ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Telser, Harry; Becker, Karolin; Zweifel, Peter

Working Paper

Validity and reliability of willingness-to-pay estimates: Evidence from two overlapping discrete-choice experiments

Working Paper, No. 0412

Provided in Cooperation with:
Socioeconomic Institute (SOI), University of Zurich

Suggested Citation: Telser, Harry; Becker, Karolin; Zweifel, Peter (2008) : Validity and reliability of willingness-to-pay estimates: Evidence from two overlapping discrete-choice experiments, Working Paper, No. 0412, University of Zurich, Socioeconomic Institute, Zurich

This Version is available at:
http://hdl.handle.net/10419/76159

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

University of Zurich

Socioeconomic Institute
Sozialökonomisches Institut

Working Paper No. 0412

Validity and Reliability of Willingness-to-Pay Estimates from Two Overlapping Discrete-Choice Experiments

Harry Telser, Karolin Becker and Peter Zweifel

Revised version March 2008

Socioeconomic Institute
University of Zurich

Working Paper No. 0412

Validity and Reliability of Willingness-to-Pay Estimates from Two Overlapping Discrete-Choice Experiments

Revised version March 2008

Author's addresses

Harry Telser
E-mail: harry.telser@polynomics.ch

Karolin Becker
E-mail: karolin.becker@polynomics.ch

Peter Zweifel
E-mail: pzweifel@soi.uzh.ch

Publisher

Sozialökonomisches Institut
Bibliothek (Working Paper)
Rämistrasse 71
CH-8006 Zürich
Phone: +41-1-634 21 37
Fax: +41-1-634 49 82
URL: www.soi.uzh.ch
E-mail: soilib@soi.uzh.ch

Validity and Reliability of Willingness-to-Pay Estimates: Evidence from Two Overlapping Discrete-Choice Experiments

Harry Telser[1, 2)], Karolin Becker[1)], and Peter Zweifel[2)]

[1)] Polynomics, Baslerstrasse 44, CH-4600 Olten, Switzerland
[2)] University of Zurich, Socioeconomic Institute, Hottingerstr. 10, 8032 Zurich, Switzerland

Correspondence to: harry.telser@polynomics.ch

This version: March 2008

Acknowledgments

Financial support by Interpharma, Santesuisse, the Association of Swiss Pharma Companies (VIPS), MSD Switzerland, the Federal Social Insurance Office (FSIO), the State Secretariat for Economic Affairs SECO, the Swiss Medical Students' Association, and the Merian-Iselin Hospital is gratefully acknowledged.

Abstract

Discrete-choice experiments, while becoming increasingly popular, have rarely been tested for validity and reliability. This contribution purports to provide some evidence of a rather unique type. Two surveys designed to measure willingness-to-accept (WTA) for reform options in Swiss health care and health insurance are used to provide independent information with regard to two elements of reform. The issue to be addressed is whether WTA values converge although the three overlapping attributes (a more restrictive drug benefit, a delayed access to medical innovation, and a change in the monthly insurance premium) are embedded in widely differing choice sets. Experiment A contains rather radical health system reform options, while experiment B concentrates on more familiar elements such as copayment and the benefit catalogue. While mean WTA values differ between experiments, they tend to vary in similar ways, suggesting at least theoretical validity and reliability.

Keywords: willingness-to-pay, discrete choice experiments, validity, reliability, framing effects

JEL Codes: C35, C93, I11

1. Introduction

In health economics, stated preference methods such as discrete-choice experiments (DCE) have been increasingly used to measure benefits. Applications of DCE to the valuation of health care programs have become numerous recently [1-3]. In a DCE, individuals are given a choice between hypothetical commodities. From the choices respondents make between the goods differing in product attributes, the researcher can derive the implicit trade-offs between these product attributes. This allows the computation of respondents' marginal utility for each product attribute. With the inclusion of a cost or price attribute, a money value can be calculated for each characteristic as well as for the entire good or program. The advantage of this approach over other stated preference methods such as e.g. the contingent-valuation method (CVM), lies in the fact that the price attribute is one among several, of which all vary in the course of the experiment. Biases that occur when individuals are asked about their willingness-to-pay (WTP) directly are less likely to be observed in DCE [4]. Applications in health economics so far mainly comprised studies of WTP for different treatment methods [5-9] or different hospital or physician services [10-11]. DCEs like the present one, i.e. dealing with the health care system as a whole, are rare [12].

