheter ved POG som metode. Dette viser igjen at vi kan få gode signaler om hvilke av de ulike egenskaper som er "viktige" for den enkelte respondent, på tvers av SP-metodene. Samtidig varierer den faktiske verdsettingen, målt i kroner pr. reise eller pr. time, betraktelig mellom metodene.

### 10.4 Frisch-sekvensene versus CV/TP

Testen av intern konsistens innenfor Frisch-metoden er blitt utvidet til å teste konsistensen mellom FM og CV/TP. Da viser det seg at 73 % av respondentene som i FM verdsetter redusert tid mellom avgangene høyere enn redusert reisetid også verdsetter redusert reisetid høyere enn redusert tid mellom avgangene i CV/TP. Men det er

bare 25 % av respondentene som i FM verdsetter redusert tid mellom avgangene lavere enn redusert reisetid som også verdsetter redusert reisetid lavere enn redusert tid mellom avgangene i CV/TP. Det er altså til dels liten grad av konsistens mellom respondenter som velger konsistent i FM og deres tilsvarende valg i CV/TP. Hvorvidt det har betydning for konsistensen mellom disse metodene om FM-sekvensene kom før eller etter CV/TP-spørsmålene, er ikke undersøkt.

Andelen med 0 i betalingsvillighet fra CV/TP-spørsmålene er for redusert reisetid og redusert tid mellom avgangene på hhv. 52,7 og 46,0 %. Dette er svært høye tall og kan bety både at respondentene ikke har betalingsvillighet for forbedringer i reisetid og frekvens eller at de svarer 0 av andre grunner (f.eks. protest mot å betale mer enn de allerede gjør eller ikke tror på de forbedringene som skisseres). Dersom det er slik at respondentene reelt sett har 0 i betalingsvillighet så skulle de i FM ha valgt laveste pris hele veien. I FM er det imidlertid bare 18,3 og 13,8 % av respondentene som hele veien velger laveste pris for hhv. redusert reisetid og redusert tid mellom avgangene.

Av dem som har valgt laveste pris i FM er det hele 86,4 og 84,4 % som også sier at de har 0 betalingsvillighet i CV/TP-spørsmålene. Dersom vi tar utgangspunkt i respondenter som konsekvent har valgt laveste prisalternativ i FM, synes det altså å være relativt godt samsvar mellom deres valg i FM og deres valg i CV/TP. Går en motsatt vei og tar utgangspunkt i respondenter som har 0 betalingsvillighet i CV/TP for redusert reisetid og redusert tid mellom avgangene, er det bare hhv. 32,5 og 27,7 % som konsekvent velger alternativet med laveste pris i FM.

Sammenlignet med CV/TP gir FM en betydelig reduksjon i andelen med 0 i betalingsvillighet, og 0 i betalingsvillighet i FM synes i større grad enn CV/TP å gi et reelt anslag på andelen som virkelig har 0 i betalingsvillighet. Hvorvidt respondentene virkelig har null betalingsvillighet eller har valgt null f.eks. som protest, kan ikke avgjøres uten kontrollspørsmål

### 10.5 Frisch-metoden, CV/TP og samvalganalyser

Tabell 10.4 viser avveininger mellom takst, reisetid og tid mellom avgangene fra FM, CV/TP og SC. Resultatene fra disse avveiningene er i samme størrelsesorden, samtidig som det er relativt store *prosentvise* forskjeller mellom verdsettingene i de ulike metodene. Følgende metodiske forskjeller og svakheter gjør at avveiningene ikke er helt enkle å sammenligne:

- I FM og CV/TP inngår bare forbedringer i reisetid og tid mellom avgangene når disse veies mot høyere takst, mens i SC inngår også forverringer. Betalingsvillighet for å unngå forverringer vil normalt være høyere enn betalingsvillighet for oppnå forbedringer. Dette kan ha gitt høyere verdsetting i SC enn i FM og CV/TP. (For eksempel er respondentene i CV/TP villig til å betale 0,75 kr/min for å unngå forverring i reisetiden.)
- I analysene av SC-dataene er det ikke tatt hensyn til at respondentene kan ha ulike preferanser og at den enkelte respondents valg kan være inkonsistente. Analyser uten å ta hensyn til variansforskjeller som skyldes inkonsistente valg kan medføre en høyere verdsetting fra SC-data enn analyser som tar hensyn til slike variansforskjeller.
- Avveiningene mellom takst, reisetid og tid mellom avgangene i FM og CV/TP er gjort ved at to og to faktorer er veid mot hverandre, mens i SC er avveiningene mellom alle tre faktorene gjort simultant. Dette kan isolert sett ha gitt lavere verdsettinger i SC enn i FM og CV/TP.
- Verdsettingene i FM og CV/TP er trolig i større grad påvirket av ekstremverdier, som f.eks. kan være utslag av strategiske valg, enn SC. Dette vil isolert sett kunne gi høyere verdsettinger fra FM og CV/TP enn SC.
- Verdsettingene i CV/TP er i stor grad påvirket av at andelen som tilkjenner 0 i verdsetting er stor.

Forskjeller mellom designen av metodene og svakheter ved metodene påvirker avveiningene som er gjort, og påvirkningene går altså i ulik retning. Det er dermed ikke uten videre grunnlag for å fastslå at den ene eller andre metoden gir for høy eller for lav verdsetting. Det er heller

ikke grunnlag for å fastslå hvilken metode som gir de mest riktige verdsettingene med bakgrunn i våre data.

Samtidig har vi påpekt godt samsvar mellom tidsverdiene som er beregnet i samvalganalysen her og tidsverdier fra tidligere studier, noe som tyder på at samvalgene gir stabile beregninger for verdsetting av kvalitet. Vi har også påpekt at avveiningene (verdsettingene) er i samme størrelsesorden (dog med stor prosentvis variasjon), og dermed at det *er* relativt godt samsvar mellom verdsettingsmetodene.

Tabell 10.4 sammenligner også andelen "0-svar." Dette er ikke noen enkel sammenlignbar størrelse på grunn av de forskjellige metodene, men kan likevel gi en pekepinn på hvordan respondentene har taklet de ulike valgsituasjonene. Vi har definert 0-svar slik:

- FM: Andel respondenter som aldri velger beste alternativ for egenskapen, men alltid alternativet med laveste pris.
- CV/TP: Andel som verdsetter egenskapen til 0 kroner.
- SC: Andelen som sorterer etter de *andre* egenskapene.

Vi ser at andelen 0-svar i stor grad avhenger av SP-metode. Når vi bare ser på 0-svarene, kan vi derfor ikke si at det er stor grad av konsistens på tvers av de ulike metodene. Vi konstaterer at alle metodene har høyere andeler 0-svar for reisetid enn for tid mellom avgangene, men at i CV/TP ligger andelen på et langt høyere nivå enn de andre metodene. Vi merker oss også at det er noen grad av samsvar mellom andelen 0-svar i FM og SC. Respondentene har trolig hatt lettere for å tilkjenne sine faktiske preferanser i disse sekvensene.

Tabell 10.4: Avveininger (betalingsvillighet)<sup>1)</sup>, mellom takst, (reduisert) reisetid og (reduisert) tid mellom avgangene, og andeler "0-svar"<sup>2)</sup> fra FM, CV/TP og SC

		FM (N=289-298)	CV/TP (N=298)	SC (N=298)
(Redusert) reisetid (kr/min)	Verdsetting	0,53 – 0,65	0,31 – 0,34	0,38
	Andel 0-svar	18,3%	55,7%	25,6%
(Redusert) tid mellom avgangene (kr/min)	Verdsetting	0,53 – 0,56	0,37 – 0,47	0,56
	Andel 0-svar	13,8%	48,2%	20,2%
Forholdet Frekvens/Reisetid		1,06 – 1,17 <sup>2)</sup>	1,19 – 1,38 <sup>3)</sup>	1,47 <sup>4)</sup>

<sup>1)</sup> Øvre grense i intervallene er avveiningene som fremkommer når alle respondentene er inkludert. Nedre grense i intervallene er avveiningene når respondenter som tilkjenner betalingsvillighet på mer enn tre ganger taksten er ekskludert

<sup>2)</sup> 0-svar er definert slik: FM: andel respondenter som aldri velger beste alternativ for egenskapen, men alltid laveste pris. CV/TP: andel med verdsetting av egenskapen lik 0 kroner. SC: Andelen som sorterer etter andre egenskaper

<sup>3)</sup> Beregnet som forholdet mellom betalingsvilligheten for hhv. redusert tid mellom avgangene og redusert reisetid

<sup>4)</sup> Eksplisitte avveininger

Når en skal trekke konklusjoner mht. valg av metode for fremtidige studier bør imidlertid to forhold telle tungt. Det første er i hvilken grad påviselige svakheter ved den enkelte metode kan reduseres og at den enkelte metode på den måten gi sikrere resultater. Det andre er i hvilken grad de ulike metodene gir mulighet for å verdsette ikke-markedsgodene i en kontekst som er i samsvar med hvordan verdsettingene er tenkt brukt (f.eks. i nytte-kostnadsanalyser).

## 10.6 Oppsummering

Det er generelt godt samsvar mellom metodene når det gjelder hvilke egenskaper som er viktigst for de enkelte respondentene. Derfor gir både samvalg, betinget verdsetting/likeverdpris og fleksibelt pakkevalg gode indikasjoner på hva den enkelte trafikant legger vekt på. For eksempel ser vi at de som har lagt stor vekt på en egenskap i POG også har gjort det i samvalgsekvensen og i CV/TP. Tilsvarende har de som sorterte etter en egenskap i samvalgene en høyere verdsetting av den

samme egenskapen i CV/TP. Av dette kan vi slutte at de enkelte trafikantene har relativt klare prioriteringer som kommer til syne uavhengig av SP-metode.

Samtidig er det store prosentvise sprik mellom de beregnede verdsettingene av kvalitetsfaktorer mellom de ulike SP-metodene. For eksempel er reisetid verdsett til kr 0,38 pr minutt i samvalganalysen mot hele kr 0,53-0,65 pr minutt i Frisch-metoden.

Det er lite sammenheng mellom 0-verdsettinger i CV/TP og FM og leksikografiske svar i samvalgene. Vi forventer at de som *ikke* tilkjenner betalingsvilje for en kvalitetsfaktor i CV/TP og FM, også legger større vekt på de øvrige kvalitetsfaktorene i samvalgsekvensen. I stor grad har dette ikke vært tilfelle, og trafikantene har således i noen grad svart inkonsistent på tvers av SP-metodene. Vi har konkludert med at egenskaper ved spørsmålstillingen i de ulike metodene, og spesielt CV/TP og POG har forårsaket dette.

# 11 Oppsummering, vurdering av SP-metodene og konklusjoner

Dette prosjektet har hatt som formål å kritisk gjennomgå design av *Stated Preference*-undersøkelser for å avdekke måter designet påvirker resultatet og gi anbefalinger til hvordan problemet kan reduseres.

I forbindelse med prosjektet er det blitt gjennomført en pilotstudie for Oslo Sporveier blant Oslos kollektivreisende. Til sammen har dette gitt et datamateriale som omfatter 364 respondenter. Det er blitt utviklet 10 ulike design for undersøkelsen, som er blitt testet og analysert i dette prosjektet. Vi har definert *design* som:

- **Utvalgs- og rekrutteringsmetode.** I studien er rekrutteringen fordelt mellom 1) telefonrekruttering (gjennomført av Gallup) av et representativt utvalg av Oslobefolkningen, og 2) rekruttering via Trafikanten og Oslo Sporveiers nettsider. Antallet respondenter har variert mellom 19 og 55 i de 10 designene.
- **Metode for datainnsamling.** Datainnsamlingen har vært fordelt mellom "tradisjonelle" hjemmeintervjuer og egenadministrerte, PC-baserte intervjuer på diskett som respondentene har fått tilsendt.
- **SP-metoder.** I undersøkelsen har følgende metoder vært benyttet:
  - Parvise samvalg (*Stated Choice*, SC)
  - Betinget verdsetting (*Contingent valuation*, CV) og likeverdspris (*Transfer Price*, TP)
  - Frisch-metoden (FM), en metode som "sirkler inn" en respondents avveining mellom to og to egenskaper. I en serie valg mellom to reisealternativer blir forskjellene i nivåene på egenskapene gradvis redusert i retning av respondentens preferanser.
  - "Fleksibel pakkesammensetning", (*Package Option Grid*, POG) som er en mulighet i programpakken MINT der respondenten selv kan velge den kombinasjonen av godeforbedringer (både hvilke goder og hvilket nivå) som gir størst nytte. Samtidig må respondenten ta hensyn til den totale kostnaden for pakken av goder som velges.
- **Struktur på spørreskjema.** Vi har variert rekkefølgen på de ulike SP-sekvensene, og hvilke SP-metoder som har inngått i intervjuet.

- **Utforming av de enkelte SP-metodene.** Her har vi konsentrert oss om de to samvalgsekvensene. Sekvens 1 inkluderte info på holdeplassen, lehus, trygghet og pris. Sekvens 2 inkluderte reisetid, frekvens sitteplass og pris. Følgende 6 varianter har vært brukt:

1. Standarddesign: 4 kvalitetsfaktorer med tre nivåer hver og opptil 9 valg. Preferanseusikkerhet (kan velge "Absolutt alternativ X" eller "Trolig alternativ X").
2. Balansering av nivåene ved hjelp av CV/TP.
3. CV/TP-balansering, samt svaralternativ "Vet ikke" og "Like gode".
4. Standarddesign, og i tillegg et eget blokkdesign med grupper av faktorer på flere nivåer.
5. CV/TP-balansering. Ta med opptil 15 valg for å studere tretthetseffekter.
6. Redusere antall kvalitetsfaktorer fra 4 til 3.

Tabell 11.1 beskriver de 10 ulike designene for spørreundersøkelsen. De fordeler seg på 6 ulike designvarianter for samvalgene, to rekrutteringsmetoder og to datainnsamlingsmetoder. I tillegg er rekkefølgen på de ulike SP-sekvensene variert.

Tabell 11.1: Oversikt over de ulike designnumrene. "Designvariant" viser til varianten av samvalgsekvensen

Design-variant for samvalg	Rekkefølge på SP-sekvensene	a) diskett pr post telefonrekruttert	b) diskett pr post selvrekruttert (internett)	c) hjemmeintervju forhåndsrekruttert
I	SC, POG, FM			1
II	CV/TP, SC, POG			2
III	CV/TP, SC, FM	3	13	
IV	SC, FM, CV/TP	4	14	
V	CV/TP, SC, FM	5	15	
VI	FM, SC, CV/TP	6	16	

©TØI 2001

I dette kapitlet vil vi oppsummere testene som er gjennomført og trekke konklusjoner på grunnlag av sammenlignende vurderinger av SP-metodene og metodene for datainnsamling.