DCEs usually are limited to a small number of attributes [1,13]. Especially when the product is defined as an entire health care system, this raises the question of whether neglected attributes influence the decisions of the respondents, causing bias in the WTP values obtained. Furthermore, it is often unclear what specific effects certain reform proposals will have in practice, which makes the hypothetical character of the experiment more problematic. However, reliable and valid WTP values are of utmost importance if policy recommendations are to be derived from DCE studies [1].

This paper adds to the literature in two ways. First, it seeks to measure and analyze WTP (or rather, willingness-to-accept, WTA) values for proposed changes to an entire health care system – a thing that has rarely been attempted thus far. Second, it benefits from the unique opportunity to conduct two parallel DCEs with two independent samples for addressing validity and reliability issues, made possible by the inclusion of three overlapping attributes in both DCEs.

To the best knowledge of the authors, comparative DCEs were only conducted by Slothuus-Skoldborg and Gyrd-Hansen [12], who analyzed WTP for screening methods for different types of cancer, and by Merino-Castellò [6], who studied the demand for two different drugs. In the field of environmental economics, DeShazo and Fermo [14] have undertaken two

DCEs concerning national park attributes in two different countries. The aim of the present study is to examine whether differences in the attribute set describing a hypothetical product have an influence on preferences and WTP values of respondents. This can be tested thanks to an overlapping subset of attributes.

This paper is structured as follows. Section 2 gives a brief overview of the underlying theory and methodology of a DCE. In section 3, reliability and validity issues are discussed and the pertinent literature reviewed. Section 4 explains the setup of the studies and the measures taken to improve validity and reliability in the analysis. The estimation results of three model specifications are presented in section 5, where WTA values are also derived, followed by a discussion of the results with respect to reliability and validity. Section 6 concludes.

2. Theoretical Background

Based on random utility theory [15-17] and Lancaster's new demand theory [18], discrete-choice experiments (DCEs) are designed to allow individuals to express their preferences for non-marketed and/or hypothetical goods that vary in their product characteristics. A rational individual will always choose the alternative with the higher level of utility. The decision-making process within a DCE can thus be seen as a comparison of utility values V_{ij} determined by

$$V_{ij} = v(a_j, p_j, y_i, s_i, \varepsilon_{ij}), \qquad (1)$$

where $v(\cdot)$ represents the indirect utility function of individual i for a good j, described by a vector of attributes a_j and a price denoted by p_j. The income of individual i is y_i, while the socioeconomic characteristics are denoted by s_i, and the error term, by ε_{ij}. Given an additive error term, the individual will choose contract j over contract l if

$$w(a_j, p_j, y_i, s_i) + \varepsilon_{ij} \geq w(a_l, p_l, y_i, s_i) + \varepsilon_{il}. \qquad (2)$$

Here, $w(\cdot)$ is the deterministic component of the utility that can be estimated, while the error term reflects unobservable factors that vary between individuals and alternatives. The utility function $v(\cdot)$ can be inferred from observed choices by assuming that the probability P_{ij} of choosing alternative j over l, given the vector of attributes, equals the probability of occurrence of the utility difference, and therefore

$$P_{ij} = \text{Prob}\,[\varepsilon_{il} - \varepsilon_{ij} \leq w(a_j, p_j, y_i, s_i) - w(a_l, p_l, y_i, s_i)]. \qquad (3)$$

The utility function is usually assumed to be linear,

$$w_j = c_0 + \sum_1^K \beta_k a_{jk} \,, \qquad (4)$$

where c_0 is a constant, $\beta_1,...\beta_K$ are the parameters to be estimated, and $a_{j1},...a_{jK}$ are the different attributes of the commodity *j*. There is empirical evidence suggesting that a linear specification leads to good predictions in the middle ranges of the utility function [13].

The marginal rate of substitution (MRS) between two attributes *k* and *m* is given by

$$MRS := -\frac{\partial v_j / \partial a_k}{\partial v_j / \partial a_m} \,. \qquad (5)$$

Denoting the *m*-th attribute as price, MRS indicates the marginal WTP for attribute *k*.

The deterministic part of the model is usually estimated by logit and probit techniques, depending on the assumption being made on the distribution of the error terms. In a DCE, participants are usually presented with a sequence of choices, which gives the underlying data a panel structure, thus making a random-effects specification appropriate. For a more detailed explanation of discrete choice models and their applications, see [13] or [19].

3. Reliability and validity issues

Reliability and validity of WTP values are an important issue with regard to stated preference elicitation methods such as DCEs. Following Jöreskog and Goldberger [20] (see also [21]), let y_1 and y_2 denote two measurements of a latent variable *x*. Since the determinants of *x* (neglected in eq. (6) below for simplicity) may change between the two observations, *x* in general will have unobserved values x_1 and x_2. However, measurements may be contaminated by an irrelevant latent influence *z* as wells as measurement errors e_1 and e_2. In the present context, y_1 and y_2 are 'observed' (calculated) WTP values, while *x* is the marginal rate of substitution defined in eq. (5) that may differ between subsamples ($x_1 \neq x_2$). However, by the maintained hypothesis, *x* should not be a function of e.g. political attitudes *z*. Therefore, observations are generated according to the measurement model

$$y_1 = \lambda_1 x_1 + \mu_1 z_1 + e_1, \quad y_2 = \lambda_2 y_2 + \mu_2 z_2 + e_2 \,, \qquad (6)$$

with (λ_1, μ_1, λ_2, μ_2) denoting the loadings of measurements on latent variables and (e_1, e_2) stochastic i.i.d. measurement errors.

There are several sources of systematic error. Relevant product attributes may not have been recognized (the *x* vector is too short to begin with), seemingly irrelevant attributes may have been excluded (the *x* vector has been erroneously shortened), or the underlying indirect utility function may have been wrongly specified (the structural model determining *x* is wrong).

Random measurement errors are always present; in DCEs, they are even part of the specification based on the random utility model [see eq.(3)].

Reliability can then be defined as the reproducibility of results on average, which means that loadings should have the same values across samples ($\lambda_1 = \lambda_2$, $\mu_1 = \mu_2$) and random errors be zero on expectation, $E(e_1) = E(e_2) = 0$. There are several ways to check for reliability (see e.g. [22]). The *test-retest method* benefits from repeated measurement; in that case, eq. (6) applies directly. *Parallel testing* involves the simultaneous use of two slightly different instruments; in this case, y_1 and y_2 are two different indicators with loadings $\lambda_1 \neq \lambda_2$, that differ in a predictable way provided $\mu_1 = \mu_2 = \text{const}$. Finally, in the *alternate-form method*, the sample is split, with part of the observations reserved for re-estimation using a variant of the measurement method. Here, y_1 and y_2 refer to the two segments of the sample that again should induce loadings λ_1 and λ_2 that differ in a predictable way.

With regard to reliability of DCE, there has been work on stability (of λ_1, μ_1, λ_2, μ_2) over time [23-24], using the same sample of respondents for two follow-up DCEs. Their test suggests temporal stability of the measurement model. Choice set design was examined in various studies on ordering effects, with mixed results. Some authors do not find evidence suggesting that results depend on the ordering of sets [9,25], whereas others do find such evidence [11]. Randomly changing the order of the choices respondents have to make therefore continues to be an accepted method of experiment design in order to avoid bias due to learning and fatigue effects [6]. Sensitivity to the choice of attribute range and attribute levels as well as to the order of presentation of attributes has been considered in several research papers [7,12,26-29]. Lloyd [30] gives an overview of the literature devoted to the analysis of the decision-making process and its influencing factors. There is considerable evidence that depending on ranges and levels of attributes, dominant preferences or lexicographic orderings are more or less likely to occur.