Hovedproblemstillingen i dette prosjektet har vært todelt. Vi har sett på muligheten for å utvikle tester som avdekker hva som er god eller dårlig design på SP-undersøkelser og om disse testene kan benyttes for å vurdere hvilke av designene i denne undersøkelsen som har fungert best, og om egenadministrerte PC-intervjuer fungerer like bra som hjemmeintervjuer og i hvilken grad dette vil gi skjevheter i utvalget.

## 11.1 Testing av utvalget

Utvalget av respondenter til en generell verdsettingsstudie må trekkes tilfeldig fra et representativt utvalg av befolkningen. Andre metoder vil medføre skjevheter.

Utvalget som er trukket av Gallup er i stor grad representativt for befolkningen, men det er en viss forskjell mellom utvalget som godtok hjemmeintervju og utvalget som fikk tilsendt disketter for egenadministrerte intervjuer.

Selv om enkelte av våre delutvalg har vært svært små, har det i seg selv ikke forårsaket store tilfeldige variasjoner i egenskaper ved respondentene og reisene de foretok. Imidlertid er det påfallende forskjeller på grunn av rekrutteringsmetode og (i mindre grad) av metode for datainnsamling.

Frafallet blant dem som fikk tilsendt diskett for egenadministrerte intervjuer ligger jevnt på ca 50 %. Denne andelen kan reduseres ved å utvikle mer brukervennlige opplegg og løsninger som ikke avhenger av operativsystem, som internettbaserte løsninger.

## 11.2 Tester av samvalgsekvensene

### 11.2.1 Verdsetting av kvalitet

I samvalgsekvensene er det til dels store sprik i verdsettingen av kvalitetsfaktorer mellom de ulike designvariantene. Tabell 11.2 viser et utdrag av beregningene (design 13-16 er utelatt) og illustrerer sprikene i verdsettinger. Blant reisetidsfaktorene (de tunge kvalitetsfaktorene) er det ingen verdsettinger som har galt fortegn, mens flere av de mykere kvalitetsfaktorene har negativ verdsetting i en del av designene. Mange av forskjellene kan skyldes tilfeldig variasjon i de små utvalgene, men vi har gjennomført flere tester for å undersøke om forskjellene skyldes egenskaper ved selve designet. (Vurdering av egenadministrerte intervjuer versus hjemmeintervjuer er beskrevet i pkt. 11.9.

### 11.2.2 Høy andel leksikografi skyldes i stor grad manglende balansering

Leksikografi er et alvorlig metodeproblem i samvalg-analyser. Når alle valgene til en respondent faller på det alternativet der den samme kvalitetsfaktoren er best, kan det ha ulike årsaker:

- Manglende balansering av egenskapsnivåene i samvalgsekvensen, slik at det er et uttrykk for respondentens faktiske (ev. ekstreme) preferanser.
- Respondenten forenkler beslutningsrutinen og lar valgene bestemmes av én kvalitetsegenskap. Dette medfører at respondenten signaliserer en større vektlegging av denne egenskapen enn vedkommendes egentlige preferanser tilsier.

Tabell 11.2: Beregnede verdsettinger (kroner pr time eller pr reise) av egenskapene fordelt på ulike design og datainnsamlingsmetoder

Kvalitetsfaktor	Design 1	Design 2	Design 3	Design 4	Design 5	Design 6	Hjemme-intervju	Egen-adm.	Alle
Reisetid, kr pr time	7	9	20	17	10	25	8	26	23
Ståplass, tillegg/time	15	17	17	15	11		16	12	13
Ventetid, kr pr time	77	88	69	61	48	44	82	66	68
Vektore, kr pr reise	2	1	-1	-1	0	2	1	0	0
Nødtelefon, kr pr reise	1	1	0	1	-3	-0	1	0,	0
Overbygg, kr pr reise	1	0	-0	0	-5		1	-1	-1
Leskur, kr pr reise	2	2	1	1	1		2	1	1
Realtid, kr pr reise	1	0	1	1	3	2	1	2	1
Rutekart, kr pr reise	1	1	-1	0	-4	-1	1	-1	-1
$\rho^2$ sekvens 1/2	0,2/0,2	0,2/0,2	0,3/0,3	0,3/0,2	0,2/0,1	0,1/0,2	0,2/0,2	0,2/0,2	0,2/0,2
Ant. respondenter	35	31	53	52	44	55	66	298	364



I samvalgsekvens 1, som omfattet de "mykere" kvalitetsfaktorene, er det gjennomgående langt høyere andeler respondenter som har sortert svarene leksikografisk enn i sekvens 2. Samtidig har sekvens 1 fungert dårlig med hensyn til estimering av kvalitetsfaktorene. En bedre balansering av nivåene på kvalitetsfaktorene i sekvens 1 ville kunne forbedre analysene betraktelig. Andelen leksikografiske svar er derfor en viktig indikator for hvordan undersøkelsen har fungert, og bør inngå som standard rapportering av SP-undersøkelser.

I noen av samvalgdesignene har variasjonen rundt basisprisen blitt bestemt av respondentenes svar i den forutgående CV/TP-sekvensen. Dette gir nærmest ingen signifikante forskjeller i verdsettingene, men det er en klar tendens til lavere andeler leksikografi i disse sekvensene. Dette opplegget for balansering av egenskapsnivåene i samvalgsekvensen har derfor hjulpet til med å redusere et viktig metodeproblem.

Det er en viss systematikk i de leksikografiske svarene, som tyder på at en del av respondentene ikke takler valgsituasjonen i samvalgene. Vi anbefaler at det gjøres ytterligere analyser av disse respondentene for å se om de har felles karakteristika, slik at det kan gjøres mulig å fange dem opp og tilrettelegge undersøkelsen til deres situasjon.

### 11.2.3 Antallet faktorer har mindre betydning

I design 6 og 16 inngår det bare tre kvalitetsfaktorer i samvalgsekvensen (i motsetning til fire i de øvrige designene). Vi har ikke funnet grunnlag ut fra våre data til å si noe om hvordan antallet egenskaper som inngår i samvalgsekvensen påvirker verdsettingen.

### 11.2.4 Lange samvalgsekvenser skaper mer usikkerhet rundt resultatene

Samvalgsekvenser med 15 valgoganger (design 5 og 15) oppfattes av respondentene som vanskeligere enn sekvenser med opptil 9 valgoganger. Disse modellene har i tillegg fungert relativt dårlig mht de vanligste indikatorene for forklaringskraft, samtidig som en rekke av de mykere kvalitetsfaktorene fått galt (altså negativt) fortegn. Dette viser at 15 valgoganger er for mange. I stedet for å gi oss mer informasjon, vil en økning i antallet valgoganger ut over 9 snarere skape mer usikkerhet rundt beregningene. Vi har videre påvist at tretthet på grunn av lange samvalgsekvenser ikke bare påvirker den enkelte samvalgsekvensen, men også senere samvalgsekvenser i samme intervju.

Vi har også studert forekomsten av avvikende og sterkt avvikende svar i samvalgene, definert som valg det

er mindre enn henholdsvis 10 og 5 prosents sjanse for at respondenten skal velge. Ut over lengden på samvalgsekvensene, målt i antallet valgrunder, er det ingen tegn til at de ulike designene som inngår i analysene har gitt økt forekomst av avvikende eller sterkt avvikende svar.

### 11.2.5 Tretthet hos respondentene

I analysene av tretthetstegn i samvalgsekvenser med 9 valgoganger har vi identifisert en større fokus på pris, og dermed lavere verdsetting av kvalitetsfaktorene i de siste tre spørsmålrundene enn i de seks forutgående spørsmålene i samvalgsekvensen. Sortering etter pris (eller en annen egenskap) kan derfor være et tretthetstegn. Vi vil anbefale at dette i fremtiden undersøkes nærmere ved sammenligning med andre datasett.

Tretthet i samvalgsekvensene med bare ni valg skyldes i stor grad at to samvalgsekvenser kom fortløpende i spørreskjemaet og at respondentene dermed ble eksponert for opptil 18 svært like skjermbilder og avveininger på en gang.

### 11.2.6 Preferanseusikkerhet gir mer informasjon

Samvalganalysene åpnet for preferanseusikkerhet. Basert på de modellene der vi har et tilstrekkelig stort data-materiale kan vi konkludere at sammenligner vi modellene med sikre og mindre sikre valg, gir modellene med respondentenes sikre valg:

- Bedre forklaringskraft
- Mer signifikante parameterestimater
- Riktig fortegn på signifikante parameterestimater (der dette tidligere var et problem)
- Lavere verdsettinger av kvalitetsfaktorene

Ved å gi respondentene muligheten til å oppgi på en skala hvor sikre de er på valgene sine, får vi muligheten til å hente mer informasjon fra hver observasjon, ved at det blir mulig å vekte valgene i analysene.

### 11.2.7 Pakkeeffekter og kontekst

Forsøk med blokkdesign for å fange opp pakkeeffekter (at verdsettingen av en *pakke med forbedringer* er forskjellig fra summen av de individuelle verdsettingene av forbedringene som inngår) viser at det kan være en stor utfordring å balansere blokker av kvalitetsfaktorer som å balansere enkeltfaktorer i en valgsekvens, og vi har ikke klart å balansere faktorene godt nok i denne studien. Til tross for dette kan vi konkludere med at:

- Faktorene relatert til reisen (sitteplass, reisetid og tid mellom avgangene) verdsettes opptil 35 % lavere når de forbedres samtidig og veies mot andre faktorer (i en pakke/blokk) enn når de bare veies mot hverandre enkeltvis.
- Faktorene relatert til holdeplassen (trygghet, informasjon og overbygg) ser også ut til å verdsettes lavere når de forbedres samtidig og veies mot andre faktorer (i en blokk) enn når de bare veies mot hverandre enkeltvis, men på grunn av dårlig balanserte faktorer er dette ikke signifikante resultater.

### 11.3 Tester av CV/TP-sekvensene

I CV/TP-sekvensene har vi fått respondentenes verdsetting av:

1. Forbedret reisetid
2. Forverret reisetid
3. Forbedring fra sitte halve reisen til sitte hele reisen
4. Forbedring stå hele til sitte hele veien
5. Forverring fra sitte halve reisen til stå hele veien
6. Forverring fra sitte hele reisen til stå hele veien
7. Redusert ventetid
8. Pakke 1: Realtidsinformasjon, overbygg/tak, nødtelefon og vektore
9. Pakke 2: Kort reisetid, sitteplass, høy frekvens

Vi har påvist følgende sammenhenger mellom egenskaper ved designet og CV/TP-verdsettinger av kvalitetsfaktorer:

- Effekten av rekrutterings- og datainnsamlingsmetoden på verdsettingene ser ut til å ha vært nøytral.
- Respondenters tendens til å avrunde svarene sine (i vårt tilfelle til 0, 5, 10, 15 osv. kroner) påvirker resultatene og skaper skjevheter i deres verdsetting av kvalitet.
- I spørsmål om betalingsvilje for en *forverret* reise kan verdsettingen av den forverrede egenskapen bli veldig høy, bl.a. fordi det ikke er spesifisert noen nedre grense for hva respondenten kan være villig til å betale for en forverret reise. Dette kan være en viktig årsak til at forverringer tillegges uforholdsmessig mye større vekt enn forbedringer i CV/TP-beregningene.
- Spørsmålsstillingen i CV/TP er direkte og har ingen elementer av avveininger, noe som gjør metoden utsatt for protestvalg og strategiske svar. Dette har gitt utslag i flere observasjoner av 0 i *betalingsvilje* (gir høy verdsetting av en egenskap som forverres),

en stor andel med 0 i *verdsetting* (opptil 86 % for en reisetidsforbedring), og en del *overbud*, som gir ekstremt høye verdsettinger av forbedringer. Disse effektene har påvirket de gjennomsnittlige verdsettingene i stor grad.

- I intervjuer hvor CV/TP-sekvensen kommer etter de andre typene verdsettingsspørsmål, er det en tendens til at CV/TP-beregningene er mer stabile og i mindre grad ekstreme. Det er også lavere andeler 0-svar i CV/TP-sekvensene i disse intervjuene.

### 11.4 Tester av Frisch-metoden

I fire ulike FM-sekvenser har respondentene gjort følgende avveininger:

1. Redusert reisetid mot lengre tid mellom avgangene (dvs. Redusert frekvens)
2. Redusert reisetid mot høyere takst
3. Mindre tid mellom avgangene (dvs. Økt frekvens) mot høyere takst
4. Mindre tid mellom avgangene (dvs. Økt frekvens) mot lengre reisetid

Mellom 1 og 3 % av respondentene valgte alternativet med høyeste betalingsvillighet i første valg. Vi vet ikke om disse 1-3 prosentene av respondentene som velger å tilkjennegi ekstremt høy betalingsvillighet, faktisk har slike preferanser. Mer bruk av kontrollspørsmål er derfor en viktig videreutvikling av metoden.

Bare 14 og 18 % av respondentene valgte laveste pris i alle FM-valgene, dvs. ville ikke godta høyere pris for å få hhv. redusert reisetid eller høyere frekvens. Vi har antatt at disse har 0 i betalingsvillighet, selv om en mindre andel av dem kan ha hatt en marginal betalingsvilje som ikke er fanget opp i designet. Dette viser at problemene med respondenter med 0 i betalingsvilje, eller som sorterer etter én kvalitetsfaktor hele veien, er betraktelig redusert i FM i forhold til de øvrige SP-metodene. En stor andel respondenter foretar faktiske avveininger i FM.

### 11.5 POG virker uegnet for verdsettingsstudier

Vi har ikke klart å gjøre nytte av POG (*Package Option Grid*, eller fleksibelt pakkevalg) i sin nåværende form. I våre analyser er POG kun egnet til å kartlegge hvilke kvalitetsfaktorer trafikantene prioriterer. Med mindre metodene knyttet til POG utvikles videre er POG derfor ikke egnet til verdsettingsstudier.

## 11.6 Noe konsistens mellom SP-metodene

Vi har sett på sammenhenger mellom hvordan respondentene har svart i Frisch-, POG-, CV/TP- og samvalgsekvansene. Hvis det er stor grad av samsvar mellom verdsettingen av de ulike egenskapene, vil det si at respondentene har taklet intervju situasjonen og at de klarer å tilkjenne sine preferanser og verdsettinger. Samtidig har vi studert respondenter som har tilkjennegitt spesielle verdsettinger, f.eks. 0-verdsettinger for å se om de svarer konsistent mellom SP-metodene.

Tabell 11.3 viser et eksempel på verdsettinger som er beregnet i FM, CV/TP og samvalgene. Resultatene fra disse metodene er i samme størrelsesorden, men likevel med prosentvise store sprik.

Det er generelt godt samsvar mellom metodene når det gjelder hvilke egenskaper som er *viktigst* for de enkelte respondentene. For eksempel har respondenter som sorterte etter én egenskap i samvalgene en høyere verdsetting av den samme egenskapen i CV/TP. Av dette slutter vi at de enkelte trafikantene har relativt klare

prioriteringer som kommer til syne uavhengig av SP-metode og design.

Samtidig er det lite sammenheng mellom 0-verdsettinger i CV/TP og FM og leksikografiske svar i samvalgene. Vi forventer at de som *ikke* tilkjenner betalingsvilje for en kvalitetsfaktor i CV/TP og FM, også legger større vekt på de øvrige kvalitetsfaktorene i samvalgsekvansen. I stor grad har dette *ikke* vært tilfelle, og trafikantene har således i noen grad svart inkonsistent på tvers av SP-metodene. Egenskaper ved spørsmålstillingen i de ulike metodene, og spesielt CV/TP og POG har forårsaket dette.

## 11.7 Struktur på spørreskjemaet

I undersøkelsen ble rekkefølge av SP-metodene i spørreskjemaet variert. Vi har ikke trukket noen klare konklusjoner med hensyn til rekkefølgen av de ulike sekvensene, men har observert at verdsettingene i CV/TP blir mer stabile, det vil si mindre grad av 0- og ekstreme svar, i design der CV/TP-sekvensen kom lenger ut i intervjuet.

Tabell 11.3: Avveininger (betalingsvillighet)<sup>1)</sup>, mellom takst, (reduisert) reisetid og (reduisert) tid mellom avgangene, og andeler "0-svar"<sup>2)</sup> fra FM, CV/TP og SC

		FM (N=289-298)	CV/TP (N=298)	SC (N=298)
(Redusert) reisetid (kr/time)	Verdsetting	32– 39	19 – 20	23
	Andel 0-svar	18%	56%	26%
(Redusert) tid mellom avgangene (kr/time)	Verdsetting	32 –34	22 – 28	34
	Andel 0-svar	14%	48%	20%
Forholdet Frekvens/Reisetid		1,06 – 1,17 <sup>4)</sup>	1,19 – 1,38 <sup>3)</sup>	1,47 <sup>4)</sup>

<sup>1)</sup> Øvre grense i intervallene er avveiningene som fremkommer når alle respondentene er inkludert. Nedre grense i intervallene er avveiningene når respondenter som tilkjennega betalingsvillighet på mer enn tre ganger taksten er ekskludert.

<sup>2)</sup> 0-svar er definert slik: FM: andel respondenter som aldri velger beste alternativ for egenskapen, men alltid laveste pris. CV/TP: andel med verdsetting av egenskapen lik 0 kroner. SC: Andelen som sorterer etter *andre* egenskaper.

<sup>3)</sup> Beregnet som forholdet mellom betalingsvilligheten for hhv. redusert tid mellom avgangene og redusert reisetid.

<sup>4)</sup> Eksplisitte avveininger.

## 11.8 Sammenlignende vurdering av metodene

### 11.8.1 Samvalgene har vesentlige fordeler i forhold til CV/TP og FM

Det er to vesentlige egenskaper ved samvalganalysen som veier tungt i valget av SP-metode, og som er viktige årsaker til at samvalganalysen oftest er foretrukket i verdsetningsstudier for kollektivtransporten.

Det første dreier seg om konteksten for verdsetningsstudien. Med unntak av POG, som vi har vist at er en lite egnet metode for verdsetningsstudier, er det bare samvalganalysene som kan brukes til å beregne verdsetting av flere enn én kvalitetsfaktor (i forhold til en annen) i samme sekvens. Konteksten for verdsettingen er helt vesentlig og legger begrensninger med hensyn til hvordan resultatene kan brukes, dvs. hvilket gyldighetsområde verdsettingene har. Samvalgene gir derfor en langt bedre mulighet til å verdsette kvalitetsegenskaper innenfor riktig kontekst.

Det andre dreier seg om problemet med taktiske svar og protestvalg. I samvalganalysen er valgalternativene presentert slik at det ikke uten videre er mulig å protestere ved å gi én type svar. Vi har sett at leksikografiske svar er et vesentlig problem innenfor samvalganalysen, men dette er ikke nødvendigvis uttrykk for protestvalg. Leksikografi kan også skyldes manglende balansering av nivåene og dermed være et uttrykk for respondentenes faktiske preferanser, eller være et resultat av forenklete beslutningsrutiner hos respondenten. Den enkle og mer direkte spørsmålsstillingen i CV/TP og FM gjør det lettere for respondentene å protestere eller svare taktisk.

### 11.8.2 Mindre problemer med 0-svar og leksikografi med Frisch-metoden

Om FM gir bedre data enn andre SP-metoder er forsøkt besvart ut fra testene av hvordan FM fungerer med hensyn til å redusere en del av de mest alvorlige problemene ved CV/TP og SC. Spørsmålet er altså om vi kan konkludere med at:

- FM reduserer problemene i CV/TP der en stor andel av respondentene angir 0 i betalingsvillighet og der en del respondenter angir urealistisk høy betalingsvillighet?
- FM reduserer problemene med leksikografiske og inkonsistente valg i SC?

Svarene på disse spørsmålene er trolig ja, men med et par gode kontrollspørsmål kunne vi vært sikrere på om denne konklusjonen var riktig.

### 11.8.3 FM ser med fordel ut til å kunne erstatte CV/TP

Resultatene så langt tyder på at FM har flere fordeler enn ulemper sammenlignet med CV/TP og at FM derfor bør kunne erstatte CV/TP i fremtidige verdsetningsstudier i kollektivtransporten.

Sammenlignet med CV/TP gir FM en betydelig reduksjon i andelen med 0 i betalingsvillighet, og 0 i betalingsvillighet i FM synes i større grad enn CV/TP å gi et reelt anslag på andelen som *virkelig har 0* i betalingsvillighet.

Denne studien gir imidlertid ikke svar på om de lavere verdsettingene fra CV/TP tyder på at respondentene i større grad tar hensyn til sin betalingsevne i CV/TP enn i FM. Dersom det er slik at FM reduserer andelen respondenter som velger null av protestgrunner, kan dette forklare forskjellene i verdsettingen mellom FM og CV/TP. En totalvurdering gir grunn til å tro at en ved bruk av FM får et riktigere anslag på respondentenes betalingsvillighet enn ved bruk av CV/TP.

### 11.8.4 FM kan oppfattes som et supplement til SC Undersøkelsen har vist at:

- Respondenter som i SC ville valgt inkonsistent vil trolig lettere kunne tilkjenne sine preferanser i FM.
- FM kan få frem verdsettinger fra respondenter som i SC velger leksikografisk for å forenkle eller på grunn av at det var for stor avstand mellom nivåene på faktorene.
- FM gjør det lettere å følge opp med kontrollspørsmål om motiver for betalingsvilligheten til respondenter som har avvikende verdsettinger fordi avveiningene mellom godene fremkommer mer direkte enn i SC.

Dette er fordeler ved FM som retter opp viktige problemer/svakheter ved SC. Det kan imidlertid være grunn til å tro at det er lettere å få til en riktig kontekst for verdsettingen ved hjelp av SC der avveining mellom flere goder kan gjøres simultant. Med riktig kontekst menes her en kontekst som er samsvarende med konteksten resultatene skal brukes i. Dersom denne konteksten er kompleks, f.eks. ved at endringer i mange goder skal avveies mot hverandre i en nytte-kostnadsanalyse, øker imidlertid faren for at verdsettingene påvirkes av både inkonsistente og leksikografiske valg i SC. En vurdering

av om SC som metode er å foretrekke fremfor FM er avhengig av om en i analysene av SC-data er i stand til å håndtere slike problematiske valg.

Inntil det er gjort flere sammenlignende metodestudier ser FM ut til å kunne fungere som et godt supplement til SC i SP-studier. FM og SC er på mange måter identiske metoder hvis en utvider antall faktorer i FM til 4. Fordelen med Frisch-metoden er fleksibiliteten som ligger i presentasjonen av nivået på kvalitetsfaktorene. Ideelt sett bør det utvikles samvalgdesign med like stor fleksibilitet i nivåene (særlig prisnivået), som gjennom algoritmer strammes inn avhengig av de valgene som foretas. Foreløpig er vi ikke kjent med noen program-pakke som gjør nettopp det.

Balansering av nivåene i SC ved hjelp av CV/TP ga en tendens til lavere andeler leksikografiske svar, samtidig som svakheter ved CV/TP-metoden gjør at slik balansering kan forverre samvalgdesignet. Med de fordelene som ligger i FM i forhold til CV/TP, anbefaler vi at FM benyttes til nivå-skalerting i samvalgsekvenser heretter.

#### 11.8.5 Respondentene oppfatter samvalgene som vanskelig

For å undersøke i hvilken grad det er forskjeller mellom metodene med hensyn til hvordan respondentene vurderer hvor vanskelig det er å gjøre de valgene metodene krever, ga vi dem et kontrollspørsmål etter at de hadde gjennomført hver enkelt metode. I kontrollspørsmålene spurte vi hvor vanskelig det var å svare på henholdsvis FM, CV/TP og SC. Den prosentvise fordelingen av svarene på disse kontrollspørsmålene er presentert i Tabell 11.4. Utvalget er de 298 respondentene som gjennomførte både FM, CV/TP og SC.

Tabell 11.4: Prosentvis fordelingen av svarene på kontrollspørsmål om hvor vanskelig respondentene synes det var å svare på hhv. Frisch-metoden, betinget verdsettning/likeverdspris og samvalg. (N=298.)

	FM	CV/TP	SC
Svært vanskelig	4,4	6,0	7,7
Litt vanskelig	27,9	34,6	47,3
Nokså lett	51,7	38,9	39,3
Svært lett	16,1	20,5	5,7

©TØI 2001

For å teste om det er signifikante forskjeller mellom metodene med hensyn til hvordan respondentene vurderer hvor vanskelig det er å gjøre de valgene metodene krever har vi benyttet en ikke-parametrisk test kalt ”Wilcoxon

*Signed Ranks Test*” i SPSS.<sup>20</sup> Dette er en test som undersøker sannsynligheten for at svarene fra de ulike kontrollspørsmålene kan sies å ha samme fordeling. Nullhypotesen er altså at svarene fra kontrollspørsmålene følger samme fordeling.

Den statistiske testen viser at nullhypotesen kan forkastes når SC testes mot både CV/TP og FM. Spørsmålene i SC oppleves altså som signifikant vanskeligere å besvare enn spørsmålene i CV/TP og FM. Nullhypotesen kan imidlertid ikke forkastes når CV/TP og FM testes mot hverandre. Respondentene har altså ikke opplevd at det å besvare spørsmålene i CV/TP har vært signifikant vanskeligere eller lettere enn i FM.

## 11.9 Egenadministrasjon fungerer tilfredsstillende

Hjemmeintervjuer er en svært ressurskrevende metode for datainnsamling. I dette prosjektet har vi derfor lagt vekt på å vurdere muligheten for, og kvaliteten av, egenadministrerte intervjuer som respondentene gjennomfører på egenhånd. Våre undersøkelser viser at egenadministrasjon av intervjuene, ved at respondentene får tilsendt en diskett med et PC-basert intervju, har fungert rimelig godt.

I rekrutteringen til undersøkelsen var det en noe større andel som vegret seg mot å få tilsendt diskett enn andelen som vegret seg mot å få besøk av en intervjuer – henholdsvis 5 og 1 % av dem som var villige å delta i en undersøkelse og som tilfredsstilte kriteriene til å delta. Noe av forskjellen kan skyldes feil koding fra Gallup. Samtidig er andelen vellykkede telefonrekrutteringer høyere i rekrutteringen til egenadministrerte intervjuer.

Det har også vært et ikke ubetydelig frafall blant dem som har fått tilsendt diskett, på rundt 50 %. En stor del av disse (om lag 10 %) har skyldtes tekniske problemer med gjennomføringen på datamaskiner (diskett, operativsystem osv.). En mer brukervennlig PC-basert løsning, som internett, ville derfor ha redusert frafallet.

Vi har funnet indikasjoner på at en del kjennetegn ved respondentene er fremtredende blant dem som administrerte intervjuene selv. De er i større grad menn, har lavere gjennomsnittsalder, reiser mer med lokaltog, har flere arbeidsreiser og benytter periodekort oftere enn utvalget for hjemmeintervjuene. Et opplegg med bare egenadministrerte intervjuer vil dermed kunne gi resul-

<sup>20</sup> Ved bruk av ”Wilcoxon Signed Ranks Test” forutsettes den *underliggende variabelen* å være kontinuerlig. I vårt tilfelle kunne vi brukt en finere responskala (f.eks. fra 0 til 100) for å kartlegge vanskelighetsgrad og på den måten fått en kontinuerlig variabel. Wilcoxon-testen er derfor egnet til tross for at vi bare hadde fire responskategorier for våre kontrollspørsmål om vanskelighetsgrad.

tater som ikke representerer hele befolkningen. Det er derfor helt nødvendig å supplere egenadministrerte intervjuer med tradisjonelle hjemmeintervjuer for å sikre representasjon fra bl.a. de høyere aldersgruppene.

I samvalgsekvensene er det ingen entydige tegn på om de egenadministrerte intervjuene har fungert bedre eller dårligere enn hjemmeintervjuene med hensyn til verdsettinger og signifikanstester. Verdsettingene av de tyngre kvalitetsfaktorene har på mange måter fungert bedre i de egenadministrerte intervjuene – lavere andel leksikografi og verdsettinger mer i tråd med tidligere studier – mens hjemmeintervjuene har fungert best for verdsettingen av de mykere kvalitetsfaktorene.

Med bakgrunn i dette mener vi at egenadministrerte intervjuer er forsvarlig både faglig sett og ikke minst økonomisk. Vi anbefaler dog at det utarbeides mer brukervennlige løsninger for egenadministrerte intervjuer, f.eks. internett- eller windows-baserte løsninger, i den videre utviklingen av egenadministrerte intervjuer. Vi vil likevel igjen understreke at det er nødvendig å supplere egenadministrerte intervjuer med tradisjonelle hjemmeintervjuer for å sikre representasjon fra hele befolkningen.