The *validity* requirement is more stringent, requiring not only $\lambda_1 = \lambda_2$ w.r.t. indicators, but also $\mu_1 = \mu_2 = 0$ in eq. (6), i.e. the exclusion of irrelevant determinants of WTP. However, $\lambda_1 \neq \lambda_2$ or $\mu_1 \neq \mu_2 \neq 0$ is also admissible provided the maintained hypothesis makes a testable prediction regarding these loadings, i.e. the relative quality of the two indicators. Thus, not only must measurement be reliable but also free of systematic (or at least uncontrolled) bias that could be caused by a variable z that is irrelevant by hypothesis. There exist different concepts of validity differing in requirements w.r.t. different populations and systematic measurement errors. In case of *internal validity*, the concept refers to the population studied in the experiment. It hinges importantly on the confirmation of prior theoretical hypotheses. Thus, *theo-*

retical validity typically tests for the expected signs of coefficients suggested by economic theory, such as diminishing marginal utility of income and differences between socioeconomic groups. Theoretical validity is a relatively weak concept since it may hold even though systematic error exists, as long as error is the same across socioeconomic groups ($\lambda_1 \neq \lambda_2$ and $\mu_1 = \mu_2 \neq 0$, where the subscripts 1 and 2 now refer to different socioeconomic groups). Thus, it is possible for a DCE to contain systematic error and yet produce theoretically valid results. Various health services researchers have tested for the theoretical validity of a DCE [5,9-11,31-32]. In most cases, results are in accordance with theoretical expectations and hence indicate internal validity of DCE.

External validity is a more generally defined concept. Given external validity, the results of a study can be generalized to different research methods, locations, groups of people, and decision-making situations. It requires the absence of systematic measurement error, i.e. $\mu_1 = \mu_2 = 0$. This condition is stronger than theoretical validity, which only requires the systematic error to be predictable in the light of (economic) theory.

External validity may be further subdivided into convergent and criterion validity; both have been addressed in health care applications of DCE. *Convergent validity* obtains if different methods that are designed to generate information about the same theoretical construct x have convergent results ($y_1 \approx y_2$). The comparator y_2 should constitute a valid elicitation technique ('gold standard'); thus $\lambda_1 = \lambda_2 = 1$ and $\mu_1 = \mu_2 = 0$ is required, although y_1 and y_2 are generated by different methods. Ryan [4] compares the results derived from a Contingent Valuation (CV) dichotomous choice study with those from a DCE concerning preferences for assisted reproductive techniques.[1] However, there are doubts about the validity of CV (see e.g. [33-34]). Therefore, conclusions w.r.t. the validity of a DCE may be unfounded.

Criterion validity is considered the strongest form of validity. It obtains if the results of a method correspond with those from a decision-making situation that is external to the experiment. For example, WTP calculated in a DCE can be compared to WTP implied by actual choices that provide the external criterion; in this case, an alternative with a known value of λ is available. Telser and Zweifel [35] compare WTP for hip protectors derived from a DCE with actual choices (that did not involve actual payment, however) the same respondents made later. The results indicate that the DCE may have criterion validity.

The present work addresses the issues of reliability and theoretical validity in a way that has to our knowledge not been considered in prior research. Here, two DCEs were carried out on two independent representative samples of the Swiss population. Each set of participants were

[1] Hanley et al. [2] give a literature overview for convergent validity in the environmental context.

presented a series of health insurance contracts with different product characteristics, except for three overlapping attributes (the health insurance premium being one of them). Given reliability and hence $\lambda_1 = \lambda_2$, $\mu_1 = \mu_2$ and $E(e_1) = E(e_2) = 0$, the WTP values derived for the overlapping attributes should be comparable in spite of the fact that experiment A otherwise revolved around more far-reaching changes than experiment B (which specifically contained Managed Care alternatives) This would indicate the absence of framing effects. For a confirmation of theoretical validity, WTP measures should vary in both experiments between different socioeconomic groups in ways predicted by economic theory. However, differences between the outcomes of the two experiments can still be persistent, caused by systematic error [$\mu_1 \neq \mu_2$ in eq. (6)]. Therefore, while WTP values derived from the two DCEs may differ in their levels, they would vary with determinants and across socioeconomic groups as predicted by economic theory.