## 11.10 Oppsummering og konklusjoner

Design av SP-undersøkelser kan ha potensial til å påvirke resultatet i stor grad. I rapporteringen av SP-analyser er det derfor nødvendig å dokumentere designet av undersøkelsen, med hensyn til:

- Rekrutteringsmetode og metode for datainnsamling
- SP-metoder
- Utforming av de enkelte SP-metodene
- Struktur på spørreskjema

Vi har gjennomført tester på i alt 10 ulike SP-design som omfattet 364 kollektivtrafikanter i Oslo. Følgende hovedkonklusjoner er trukket:

- Utvalgsstrategi og frafall kan påvirke resultatet. Derfor må dette dokumenteres ved enhver SP-studie.
- Pilotstudier og grundig gjennomgang av undersøkelsen med "friske øyne" er viktig. Det kan synliggjøre svakheter i designet, og skjevheter i balanseringen av nivåene i samvalgsekvenser.
- Leksikografi er et alvorlig problem i samvalgundersøkelser. Andelen leksikografi gir en god indi-

kasjon på hvor godt samvalgdesignet har fungert, og må rapporteres.

- 0-svar og ekstreme verdsettinger i CV/TP og FM er metodisk sett like problematiske som leksikografiske svar i samvalg. Disse andelene må derfor rapporteres i SP-studier. FM har vist seg å ha langt lavere andeler slike svar, og er derfor funnet å være en bedre verdsettingsmetode enn CV/TP.
- Rekkefølgen på de ulike SP-sekvensene har liten betydning for resultatene. Samvalgsekvenser bør om mulig ikke komme fortløpende i spørreskjemaet fordi det kan forårsake tretthet.
- Det er et vesentlig forbedringspotensial i å utvikle mer fleksibelt design av samvalgene, f.eks. ved å balansere nivåene i samvalgsekvensen ut fra betalingsvilje i andre typer verdsettingsspørsmål. I denne studien er CV/TP-sekvensen brukt til dette. Dette gir en klar tendens til lavere andeler leksikografi. På grunn av metodiske svakheter ved CV/TP anbefaler vi å benytte FM istedenfor CV/TP til å balansere samvalgsekvensen i fremtidige undersøkelser.
- Antallet kvalitetsfaktorer som inngår i samvalgsekvensen har ved tidligere studier vist seg å påvirke resultatet. I denne studien har imidlertid virkningen av antallet faktorer vært nøytral idet vi ikke har kunnet påvise forskjeller i resultatene mellom samvalgsekvenser med henholdsvis tre og fire kvalitetsfaktorer.
- Hjemmeintervjuer er en svært ressurskrevende metode for datainnsamling. Våre undersøkelser viser at egenadministrasjon av intervjuene, som er langt mindre kostnads-krevende, har fungert godt. En rekke tester av utvalgene med egenadministrerte intervjuer viser at datainnsamlingsmetoden har hatt svært liten innvirkning på resultatene. Imidlertid er det helt nødvendig å supplere egenadministrerte intervjuer med tradisjonelle hjemmeintervjuer blant grupper som er underrepresentert i egenadministrerte utvalg (bl.a. kvinner og de høyere aldersgruppene).
- Egenadministrerte intervjuer må gjøres mer brukervennlige enn den MS-DOS-baserte teknologien som er benyttet i denne undersøkelsen. Vi anbefaler at det utvikles Windows- og internettbaserte løsninger.
- Det anbefales økt bruk av kontrollspørsmål i alle typer SP-undersøkelser. Kontrollspørsmål kan avdekke årsaker til at respondentene tilkjenner ekstremer preferanser, og forekomsten av protestvalg og forenklete beslutningsrutiner.

# Litteratur

- Axhausen, K.W. (1998)  
Experiments with SP and CA approaches to mode choice. In *PTRC Seminar D Transportation Planning Methods Vol 1* 1998
- Bates J., T. Fowkes, P. Mackie, M. Wardman og G. Whelan (2001)  
*Three controversies in the valuation of travel time savings*. Paper presented at the European Transport Conference (PTRC), Cambridge, UK 2001
- Bradley, M. og A. Daly (1994)  
Use of logit scaling approach to test for rank-order and fatigue effects in Stated Preference data. I *Transportation* 21(2) 1994, side 167-184
- Carlquist, E. og N. Fearnley (2001)  
*Samfunnseffektiv kollektivtransport? En analyse av utviklingen i sju norske byer*. TØI-rapport 508/2001
- Fearnley, N. (2000)  
*Litteraturgjennomgang SP-design*. Oslo, Transportøkonomisk institutt. Arbeidsdokument PT/1406/2000)
- Frisch, R. (1972)  
*Cooperation between politicians and econometricians on the formalization of political preferences*. Oslo, Institute of Economics, University of Oslo. Reprint series no 90
- Gunn, H. og R. Sheldon (2001)  
The value of time. Paper presented to the Transport Economists' Group University of Westminster. In: *The Transport Economist* Vol.28 No.1, 2001
- Killi, M., H. Samstad og K. Sælensminde (2001)  
*Trafikantenes verdsetting av trafikkinformasjon*. TØI-rapport 537/2001, Transportøkonomisk institutt, Oslo
- Kottenhoff, K and Schmidt, L. (2000)  
Swedish studies of reasons and scale of the package effects in Stated Preference studies. 2000. Paper presented at the *European Transport Conference*, 11-13 September 2000, Cambridge, UK
- Laitila, T. og K. Westin (1997)  
*Utværdering av nåra analysteknikker vid Stated Preference undersökningar*. KFB-Meddelande 1997:14
- Nelson, P.S. og J.G. Towris (1995)  
Stated Preference Techniques and the Representation of Attributes. In *PTRC Transportation Planning Methods* 1995
- Nergård, Vidar (2001)  
*Vurdering av pilotstudien : En vurdering av utvalget i pilotstudien for SP-design og fordelingen av svarene gitt på de obligatoriske spørsmål*. Oslo, Transportøkonomisk institutt. Arbeidsdokument PT/1500/2001
- Norheim, B. (1996)  
*Bedre kollektivtransport : Samvalganalyse i Oslo – Metodetester og etterspørselsberegninger*. Oslo, Transportøkonomisk institutt. TØI rapport 327/1996
- Olsson, C. (1997)  
*Metodetester för mätningar av bilisters betalingsvilja för drift- och underhållsåtgärder i vägnätet*. KTH forskningsrapport TRITA-IP FR 97-24
- Ortúzar, J. de og L.G. Willumsen (1996)  
*Modelling Transport* 2<sup>nd</sup> edition. J.Wiley and Sons
- Pearmain, D. og E. Kroes (1990)  
*Stated Preference Techniques : A Guide to Practice*. Steer Davies Gleave and Hague Consulting Group
- Pearmain, D, J. Swanson, E. Kroes og M. Bradlet (1991)  
*Stated Preference Techniques : A Guide to Practise 2<sup>nd</sup> Edition*. Steer Davies Gleave og Hague Consulting Group
- Sælensminde, K. (1995a)  
*Kunnskapsoversikt SP-metoder*. Oslo, Transportøkonomisk institutt. TØI rapport 294/1995
- Sælensminde, K. (1995b)  
*Verdsetting av reisetid. Resultater fra pilotundersøkelsen nov-des 1994*. Oslo, Transportøkonomisk institutt. TØI arbeidsdokument TP/0783/1995

- Sælensminde, K. (1997)  
*Metodeproblemer ved verdsetting av ikke-markedsgoder: Resultater fra forskningsaktivitet finansiert av LOKTRA*. Oslo, Transportøkonomisk institutt. TØI arbeidsdokument TP/1108/97
- Sælensminde, K. (1998)  
Faglig rapport til Norges forskningsråd : Prosjekt: *Verdsetting av ikke-markedsgoder for bruk i nytte-kostnadsanalyser* (P.nr. 117025/730) : Rapporteringsperiode 1.8.1997 - 1.8.1998. Oslo, Transportøkonomisk institutt. TØI arbeidsdokument SM/0941/1998
- Sælensminde, K. (2000)  
*Valuation of nonmarket goods for use in cost-benefit analyses : Methodological issues*. Dissertation for the Degree of Doctor Scientiarum, Department of Economics and Social Sciences, Agricultural University of Norway 1999. Oslo, Transportøkonomisk institutt. TOI report 491/2000
- Sælensminde, K. (2001a)  
*Designforbedringer i SP-studier. Tester for å undersøke om kvaliteten på SP-data kan forbedres samtidig som kostnadene reduseres*. Oslo, Transportøkonomisk institutt. TØI Arbeidsdokument SM/1232/2001
- Sælensminde, K. (2001b)  
*Metodestudier knyttet til målet om konsistente verdsettinger av ikke-markedsgoder*. Oslo, Transportøkonomisk institutt. TØI Arbeidsdokument SM/1259/2001
- Sælensminde, K. (2001c)  
*"Frisch-metoden". En studie av hvordan en "Stated Preference"-metode foreslått av Ragnar Frisch fungerer til verdsetting av standardfaktorer i kollektivtransporten*. Oslo, Transportøkonomisk institutt. TØI Arbeidsdokument SM/1295/2001
- Sælensminde, K. (2001d)  
*Verdsetting av standardfaktorer i kollektivtransporten ved hjelp av "blokkdesign". En studie av hvordan konteksten i "Stated Choice"-studier påvirker verdsettingsresultatene*. Oslo, Transportøkonomisk institutt. TØI arbeidsdokument SM/1325/2001
- Sælensminde, K. (2001e)  
*Preferanseusikkerhet i samvalgundersøkelser. En studie av hvordan respondentene har brukt muligheten til å tilkjennegi hvor sikre de er på sine valg og hvordan dette påvirker verdsettingene*. Oslo, Transportøkonomisk institutt. TØI arbeidsdokument SM/1331/2001
- Sælensminde, K. (2001f)  
Inconsistent choices in Stated Choice data : Use of logit scaling approach to handle resulting variance increases. I *Transportation* nr. 28, 2001, side 269-296
- Sælensminde, K. og F. Hammer (1993)  
*Samvalganalyse som metode for verdsetting av miljøgoder. pilotundersøkelse*. Oslo, Transportøkonomisk institutt. Oslo, Transportøkonomisk institutt. TØI rapport nr. 184/1993
- Sjöstrand, H. (1999)  
*Värdering av kvalitet i lokal kollektivtrafik med Stated Preferences-metoden*. Lunds Tekniska Högskola Bulletin 175
- Stopher, P.R. og P.M. Jones (2001)  
*Developing Standards of Transport Survey Quality*. Keynote paper presented for the international conference on transport survey quality and innovation: How to recognize it and how to achieve it. Kruger Park, Sør-Afrika 2001
- Swanson, J. (1998)  
Factors Affecting the Validity of Stated Preferences. *PTRC*, Seminar D 1998
- Swanson, J., D. Pearmain and D. Holden (1993)  
New Approaches to the Design of Stated Preference Experiments. I: *PTRC Seminar D Transportation Planning Methods* 1993
- Wardman, M. (1998)  
The value of Time: A Review of British Evidence. I: *Journal of Transport Economics and Policy*, Vol 32 Part 3 1998 s. 285-316
- Wardman, M. (1999a)  
*Stated Preference Analysis Methods*. Stensiler. Leeds University ITS
- Wardman, M. (1999b)  
Evaluation of Railway Rolling Stock. Talk given to Transport Studies Group, university of Westminster. I: *The Transport Economist* Vol 26:3 1999



Watson, S.M., J.P. Toner og M. Wardman (1996)

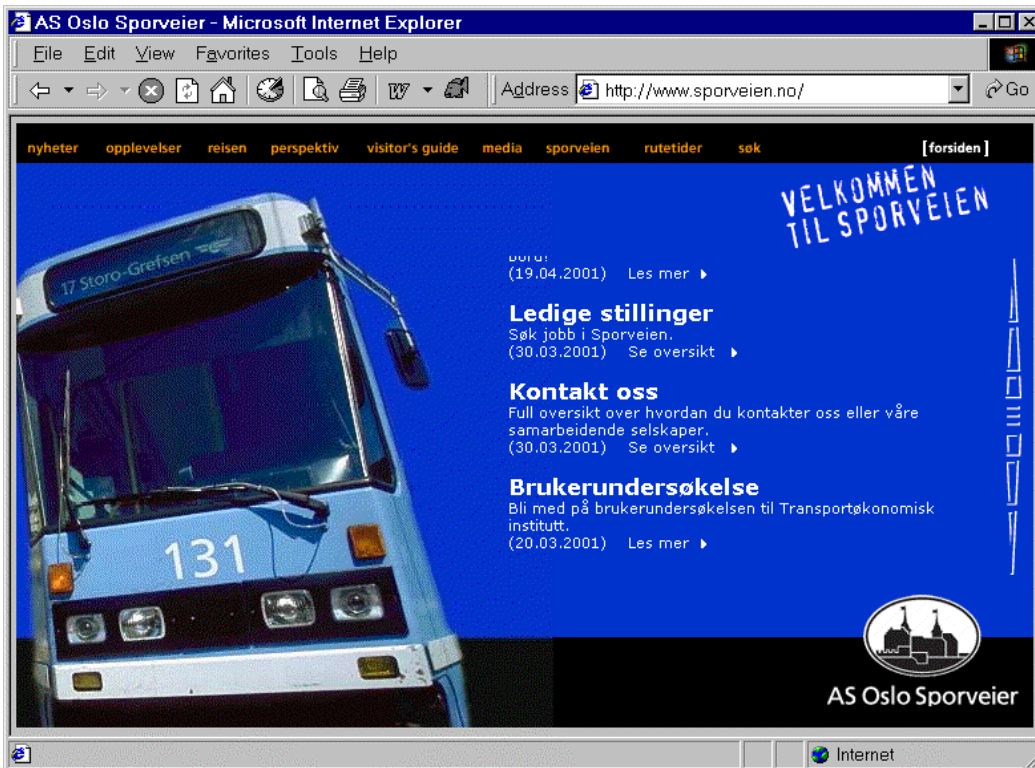
Efficiency Properties of Orthogonal Stated Preference Designs. I: *PTRC Seminar D Transportation Planning Methods* 1996

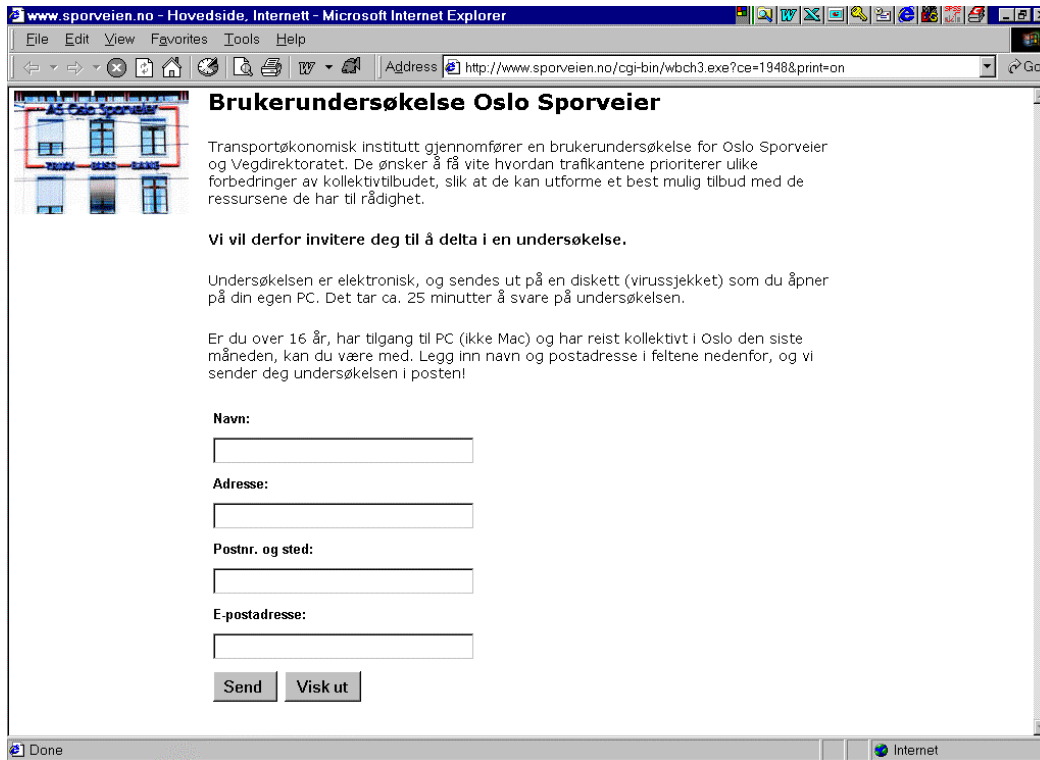
Widlert, S. (1994)

Stated Preference Studies: The Design Affects the Results. Paper submitted to the 7<sup>th</sup> *International Conference on Travel Behaviour*, Valle Nevado, Santiago, Chile 12–16 June 1994



# Vedlegg 1: Eksempler på rekruttering av respondenter via Trafikanten og Oslo Sporveiers hjemmesider





www.sporveien.no - Hovedside, Internett - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address http://www.sporveien.no/cgi-bin/wbch3.exe?ce=1948&print=on

## Brukerundersøkelse Oslo Sporveier

Transportøkonomisk institutt gjennomfører en brukerundersøkelse for Oslo Sporveier og Vegdirektoratet. De ønsker å få vite hvordan trafikantene prioriterer ulike forbedringer av kollektivtilbudet, slik at de kan utforme et best mulig tilbud med de ressursene de har til rådighet.

**Vi vil derfor invitere deg til å delta i en undersøkelse.**

Undersøkelsen er elektronisk, og sendes ut på en diskett (virusssjekk) som du åpner på din egen PC. Det tar ca. 25 minutter å svare på undersøkelsen.

Er du over 16 år, har tilgang til PC (ikke Mac) og har reist kollektivt i Oslo den siste måneden, kan du være med. Legg inn navn og postadresse i feltene nedenfor, og vi sender deg undersøkelsen i posten!

**Navn:**

**Adresse:**

**Postnr. og sted:**

**E-postadresse:**

Done Internet

## Vedlegg 2: Frafallsregistrering: Telefonrekruttering for egenadministrerte intervjuer

### FAKTA OM UNDERSØKELSEN

UNDERSØKELSE:	24159 Rekruttering
OPPDRAGSGIVER:	Transportøkonomisk institutt
DATAINNSAMLINGSMETODE:	Dataassisterte telefonintervju (CATI)
ANTALL INTERVJU:	550
ANTALL REKRUTTERTE:	429
GEOGRAFI:	Oslo
MÅLGRUPPE:	Befolkning over 18 år
INTERVJUTID:	12. - 20- mars 2001
ANSVARLIG KONSULENT:	Truls Nedregård

<b>FRAFALL</b>	<b>ANTALL</b>	<b>PROSENT</b>
ANTALL I UTVALG:	5000	
ANTALL GJENVÆRENDE UTVALG:	1341	
ANTALL GJENNOMFØRTE INTERVJU:	<b>429</b>	18
ANTALL BRUTTO FRAFALL:	3109	
ANTALL NETTO FRAFALL:	<b>1961</b>	82
<b>FORDELING AV FRAFALL</b>		
NEKT TLF.INTERVJU:	692	35
NEKT REKRUTTERING:	86	4
IKKE TILGJENGELIG I INTERVJUPERIODEN:	71	4
IKKE SVAR TELEFONSVARER:	1044	53
NUMMERET ER OPPTATT:	68	3
<b>SUM REELLE FRAFALL:</b>	<b>1961</b>	100
NUMMER IKKE I BRUK:	113	
IKKE KORREKT NUMMER:	27	
IKKE I MÅLGRUPPE (IKKE REIST):	497	
IKKE I MÅLGRUPPE (IKKE PC):	511	
IKKE I MÅLGRUPPE ELLERS:	81	
<b>SUM ANNET FRAFALL:</b>	<b>1148</b>	
<b>NEKTÅRSAKER - TLF</b>		
Andre nekte for IO	21	3
IO er syk	18	3
Sykdom/dødsfall i IOs familie	1	0
IO er bortreist/ferie	4	1
Ikke tid	188	27
For langt intervju	1	0
Deltar ikke i telefonintervju	118	17
Ikke interessert i temaet	105	15
Ikke kompetent	44	6
Ingen nektgrunn oppgitt	111	16
Annet	81	12



## Vedlegg 3: Frafallsregistrering: Telefonrekruttering for hjemmeintervjuer

### FAKTA OM UNDERSØKELSEN

UNDERSØKELSE:	24160 Rekruttering til hjemmeintervju
OPPDRAKSGIVER:	Transportøkonomisk institutt
DATAINNSAMLINGSMETODE:	Dataassisterte telefonintervju (CATI)
ANTALL INTERVJU:	66
ANTALL REKRUTTERTE:	74
GEOGRAFI:	Oslo
MÅLGRUPPE:	Befolkning over 18 år
INTERVJUTID:	12. - 16- mars 2001
ANSVARLIG KONSULENT:	Truls Nedregård

<b>FRAFALL</b>	<b>ANTALL</b>	<b>PROSENT</b>
ANTALL I UTVALG:	2000	
ANTALL GJENVÆRENDE UTVALG:	970	
ANTALL GJENNOMFØRTE INTERVJU:	78	12
ANTALL REKRUTTERTE:	74	
ANTALL BRUTTO FRAFALL:	952	
ANTALL NETTO FRAFALL:	596	88

### FORDELING AV FRAFALL

NEKT TLF.INTERVJU:	292	49
NEKT REKRUTTERING:	4	1
IKKE TILGJENGELIG I INTERVJUPERIODEN:	15	3
IKKE SVAR TELEFONSVARER:	269	45
NUMMERET ER OPPTATT:	16	3
<b>SUM REELLE FRAFALL:</b>	<b>596</b>	<b>100</b>
NUMMER IKKE I BRUK:	42	
IKKE KORREKT NUMMER:	5	
IKKE I MÅLGRUPPE (IKKE REIST):	309	
IKKE I MÅLGRUPPE ELLERS:	0	
<b>SUM ANNET FRAFALL:</b>	<b>356</b>	

### NEKTÅRSAKER - TLF

Andre neker for IO	6	2
IO er syk	5	2
Sykdom/dødsfall i IOs familie	1	0
IO er bortreist/ferie	0	0
Ikke tid	79	27
For langt intervju	58	20
Deltar ikke i telefonintervju	0	0
Ikke interessert i temaet	33	11
Ikke kompetent	26	9
Ingen nektgrunn oppgitt	28	10
Annet	56	19







## Vedlegg 4: Frafall i felt, dvs. frafall ved hjemmebesøk for hjemmeintervjuer



### FRAFALLSREGISTRERING - 24160

Hovedårsak	Antall
Respondenten avlyste intervjuet på forhånd	
Respondenten var ikke hjemme	4
Respondenten e.a. var syk e.a.	2
Respondenten nektet	
Jeg rakk ikke frem i tide	2
Jeg fant ikke frem til adressen	
Annet, noter:	
Annet, noter:	
Annet, noter:	
Annet, noter:	
Annet, noter:	
Annet, noter:	
Annet, noter:	
Annet, noter:	
Annet, noter:	
Annet, noter:	

Norsk Gallup Institutt AS – avd. Midt-Norge  
 Fortunalia 106, 7057 Jonsvatnet  
 Tlf: +47 73 90 55 11, Telefax: +47 73 90 55 12



## Vedlegg 5: Frekvenstabeller spørreskjemaet

### DESIGN

#### DESIGN

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1,00	35	9,6	9,6	9,6
2,00	31	8,5	8,5	18,1
3,00	53	14,6	14,6	32,7
4,00	52	14,3	14,3	47,0
5,00	44	12,1	12,1	59,1
6,00	55	15,1	15,1	74,2
13,00	19	5,2	5,2	79,4
14,00	28	7,7	7,7	87,1
15,00	27	7,4	7,4	94,5
16,00	20	5,5	5,5	100,0
Total	364	100,0	100,0	

### Har respondendenten førerkort?

#### FKORT

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1,00 Ja	301	82,7	82,7	82,7
2,00 Nei	63	17,3	17,3	100,0
Total	364	100,0	100,0	

### Har dere bil i husstanden?

#### BIL

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1,00 Ja	252	69,2	69,2	69,2
2,00 Nei	112	30,8	30,8	100,0
Total	364	100,0	100,0	

### Hvor mange biler i husstanden?

**ANTBIL**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	,00	4	1,1	1,6	1,6
	1,00	191	52,5	75,8	77,4
	2,00	50	13,7	19,8	97,2
	3,00	5	1,4	2,0	99,2
	4,00	2	,5	,8	100,0
	Total	252	69,2	100,0	
Missing	-1,00	112	30,8		
Total		364	100,0		

**Har du reist kollektivt de siste 7 dagene?**

**KOLLDAG**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00 Ja	326	89,6	89,6	89,6
	2,00 Nei	38	10,4	10,4	100,0
Total		364	100,0	100,0	

**Har du problemer med å huske den siste dagen du reiste kollektivt?**

**HUSK**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00 Ja	16	4,4	42,1	42,1
	2,00 Nei	22	6,0	57,9	100,0
	Total	38	10,4	100,0	
Missing	-1,00	326	89,6		
Total		364	100,0		

**Hvilken dag foretok du den siste reisen?**

**DAG**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00 Mandag-Fredag	323	88,7	88,7	88,7
	2,00 Lørdag	19	5,2	5,2	94,0
	3,00 Søndag	17	4,7	4,7	98,6
	4,00 Husker ikke	5	1,4	1,4	100,0
Total		364	100,0	100,0	

**Hva var hovedformålet med denne reisen?**

**FORMÅL**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1,00 Til/fra arbeid/skole	237	65,1	65,1	65,1
2,00 Reise i arbeidet	11	3,0	3,0	68,1
3,00 Følge/hente barn	3	,8	,8	69,0
5,00 Innkjøp dagligvarer	6	1,6	1,6	70,6
6,00 Andre innkjøp	14	3,8	3,8	74,5
7,00 Privat ærend/service	38	10,4	10,4	84,9
8,00 Egen fritidsaktivitet	32	8,8	8,8	93,7
9,00 Privat besøk	16	4,4	4,4	98,1
10,00 Annet formål (Forklar)	7	1,9	1,9	100,0
Total	364	100,0	100,0	

**Reiste du alene eller sammen med andre?****SAMMENX**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1,00 Alene	311	85,4	85,4	85,4
2,00 Sammen med andre	53	14,6	14,6	100,0
Total	364	100,0	100,0	

**Hvor mange reiste du sammen med?****SAMMEN1**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid ,00	1	,3	1,9	1,9
1,00	32	8,8	60,4	62,3
2,00	12	3,3	22,6	84,9
3,00	6	1,6	11,3	96,2
7,00	1	,3	1,9	98,1
9,00	1	,3	1,9	100,0
Total	53	14,6	100,0	
Missing -1,00	311	85,4		
Total	364	100,0		

**Hvor mange barn 0-6 år var med på reisen?**

**BARN1**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid ,00	39	10,7	73,6	73,6
1,00	11	3,0	20,8	94,3
2,00	1	,3	1,9	96,2
3,00	1	,3	1,9	98,1
9,00	1	,3	1,9	100,0
Total	53	14,6	100,0	
Missing -1,00	311	85,4		
Total	364	100,0		

**Hvordan var været?**

**NEDBØR**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1,00 Husker ikke	15	4,1	4,1	4,1
2,00 Solskinn/skyfri himmel	198	54,4	54,4	58,5
3,00 Overskyet-lett skydekke	109	29,9	29,9	88,5
4,00 Overskyet-lette regnbyger	18	4,9	4,9	93,4
5,00 Regnvær	12	3,3	3,3	96,7
6,00 Sludd	4	1,1	1,1	97,8
7,00 Snøvær	3	,8	,8	98,6
8,00 Tåke	1	,3	,3	98,9
9,00 Annet, (forklar)	4	1,1	1,1	100,0
Total	364	100,0	100,0	

**Hva var temperaturen (i celsius)?**

**CELSIUS**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1,00 Under 10 kuldegrader	19	5,2	5,2	5,2
2,00 Mellom 5 og 10 kuldegrade	89	24,5	24,5	29,7
3,00 Mellom 0 og 5 kuldegrader	132	36,3	36,3	65,9
4,00 Mellom 0 og 5 varmegrader	81	22,3	22,3	88,2
5,00 Mellom 5 og 10 varmegrade	29	8,0	8,0	96,2
6,00 Over 10 varmegrader	5	1,4	1,4	97,5
7,00 Husker ikke	9	2,5	2,5	100,0
Total	364	100,0	100,0	

**Hvordan kom du til holdeplassen?**

**TILHOLD**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1,00 Gikk	347	95,3	95,3	95,3
2,00 Syklet	2	,5	,5	95,9
3,00 Ble kjørt	8	2,2	2,2	98,1
4,00 Kjørte bil selv	3	,8	,8	98,9
5,00 Annet, (forklar)	4	1,1	1,1	100,0
Total	364	100,0	100,0	

**Hvor lang tid mellom avgangene ved startstasjon/-plass?**

**AVGANGER**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1,00	1	,3	,3	,3
2,00	2	,5	,5	,8
3,00	3	,8	,8	1,6
4,00	4	1,1	1,1	2,7
5,00	15	4,1	4,1	6,9
6,00	1	,3	,3	7,1
7,00	22	6,0	6,0	13,2
8,00	14	3,8	3,8	17,0
10,00	21	5,8	5,8	22,8
12,00	1	,3	,3	23,1
15,00	220	60,4	60,4	83,5
18,00	1	,3	,3	83,8
20,00	10	2,7	2,7	86,5
22,00	1	,3	,3	86,8
25,00	2	,5	,5	87,4
30,00	39	10,7	10,7	98,1
40,00	1	,3	,3	98,4
60,00	4	1,1	1,1	99,5
100,00	1	,3	,3	99,7
120,00	1	,3	,3	100,0
Total	364	100,0	100,0	

**Hvor mange transportmidler brukte du på reisen?**

**ANTKOLL**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1,00 1	246	67,6	67,6	67,6
2,00 2	105	28,8	28,8	96,4
3,00 3 eller flere	13	3,6	3,6	100,0
Total	364	100,0	100,0	

**Hvilke transportmidler var dette(1.)?**

**KOLLMID1**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1,00 BUSS	113	31,0	31,0	31,0
2,00 TRIKK	66	18,1	18,1	49,2
3,00 T-BANE	161	44,2	44,2	93,4
4,00 LOKALTOG	24	6,6	6,6	100,0
Total	364	100,0	100,0	

**Hvilke andre transportmiddel (2.)?**

**KOLLMID2**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1,00 BUSS	46	12,6	39,0	39,0
2,00 TRIKK	28	7,7	23,7	62,7
3,00 T-BANE	34	9,3	28,8	91,5
4,00 LOKALTOG	10	2,7	8,5	100,0
Total	118	32,4	100,0	
Missing -1,00	246	67,6		
Total	364	100,0		

**Hvilke andre transportmiddel (3.)?**

**KOLLMID3**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1,00 BUSS	7	1,9	53,8	53,8
2,00 TRIKK	3	,8	23,1	76,9
3,00 T-BANE	1	,3	7,7	84,6
4,00 LOKALTOG	2	,5	15,4	100,0
Total	13	3,6	100,0	
Missing -1,00	351	96,4		
Total	364	100,0		

**Er det flere (4.)?**

**KOLLMID4**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1,00 BUSS	1	,3	7,7	7,7
5,00 Ingen flere	12	3,3	92,3	100,0
Total	13	3,6	100,0	
Missing -1,00	351	96,4		
Total	364	100,0		

**Var T-banestasjonen hvor reisen startet, betjent eller ubetjent?**



**BETJENT**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00 Betjent	37	10,2	23,0	23,0
	2,00 Ubetjent	124	34,1	77,0	100,0
	Total	161	44,2	100,0	
Missing	-1,00	203	55,8		
Total		364	100,0		

**Hvor lenge måtte du vente på den første holdeplassen?**

**VENTE**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	,00	15	4,1	4,1	4,1
	1,00	25	6,9	6,9	11,0
	2,00	46	12,6	12,6	23,6
	3,00	50	13,7	13,7	37,4
	4,00	25	6,9	6,9	44,2
	5,00	103	28,3	28,3	72,5
	6,00	8	2,2	2,2	74,7
	7,00	15	4,1	4,1	78,8
	8,00	7	1,9	1,9	80,8
	9,00	2	,5	,5	81,3
	10,00	32	8,8	8,8	90,1
	12,00	6	1,6	1,6	91,8
	13,00	1	,3	,3	92,0
	14,00	1	,3	,3	92,3
	15,00	12	3,3	3,3	95,6
	17,00	1	,3	,3	95,9
	20,00	10	2,7	2,7	98,6
	25,00	3	,8	,8	99,5
	30,00	2	,5	,5	100,0
Total		364	100,0	100,0	

**Hvor lang var reisetiden?**

**REISETID**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid ,00	1	,3	,3	,3
1,00	1	,3	,3	,5
3,00	2	,5	,5	1,1
4,00	2	,5	,5	1,6
5,00	10	2,7	2,7	4,4
6,00	4	1,1	1,1	5,5
7,00	13	3,6	3,6	9,1
8,00	6	1,6	1,6	10,7
9,00	2	,5	,5	11,3
10,00	32	8,8	8,8	20,1
11,00	3	,8	,8	20,9
12,00	7	1,9	1,9	22,8
13,00	5	1,4	1,4	24,2
14,00	1	,3	,3	24,5
15,00	43	11,8	11,8	36,3
16,00	1	,3	,3	36,5
17,00	4	1,1	1,1	37,6
18,00	8	2,2	2,2	39,8
19,00	3	,8	,8	40,7
20,00	51	14,0	14,0	54,7
21,00	4	1,1	1,1	55,8
22,00	6	1,6	1,6	57,4
24,00	1	,3	,3	57,7
25,00	38	10,4	10,4	68,1
26,00	1	,3	,3	68,4
28,00	1	,3	,3	68,7
30,00	25	6,9	6,9	75,5
34,00	1	,3	,3	75,8
35,00	17	4,7	4,7	80,5
39,00	1	,3	,3	80,8
40,00	10	2,7	2,7	83,5
42,00	1	,3	,3	83,8
44,00	1	,3	,3	84,1
45,00	25	6,9	6,9	90,9
47,00	1	,3	,3	91,2
50,00	12	3,3	3,3	94,5
54,00	1	,3	,3	94,8
55,00	2	,5	,5	95,3
60,00	7	1,9	1,9	97,3
63,00	1	,3	,3	97,5
65,00	1	,3	,3	97,8
70,00	1	,3	,3	98,1
75,00	1	,3	,3	98,4
80,00	1	,3	,3	98,6
85,00	1	,3	,3	98,9
90,00	2	,5	,5	99,5
120,00	2	,5	,5	100,0
Total	364	100,0	100,0	

**Fikk du sitteplass?**

**SITTEPL**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1,00 sitteplass hele veien	253	69,5	69,5	69,5
2,00 sitteplass deler av veien	65	17,9	17,9	87,4
3,00 ståplass hele veien	46	12,6	12,6	100,0
Total	364	100,0	100,0	

## Reiste du på honnørbillett?

### HONNØR

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1,00 Ja	38	10,4	10,4	10,4
2,00 Nei	326	89,6	89,6	100,0
Total	364	100,0	100,0	

## Hva slags billettype brukte du?

### BILLETT

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1,00 Enkeltbillett	45	12,4	12,4	12,4
2,00 Flexikort	127	34,9	35,0	47,4
3,00 Dagskort	13	3,6	3,6	51,0
4,00 7-dagerskort	20	5,5	5,5	56,5
5,00 Månedskort	83	22,8	22,9	79,3
6,00 Tremånederskort	46	12,6	12,7	92,0
7,00 Ungdomskort (Måned)	4	1,1	1,1	93,1
8,00 Ungdomskort (Tremåneder)	2	,5	,6	93,7
9,00 Annet, (forklar)	23	6,3	6,3	100,0
Total	363	99,7	100,0	
Missing -1,00	1	,3		
Total	364	100,0		

## Hadde du bil som du kunne ha brukt på denne reisen istedenfor kollektivtrafikk?

### BILTIL

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1,00 Ja	137	37,6	60,9	60,9
2,00 Nei	88	24,2	39,1	100,0
Total	225	61,8	100,0	
Missing -1,00	139	38,2		
Total	364	100,0		

## Hvorfor brukte du ikke bilen?

**HVORFOR**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00 Dyrere å kjøre bil	9	2,5	6,6	6,6
	3,00 Bilkøer	15	4,1	11,0	17,6
	4,00 Raskere å reise kollektiv	17	4,7	12,5	30,1
	5,00 Problemer/dyrt med parker	55	15,1	40,4	70,6
	6,00 Vinterføre/glatte veier	1	,3	,7	71,3
	7,00 Miljøhensyn	15	4,1	11,0	82,4
	8,00 Annet:	24	6,6	17,6	100,0
	Total	136	37,4	100,0	
Missing	-1,00	228	62,6		
Total		364	100,0		

**Er det flere grunner?**

**FLERE**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00 Dyrere å kjøre bil	11	3,0	8,0	8,0
	2,00 Problemer med bilen	1	,3	,7	8,8
	3,00 Bilkøer	19	5,2	13,9	22,6
	4,00 Raskere å reise kollektiv	16	4,4	11,7	34,3
	5,00 Problemer/dyrt med parker	20	5,5	14,6	48,9
	6,00 Vinterføre/glatte veier	2	,5	1,5	50,4
	7,00 Miljøhensyn	22	6,0	16,1	66,4
	8,00 Annet:	14	3,8	10,2	76,6
	9,00 Ingen flere	32	8,8	23,4	100,0
	Total	137	37,6	100,0	
Missing	-1,00	227	62,4		
Total		364	100,0		

**Er det enda flere grunner?**

**FLERE2**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00 Dyrere å kjøre bil	7	1,9	5,7	5,7
	3,00 Bilkøer	9	2,5	7,3	13,0
	4,00 Raskere å reise kollektiv	8	2,2	6,5	19,5
	5,00 Problemer/dyrt med parker	9	2,5	7,3	26,8
	6,00 Vinterføre/glatte veier	2	,5	1,6	28,5
	7,00 Miljøhensyn	15	4,1	12,2	40,7
	8,00 Annet:	7	1,9	5,7	46,3
	9,00 Ingen flere	66	18,1	53,7	100,0
	Total	123	33,8	100,0	
Missing	-1,00	241	66,2		
Total		364	100,0		

**Hvor gammel er du?**

ALDER

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 16,00	6	1,6	1,6	1,6
17,00	3	,8	,8	2,5
18,00	10	2,7	2,7	5,2
19,00	2	,5	,5	5,8
20,00	8	2,2	2,2	8,0
21,00	6	1,6	1,6	9,6
22,00	10	2,7	2,7	12,4
23,00	4	1,1	1,1	13,5
24,00	11	3,0	3,0	16,5
25,00	13	3,6	3,6	20,1
26,00	8	2,2	2,2	22,3
27,00	12	3,3	3,3	25,5
28,00	8	2,2	2,2	27,7
29,00	18	4,9	4,9	32,7
30,00	9	2,5	2,5	35,2
31,00	17	4,7	4,7	39,8
32,00	13	3,6	3,6	43,4
33,00	12	3,3	3,3	46,7
34,00	10	2,7	2,7	49,5
35,00	8	2,2	2,2	51,6
36,00	7	1,9	1,9	53,6
37,00	5	1,4	1,4	54,9
38,00	6	1,6	1,6	56,6
39,00	7	1,9	1,9	58,5
40,00	6	1,6	1,6	60,2
41,00	8	2,2	2,2	62,4
42,00	8	2,2	2,2	64,6
43,00	10	2,7	2,7	67,3
44,00	8	2,2	2,2	69,5
45,00	7	1,9	1,9	71,4
46,00	9	2,5	2,5	73,9
47,00	5	1,4	1,4	75,3
48,00	3	,8	,8	76,1
49,00	8	2,2	2,2	78,3
50,00	6	1,6	1,6	79,9
51,00	3	,8	,8	80,8
52,00	6	1,6	1,6	82,4
53,00	8	2,2	2,2	84,6
54,00	4	1,1	1,1	85,7
55,00	5	1,4	1,4	87,1
56,00	4	1,1	1,1	88,2
57,00	7	1,9	1,9	90,1
58,00	1	,3	,3	90,4
59,00	3	,8	,8	91,2
60,00	3	,8	,8	92,0
61,00	4	1,1	1,1	93,1
62,00	1	,3	,3	93,4
63,00	2	,5	,5	94,0
64,00	1	,3	,3	94,2
65,00	1	,3	,3	94,5
67,00	2	,5	,5	95,1
69,00	2	,5	,5	95,6
70,00	1	,3	,3	95,9
72,00	3	,8	,8	96,7
73,00	2	,5	,5	97,3
74,00	1	,3	,3	97,5
77,00	4	1,1	1,1	98,6
78,00	1	,3	,3	98,9
80,00	1	,3	,3	99,2
81,00	1	,3	,3	99,5
85,00	1	,3	,3	99,7
87,00	1	,3	,3	100,0
Total	364	100,0	100,0	

**Hvilken sivilstand?**

**SIVILST**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00 Ugift	113	31,0	31,0	31,0
	2,00 Gift/samboende	222	61,0	61,0	92,0
	3,00 Skilt/separeert	19	5,2	5,2	97,3
	4,00 Enke/enkemann	10	2,7	2,7	100,0
	Total	364	100,0	100,0	

**Hvor mange voksne og barn i husholdningen inkl. deg selv?**

**HUSHTYP**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00	84	23,1	23,1	23,1
	2,00	141	38,7	38,7	61,8
	3,00	64	17,6	17,6	79,4
	4,00	55	15,1	15,1	94,5
	5,00	14	3,8	3,8	98,4
	6,00	4	1,1	1,1	99,5
	7,00	1	,3	,3	99,7
	8,00	1	,3	,3	100,0
	Total	364	100,0	100,0	

**Hvor mange barn i husholdningen, 0-6 år?**

**BARN06**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	,00	212	58,2	75,7	75,7
	1,00	50	13,7	17,9	93,6
	2,00	14	3,8	5,0	98,6
	3,00	3	,8	1,1	99,6
	9,00	1	,3	,4	100,0
	Total	280	76,9	100,0	
Missing	-1,00	84	23,1		
	Total	364	100,0		

**Hvor mange barn i husholdningen, 7-12 år?**

**BARN12**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	,00	222	61,0	79,3	79,3
	1,00	37	10,2	13,2	92,5
	2,00	21	5,8	7,5	100,0
	Total	280	76,9	100,0	
Missing	-1,00	84	23,1		
	Total	364	100,0		

**Hvor mange barn i husholdningen, 13-15 år?**

**BARN16**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	,00	255	70,1	91,1	91,1
	1,00	23	6,3	8,2	99,3
	2,00	2	,5	,7	100,0
	Total	280	76,9	100,0	
Missing	-1,00	84	23,1		
Total		364	100,0		

**Hvor mange voksne i husholdningen, 16 år og over?**

**VOKSNE1**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	,00	55	15,1	19,6	19,6
	1,00	22	6,0	7,9	27,5
	2,00	160	44,0	57,1	84,6
	3,00	27	7,4	9,6	94,3
	4,00	15	4,1	5,4	99,6
	5,00	1	,3	,4	100,0
	Total	280	76,9	100,0	
Missing	-1,00	84	23,1		
Total		364	100,0		

**Hvilket yrke (hovedgeskjeft)?**

**HOVEDGJ**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00 Yrkesaktiv	260	71,4	71,4	71,4
	2,00 Elev, student	65	17,9	17,9	89,3
	3,00 Hjemmearbeidende	9	2,5	2,5	91,8
	4,00 Arbeidsledig, på tiltak	2	,5	,5	92,3
	5,00 Pensjonist, trygdet	23	6,3	6,3	98,6
	6,00 Annet....	5	1,4	1,4	100,0
	Total	364	100,0	100,0	

**Er du yrkesaktiv på heltid eller deltid?**

**YRKESAKT**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00 Heltid,( > 35 t/uke)	248	68,1	95,4	95,4
	2,00 Deltid,( < 35 t/uke)	12	3,3	4,6	100,0
	Total	260	71,4	100,0	
Missing	-1,00	104	28,6		
Total		364	100,0		

**Hva slags arbeidstidsordning?**



**ARBTID**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00 Fast arbeidstid	118	32,4	45,4	45,4
	2,00 Fleksibel arbeidstid	124	34,1	47,7	93,1
	3,00 Skift, turnusordning, nat	17	4,7	6,5	99,6
	4,00 Annet.....	1	,3	,4	100,0
	Total	260	71,4	100,0	
Missing	-1,00	104	28,6		
Total		364	100,0		

**Hvilket kjønn?**

**KJØNN**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00 Mann	184	50,5	50,5	50,5
	2,00 Kvinne	180	49,5	49,5	100,0
	Total	364	100,0	100,0	



# Vedlegg 6: Beregninger av samvalgene

## 6.A: Samvalgsekvens 1. Hovedmodell for alle respondentene

```
.mt0.17"
.mb1.00"
.op
.po8
.cw10
```

```
Hague Consulting Group
ALOGIT Version 3F/2 (904)
```

```
Page 1
15:10:50 on 6 Nov 01
```

```
Alle-1ny.PIN: Game 1 ALOGIT setup Date:2506 Time:1058
```

```
Last input data item in transformations or utilities 22
```

```
Number of temporary stores used 12
```

```
60 transformation codes; maximum 5000
```

```
Utility transformations
```

```
Temp001 = IFEQ( Data0006 ,1 )
Temp002 = IFEQ( Data0006 ,2 )
Temp003 = IFEQ( Data0008 ,1 )
Temp004 = IFEQ( Data0008 ,2 )
Temp005 = IFEQ( Data0010 ,1 )
Temp006 = IFEQ( Data0010 ,2 )
Temp007 = IFEQ( Data0015 ,1 )
Temp008 = IFEQ( Data0015 ,2 )
Temp009 = IFEQ( Data0017 ,1 )
Temp010 = IFEQ( Data0017 ,2 )
Temp011 = IFEQ( Data0019 ,1 )
Temp012 = IFEQ( Data0019 ,2 )
```

```
INFORMATION: DATA input set by user from 22 to 198
```

```
Maximum Iterations 10
```

```
Convergence criterion is .10E-01 Option 3
```

```
INFORMATION: No explicit specification - base file read with default format
```

```
OBSERVATION REJECTED: 10030004 choice 0 not defined
OBSERVATION REJECTED: 10030007 choice 0 not defined
OBSERVATION REJECTED: 10040003 choice 0 not defined
OBSERVATION REJECTED: 10110003 choice 0 not defined
OBSERVATION REJECTED: 10110005 choice 0 not defined
OBSERVATION REJECTED: 10110009 choice 0 not defined
OBSERVATION REJECTED: 10140006 choice 0 not defined
OBSERVATION REJECTED: 10140007 choice 0 not defined
OBSERVATION REJECTED: 10140009 choice 0 not defined
OBSERVATION REJECTED: 10170003 choice 0 not defined
OBSERVATION REJECTED: 10180006 choice 0 not defined
OBSERVATION REJECTED: 10180008 choice 0 not defined
OBSERVATION REJECTED: 10260005 choice 0 not defined
```

OBSERVATION REJECTED: 10260008 choice 0 not defined  
 .pa

Hague Consulting Group Page 2  
 ALOGIT Version 3F/2 (904) 15:10:50 on 6 Nov 01  
 Alle-1ny.PIN: Game 1 ALOGIT setup Date:2506 Time:1058

OBSERVATION REJECTED: 10360001 choice 0 not defined  
 OBSERVATION REJECTED: 10360002 choice 0 not defined  
 OBSERVATION REJECTED: 10360003 choice 0 not defined  
 OBSERVATION REJECTED: 10360004 choice 0 not defined  
 OBSERVATION REJECTED: 10360005 choice 0 not defined  
 OBSERVATION REJECTED: 10360006 choice 0 not defined  
 OBSERVATION REJECTED: 10360007 choice 0 not defined  
 OBSERVATION REJECTED: 10360008 choice 0 not defined  
 OBSERVATION REJECTED: 10360009 choice 0 not defined  
 OBSERVATION REJECTED: 10370002 choice 0 not defined  
 OBSERVATION REJECTED: 10370003 choice 0 not defined  
 Further diagnostics omitted after 25 rejections

Report of user selections  
 0 Observations rejected because item 2012 = 1.00  
 44 Observations rejected because choice improper

DATA INPUT COMPLETED  
 from data file : C:\2584\ALLE\SPILL1\DATA1.DAT  
 Total observations read from file : 3669  
 Observations rejected by user tests : 0  
 Observations rejected automatically : 44  
 Observations accepted for processing : 3625  
 Sum of weights of observations : 3625.00

SPECIFICATION OF MODEL and DATA STATISTICS

Alternative 1: chosen 1856.0 of available 3625.0 observations

Coefficient Vektore +N-telefon + Overbygg + Leskur + Realtid + Rutekart + Pris + Konstant

Number (Con)	1 (F)	2 (F)	3 (F)	4 (F)	5 (F)	6 (F)	7 (F)	8 (F)
Start Value	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
Data Item	*Temp001	*Temp002	*Temp003	*Temp004	*Temp005	*Temp006	*Data0013	*Const.=1
% Non-Zero	23.8	37.6	28.3	21.5	33.4	28.1	100.0	100.0
Mean (N-Z)	1.00	1.00	1.00	1.00	1.00	1.00	14.50	1.00
C. of V. %	.0	.0	.0	.0	.0	.0	42.9	.0

.pa

Hague Consulting Group Page 3  
 ALOGIT Version 3F/2 (904) 15:10:50 on 6 Nov 01  
 Alle-1ny.PIN: Game 1 ALOGIT setup Date:2506 Time:1058

Alternative 2: chosen 1769.0 of available 3625.0 observations

Coefficient Number (Con)	Vektore 1 (F)	+N-telefon 2 (F)	+ Overbygg 3 (F)	+ Leskur 4 (F)	+ Realtid 5 (F)	+ Rutekart 6 (F)	+ Pris 7 (F)
Start Value	.0000	.0000	.0000	.0000	.0000	.0000	.0000

Data Item	*Temp007	*Temp008	*Temp009	*Temp010	*Temp011	*Temp012	*Data0022
% Non-Zero	23.6	37.2	28.4	19.7	34.2	27.7	100.0
Mean (N-Z)	1.00	1.00	1.00	1.00	1.00	1.00	14.46
C. of V. %	.0	.0	.0	.0	.0	.0	41.4

RANGES OF INDEPENDENT VARIABLES

Variable	Vektore	N-telefon	Overbygg	Leskur	Realtid	Rutekart	Pris	Konstant
Chsn Min	.00	.00	.00	.00	.00	.00	1.00	.00
Max	1.00	1.00	1.00	1.00	1.00	1.00	99.00	1.00
Diff Min	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-82.00	-1.00
Max	1.00	1.00	1.00	1.00	1.00	1.00	87.00	1.00

MEANS OF INDEPENDENT VARIABLES

Variable	Vektore	N-telefon	Overbygg	Leskur	Realtid	Rutekart	Pris	Konstant
Chosen	.20	.38	.23	.22	.36	.22	13.39	.51
Difference	.07	-.02	.10	-.04	-.05	.11	2.18	-.02

STANDARD DEVIATIONS OF INDEPENDENT VARIABLES

Variable	Vektore	N-telefon	Overbygg	Leskur	Realtid	Rutekart	Pris	Konstant
Chosen	.40	.49	.42	.42	.48	.42	5.53	.50
Difference	.63	.73	.64	.59	.73	.67	5.79	1.00

CORRELATION MATRIX of DIFFERENCES from CHOSEN VALUE

Variable	Vektore	N-telefon	Overbygg	Leskur	Realtid	Rutekart	Pris
N-telefon	-.4083						
Overbygg	-.1865	-.0064					
Leskur	-.0343	.2067	-.4117				
Realtid	-.3004	.0510	-.3680	.2117			
Rutekart	-.0574	.0305	.1059	-.1448	-.4479		
Pris	.1067	.0258	-.0012	.1772	.1004	.0504	
Konstant	.0064	.0050	.0029	.0290	-.0116	.0098	.0160

Data preparation completed

.pa

Linear ("Quick") algorithm being used

.pa

Hague Consulting Group

Page 5

ALOGIT Version 3F/2 (904)

15:10:50 on 6 Nov 01

Alle-1ny.PIN: Game 1 ALOGIT setup

Date:2506 Time:1058

Convergence achieved after 5 iterations

Analysis is based on 3625 observations

Likelihood with Zero Coefficients = -2512.6585

Likelihood with Constants only = -2511.6144

Initial Likelihood = -2512.6585

Final value of Likelihood = -2039.1884

"Rho-Squared" w.r.t. Zero = .1884

"Rho-Squared" w.r.t. Constants = .1881

ESTIMATES OBTAINED AT ITERATION 5

Likelihood = -2039.1884

	Vektore	N-telefon	Overbygg	Leskur	Realtid	Rutekart	Pris	Konstant
Estimate	.7145E-01	.9082E-01	-.1845	.4002	.2082	-.1866	-.2808	.6394E-01
Std. Error	.837E-01	.591E-01	.787E-01	.753E-01	.763E-01	.693E-01	.132E-01	.380E-01
"T" Ratio	.9	1.5	-2.3	5.3	2.7	-2.7	-21.3	1.7

Correlation of Estimates (multiplied by 1000)

	1	2	3	4	5	6	7	
N-telefon	2	456						
Overbygg	3	437	106					
Leskur	4	172	-108	427				
Realtid	5	568	207	519	134			
Rutekart	6	328	48	302	220	547		
Pris	7	-440	-189	-328	-308	-427	-299	
Konstant	8	1	3	-11	-30	17	-3	-22

Convergence (option 3) value is .1487E-02

Normal finish after 0 mins. 02.5 secs.

6.B: Samvalgsekvens 2. Hovedmodell for alle respondentene

.mt0.17"  
 .mb1.00"  
 .op  
 .po8  
 .cw10

Hague Consulting Group Page 1  
 ALOGIT Version 3F/2 (904) 10:08:12 on 7 Nov 01  
 Alle-2.PIN: Game 2 ALOGIT setup Date:2506 Time:1058

Last input data item in transformations or utilities 22

Number of temporary stores used 4  
 50 transformation codes; maximum 5000

Utility transformations  
 Temp001 = IFEQ( Data0006 ,2 )  
 Temp001 = 0.5 \*Temp001  
 Temp001 = Temp001 \*Data0009  
 Temp002 = IFEQ( Data0006 ,3 )  
 Temp002 = Temp002 \*Data0009  
 Temp003 = IFEQ( Data0015 ,2 )  
 Temp003 = 0.5 \*Temp003  
 Temp003 = Temp003 \*Data0018  
 Temp004 = IFEQ( Data0015 ,3 )  
 Temp004 = Temp004 \*Data0018

INFORMATION: DATA input set by user from 22 to 198

Maximum Iterations 10  
 Convergence criterion is .10E-01 Option 3

INFORMATION: No explicit specification - base file read with default format

OBSERVATION REJECTED: 10020006 choice 0 not defined  
 OBSERVATION REJECTED: 10030007 choice 0 not defined  
 OBSERVATION REJECTED: 10040009 choice 0 not defined  
 OBSERVATION REJECTED: 10050001 choice 0 not defined  
 OBSERVATION REJECTED: 10110004 choice 0 not defined  
 OBSERVATION REJECTED: 10130003 choice 0 not defined  
 OBSERVATION REJECTED: 10180002 choice 0 not defined  
 OBSERVATION REJECTED: 10180009 choice 0 not defined  
 OBSERVATION REJECTED: 10220002 choice 0 not defined  
 OBSERVATION REJECTED: 10250003 choice 0 not defined  
 OBSERVATION REJECTED: 10250004 choice 0 not defined  
 OBSERVATION REJECTED: 10260002 choice 0 not defined  
 OBSERVATION REJECTED: 10280005 choice 0 not defined  
 OBSERVATION REJECTED: 10330001 choice 0 not defined  
 OBSERVATION REJECTED: 10330002 choice 0 not defined  
 .pa

OBSERVATION REJECTED: 10330005 choice 0 not defined  
 OBSERVATION REJECTED: 10330006 choice 0 not defined  
 OBSERVATION REJECTED: 10330009 choice 0 not defined  
 OBSERVATION REJECTED: 10360001 choice 0 not defined  
 OBSERVATION REJECTED: 10360002 choice 0 not defined  
 OBSERVATION REJECTED: 10360003 choice 0 not defined  
 OBSERVATION REJECTED: 10360004 choice 0 not defined  
 OBSERVATION REJECTED: 10360005 choice 0 not defined  
 OBSERVATION REJECTED: 10360006 choice 0 not defined  
 OBSERVATION REJECTED: 10360007 choice 0 not defined

Further diagnostics omitted after 25 rejections

Report of user selections

0 Observations rejected because item 2012 = 1.00  
 57 Observations rejected because choice improper

DATA INPUT COMPLETED

from data file : C:\2584\ALLE\SPILL2\DATA2.DAT  
 Total observations read from file : 3683  
 Observations rejected by user tests : 0  
 Observations rejected automatically : 57  
 Observations accepted for processing : 3626  
 Sum of weights of observations : 3626.00

SPECIFICATION OF MODEL and DATA STATISTICS

Alternative 1: chosen 1796.0 of available 3626.0 observations

Coefficient	TidStaplas	+TidStaplas	+ Reisetid	+ Frekvens	+ Pris	+ Konstant
Number (Con)	1 (F)	1 (F)	3 (F)	4 (F)	5 (F)	6 (F)
Start Value	.0000	.0000	.0000	.0000	.0000	.0000
Data Item	*Temp001	*Temp002	*Data0009	*Data0011	*Data0013	*Const.=1
% Non-Zero	30.4	32.2	99.8	100.0	100.0	100.0
Mean (N-Z)	12.55	25.23	25.74	17.36	14.55	1.00
C. of V. %	72.8	67.6	72.7	70.8	45.1	.0

.pa

Alternative 2: chosen 1830.0 of available 3626.0 observations

Coefficient	TidStaplas	+TidStaplas	+ Reisetid	+ Frekvens	+ Pris
Number (Con)	1 (F)	1 (F)	3 (F)	4 (F)	5 (F)
Start Value	.0000	.0000	.0000	.0000	.0000
Data Item	*Temp003	*Temp004	*Data0018	*Data0020	*Data0022
% Non-Zero	31.7	30.4	99.8	100.0	100.0
Mean (N-Z)	12.60	25.66	25.92	17.20	14.53
C. of V. %	69.9	69.5	71.8	72.5	44.6

RANGES OF INDEPENDENT VARIABLES



Variable	TidStaplas	Reisetid	Frekvens	Pris	Konstant
Chsn Min	.00	.00	1.00	1.00	.00
Max	150.00	150.00	130.00	99.00	1.00
Diff Min	-150.00	-60.00	-70.00	-54.00	-1.00
Max	85.00	60.00	70.00	89.00	1.00

MEANS OF INDEPENDENT VARIABLES

Variable	TidStaplas	Reisetid	Frekvens	Pris	Konstant
Chosen	11.85	25.64	16.10	13.56	.50
Difference	.03	.25	2.37	1.96	.01

STANDARD DEVIATIONS OF INDEPENDENT VARIABLES

Variable	TidStaplas	Reisetid	Frekvens	Pris	Konstant
Chosen	14.52	17.83	11.90	6.08	.50
Difference	16.94	9.71	11.96	7.42	1.00

CORRELATION MATRIX of DIFFERENCES from CHOSEN VALUE

Variable	TidStaplas	Reisetid	Frekvens	Pris
Reisetid	.0618			
Frekvens	-.3995	-.3012		
Pris	-.1461	-.2029	-.2122	
Konstant	.0095	-.0191	.0115	.0010

Data preparation completed

Linear ("Quick") algorithm being used

.pa

Hague Consulting Group

Page 4

ALOGIT Version 3F/2 (904)

10:08:12 on 7 Nov 01

Alle-2.PIN: Game 2 ALOGIT setup

Date:2506 Time:1058

Convergence achieved after 5 iterations

Analysis is based on 3626 observations

Likelihood with Zero Coefficients = -2513.3517

Likelihood with Constants only = -2513.1923

Initial Likelihood = -2513.3517

Final value of Likelihood = -2111.9598

"Rho-Squared" w.r.t. Zero = .1597

"Rho-Squared" w.r.t. Constants = .1597

ESTIMATES OBTAINED AT ITERATION 5

Likelihood = -2111.9598

	TidStaplas	Reisetid	Frekvens	Pris	Konstant
Estimate	-.3202E-01	-.5939E-01	-.8714E-01	-.1531	-.1363E-01
Std. Error	.276E-02	.467E-02	.434E-02	.723E-02	.372E-01
"T" Ratio	-11.6	-12.7	-20.1	-21.2	-.4

Correlation of Estimates (multiplied by 1000)

	1	3	4	5	
Reisetid	3	189			
Frekvens	4	501	447		
Pris	5	368	407	488	
Konstant	6	-19	16	-13	1

Convergence (option 3) value is .1569E-04  
 Normal finish after 0 mins. 02.4 secs.

### 6.C Modell med skalering

.mt0.17"  
 .mb1.00"  
 .op  
 .po8  
 .cw10

Hague Consulting Group  
 ALOGIT Version 3F/2 (904)

Page 1  
 11:30:12 on 28 Nov 01  
 Date:2506 Time:1058

DES3-2.PIN: Game 2 ALOGIT setup

Workfile created at 11:29:58 on 28 Nov 01 from  
 Base data file: H:\2584\PILOT\SCALING\DATA2.DAT

#### COEFFICIENTS DEFINED

No.	Coefficient	C	Init. Value
1	SittHalv	F	.507681E+00
2	SittHel	F	.991875E+00
3	Reisetid	F	-.703839E-01
4	Frekvens	F	-.943513E-01
5	Pris	F	-.150380E+00
6	Konstant	F	-.895803E-02

Restart after 4 previous iterations at 11:29:58 on 28 Nov 01

Dummy alternative 19 introduced into structure  
 Dummy alternative 20 introduced into structure  
 Dummy alternative 21 introduced into structure  
 Dummy alternative 22 introduced into structure  
 Dummy alternative 23 introduced into structure  
 Dummy alternative 24 introduced into structure  
 Dummy alternative 25 introduced into structure  
 Dummy alternative 26 introduced into structure  
 Dummy alternative 27 introduced into structure  
 Dummy alternative 28 introduced into structure  
 Dummy alternative 29 introduced into structure  
 Dummy alternative 30 introduced into structure  
 Dummy alternative 31 introduced into structure  
 Dummy alternative 32 introduced into structure  
 Dummy alternative 33 introduced into structure  
 Dummy alternative 34 introduced into structure  
 Dummy alternative 35 introduced into structure  
 Dummy alternative 36 introduced into structure  
 Root alternative 37 has been defined

.pa

Hague Consulting Group
ALOGIT Version 3F/2 (904)

Page 2
11:30:12 on 28 Nov 01

DES3-2.PIN: Game 2 ALOGIT setup Date:2506 Time:1058

Tree from root to level 3

- 10000000 19000000
20000000 20000000
30000000 21000000
40000000 22000000
50000000 23000000
60000000 24000000
70000000 25000000
80000000 26000000
90000000 27000000
100000000 28000000
110000000 29000000
120000000 30000000
130000000 31000000
140000000 32000000
150000000 33000000
160000000 34000000
170000000 35000000
180000000 360000000 37

Maximum Iterations 10

Convergence criterion is .10E-01 Option 3

Non-linear algorithm being used

Maximum tree levels possible = 60; needed = 3

.pa

Hague Consulting Group
ALOGIT Version 3F/2 (904)

Page 3
11:30:12 on 28 Nov 01

DES3-2.PIN: Game 2 ALOGIT setup Date:2506 Time:1058

Convergence achieved after 9 iterations

Analysis is based on 1310 observations

Likelihood with Zero Coefficients = -908.0228

Likelihood with Constants only = -901.7665

Initial Likelihood = -743.2549

Final value of Likelihood = -736.9511

"Rho-Squared" w.r.t. Zero = .1884

"Rho-Squared" w.r.t. Constants = .1828

ESTIMATES OBTAINED AT ITERATION 9

Likelihood = -736.9511

SittHalv SittHel Reisetid Frekvens Pris Konstant s1 s2 s3 s4

Tester av *Stated Preference*-teknikker og samvalgdesign

Estimate	.7827	1.540	-.1101	-.1360	-.2071	-.3462E-01	1.000	.7387	.6400	.7450
Std. Error	.229	.412	.294E-01	.300E-01	.424E-01	.884E-01	.000	.233	.20	.206
"T" Ratio	3.4	3.7	-3.8	-4.5	-4.9	-.4	.0	3.2	3.2	3.6

	s5	s6	s7	s8	s9
Estimate	1.128	.6076	.5959	.6988	.4903
Std. Error	.307	.175	.181	.200	.174
"T" Ratio	3.7	3.5	3.3	3.5	2.8

.pa

Hague Consulting Group

Page 4

ALOGIT Version 3F/2 (904)

11:30:12 on 28 Nov 01

DES3-2.PIN: Game 2 ALOGIT setup

Date:2506 Time:1058

Correlation of Estimates (multiplied by 1000)

		1	2	3	4	5	6	81	82	83	84	85	86	87	88
SittHel	2	813													
Reisetid	3	-778	-849												
Frekvens	4	-832	-876	939											
Pris	5	-827	-833	913	923										
Konstant	6	-160	-145	171	157	160									
s1	81	9999	9999	9999	9999	9999	9999								
s2	82	-615	-552	573	586	668	143	9999							
s3	83	-670	-590	712	713	721	229	9999	487						
s4	84	-639	-559	696	722	737	70	9999	485	568					
s5	85	-695	-664	776	785	790	151	9999	519	611	616				
s6	86	-621	-617	706	720	727	156	9999	475	553	558	600			
s7	87	-659	-680	712	732	715	182	9999	472	548	536	593	544		
s8	88	-669	-666	713	738	736	93	9999	487	553	561	606	553	557	
s9	89	-589	-618	689	682	649	107	9999	415	506	500	552	502	512	515

Convergence (option 3) value is .2415E-03

Normal finish after 0 mins. 03.5 secs.

## Vedlegg 7: Oversikt over leksikografiske svar

### 7.A.Samvalgsekvens 1 fordelt på ulike design

DESIGN			Frequency	Percent	Valid Percent	Cumulative Percent
1	Valid	Ikke leksikografisk	21	60,0	60,0	60,0
		Pris	14	40,0	40,0	100,0
		Total	35	100,0	100,0	
2	Valid	Ikke leksikografisk	17	54,8	54,8	54,8
		Trygghet	1	3,2	3,2	58,1
		Leskur	1	3,2	3,2	61,3
		Informasjon på holdeplassen	1	3,2	3,2	64,5
		Pris	11	35,5	35,5	100,0
		Total	31	100,0	100,0	
3	Valid	Ikke leksikografisk	27	50,9	50,9	50,9
		Informasjon på holdeplassen	1	1,9	1,9	52,8
		Pris	25	47,2	47,2	100,0
		Total	53	100,0	100,0	
4	Valid	Ikke leksikografisk	24	46,2	46,2	46,2
		Informasjon på holdeplassen	1	1,9	1,9	48,1
		Pris	27	51,9	51,9	100,0
		Total	52	100,0	100,0	
5	Valid	Ikke leksikografisk	25	56,8	56,8	56,8
		Pris	19	43,2	43,2	100,0
		Total	44	100,0	100,0	
6	Valid	Ikke leksikografisk	21	38,2	50,0	50,0
		Trygghet	1	1,8	2,4	52,4
		Pris	20	36,4	47,6	100,0
		Total	42	76,4	100,0	
		Missing	-1	13	23,6	
	Total	55	100,0			
13	Valid	Ikke leksikografisk	12	63,2	63,2	63,2
		Trygghet	1	5,3	5,3	68,4
		Pris	6	31,6	31,6	100,0
		Total	19	100,0	100,0	
14	Valid	Ikke leksikografisk	13	46,4	48,1	48,1
		Trygghet	1	3,6	3,7	51,9
		Informasjon på holdeplassen	1	3,6	3,7	55,6
		Pris	12	42,9	44,4	100,0
		Total	27	96,4	100,0	
	Missing	-1	1	3,6		
Total	28	100,0				
15	Valid	Ikke leksikografisk	20	74,1	74,1	74,1
		Pris	7	25,9	25,9	100,0
		Total	27	100,0	100,0	
16	Valid	Ikke leksikografisk	11	55,0	73,3	73,3
		Pris	4	20,0	26,7	100,0
		Total	15	75,0	100,0	
	Missing	-1	5	25,0		
	Total	20	100,0			

### 7.B Samvalgsekvens 2 fordelt på ulike design

DESIGN			Frequency	Percent	Valid Percent	Cumulative Percent
1	Valid	Ikke sortert leksikografisk	22	62,86	62,86	62,86
		Sortert etter sitteplass	1	2,86	2,86	65,71
		Sortert etter frekvens	3	8,57	8,57	74,29
		Sortert etter pris	9	25,71	25,71	100,00
		Total	35	100,00	100,00	
2	Valid	Ikke sortert leksikografisk	16	51,61	53,33	53,33
		Sortert etter sitteplass	3	9,68	10,00	63,33
		Sortert etter reisetid	1	3,23	3,33	66,67
		Sortert etter frekvens	7	22,58	23,33	90,00
		Sortert etter pris	3	9,68	10,00	100,00
		Total	30	96,77	100,00	
	Missing -1	1	3,23			
Total	31	100,00				
3	Valid	Ikke sortert leksikografisk	45	84,91	84,91	84,91
		Sortert etter reisetid	1	1,89	1,89	86,79
		Sortert etter frekvens	1	1,89	1,89	88,68
		Sortert etter pris	6	11,32	11,32	100,00
		Total	53	100,00	100,00	
4	Valid	Ikke sortert leksikografisk	33	63,46	64,71	64,71
		Sortert etter sitteplass	2	3,85	3,92	68,63
		Sortert etter reisetid	1	1,92	1,96	70,59
		Sortert etter frekvens	1	1,92	1,96	72,55
		Sortert etter pris	14	26,92	27,45	100,00
	Total	51	98,08	100,00		
Missing -1	1	1,92				
Total	52	100,00				
5	Valid	Ikke sortert leksikografisk	38	86,36	86,36	86,36
		Sortert etter sitteplass	1	2,27	2,27	88,64
		Sortert etter pris	5	11,36	11,36	100,00
		Total	44	100,00	100,00	
6	Valid	Ikke sortert leksikografisk	31	56,36	64,58	64,58
		Sortert etter frekvens	8	14,55	16,67	81,25
		Sortert etter pris	9	16,36	18,75	100,00
		Total	48	87,27	100,00	
	Missing -1	7	12,73			
Total	55	100,00				
13	Valid	Ikke sortert leksikografisk	14	73,68	73,68	73,68
		Sortert etter sitteplass	2	10,53	10,53	84,21
		Sortert etter frekvens	1	5,26	5,26	89,47
		Sortert etter pris	2	10,53	10,53	100,00
		Total	19	100,00	100,00	
14	Valid	Ikke sortert leksikografisk	16	57,14	59,26	59,26
		Sortert etter sitteplass	1	3,57	3,70	62,96
		Sortert etter reisetid	4	14,29	14,81	77,78
		Sortert etter frekvens	2	7,14	7,41	85,19
		Sortert etter pris	4	14,29	14,81	100,00
		Total	27	96,43	100,00	
	Missing -1	1	3,57			
Total	28	100,00				
15	Valid	Ikke sortert leksikografisk	24	88,89	88,89	88,89
		Sortert etter reisetid	1	3,70	3,70	92,59
		Sortert etter frekvens	2	7,41	7,41	100,00
		Total	27	100,00	100,00	
16	Valid	Ikke sortert leksikografisk	13	65,00	76,47	76,47
		Sortert etter reisetid	1	5,00	5,88	82,35
		Sortert etter frekvens	3	15,00	17,65	100,00
		Total	17	85,00	100,00	
	Missing -1	3	15,00			
Total	20	100,00				