4. Experiment Design

To elicit preferences of the Swiss residential population with regard to proposed changes in the health care system, two DCEs were designed featuring hypothetical insurance contracts. Their attributes should reflect the reforms that are debated at present by policy makers. These contract attributes were preselected in expert sessions with representatives of the Swiss health care system and their relevance checked in a pretest. The nine characteristics retained (plus PREMIUM as the price attribute) are listed in Table 1. With regard to those attributes that were not taken into account, participants were told that the status quo and the alternative were identical in this regards in order to avoid omitted variable bias in the econometric analysis.

The possibilities considered are the following. In experiment A, free choice of physician is restricted to a list of contract providers (PHYSLIST). The list can be made up applying different selection criteria, viz. cost, quality, or efficiency, defined as the quality-cost ratio (PHYSCOST, PHYSQUAL, and PHYSEFF). The number of hospitals available is reduced by closing small local hospitals in favor of larger centralized ones (HOSPITAL). At present, long-term care is only partially covered by mandatory health insurance in Switzerland. The proposed change comprises full coverage of long-term care, to be financed by those over 50 years old (LTCARE). The current drug benefit is very comprehensive; it would be changed by excluding drugs for minor illnesses such as the common cold (MINOR) or reimbursing only the cheapest drug available, usually a generic (GENERICS).

Experiment B was devoted to more conventional insurance parameters. The existing annual deductible (CHF 230 at the time, with 1 CHF = 0.8 US$ in 2004) and copayment (10 percent)

are varied (DEDUCTIBLE, COPAYMENT). The coverage of alternative medicine is expanded or reduced compared to the status quo, in which only few therapies are covered (ALTMED). Access to innovative treatments (currently immediate after a decision by an expert committee) is delayed by two years after approval by official authorities (INNOVATION). Finally, each insurance contract is characterized by an absolute change in the monthly insurance premium (PREMIUM).

Table 1 Product attributes and levels in experiments A and B

Attribute	Labels	Levels [1]
Experiment A		
List of contract providers	PHYSLIST PHYSCOST PHYSQUAL PHYSEFF	- Status quo: free choice of physician in the home canton - List of providers: Cost criterion, Quality criterion, Cost-quality (efficiency) criterion
Centralization of hospitals	HOSPITAL	- Status quo: existing hospitals - Closing of local hospitals
Long-term care	LTCARE	- Status quo: nursing care is only partially covered - Coverage of long-term care, financed by those aged over 50
Medication for minor illnesses	MINOR	- Status quo: All drugs on the official list are reimbursed - Medications for minor diseases such as the common cold have to be paid out-of-pocket
Experiment B		
Deductible	DEDUCTIBLE	- Status quo: CHF 230, 400, 600, 1,200, 1,500 per year [2] - CHF 0, 2,400, 4,800 per year [2]
Copayment	COPAYMENT	- Status quo: 10% (max. CHF 600) [2] - 20% (max. CHF 700) [2]
Alternative medicine	ALTMED	- Status quo: some treatment methods are covered - Additional alternative treatment methods are covered - Fewer alternative treatment methods are covered
Joint Attributes		
Generics	GENERICS	- Status quo: all drugs on the list are reimbursed - The cheapest product on the market is covered
Innovation	INNOVATION	- Status quo: all treatment methods are covered as soon as they get approved - Innovative treatments are covered only two years after introduction
Premium	PREMIUM	- Reduction of the monthly premium by CHF 10, 25, 60 [2] - Increase/ reduction of the monthly premium by +/- CHF 50, 25 or 10 [2] (Experiment B)

[1] Coding for the dummy variables: status quo=0, alternative=1 (in the case of ALTMED: 0=fewer covered, 1=additional covered)

[2] 1 CHF=0.8 US$ at 2004 exchange rates

Such a high number of attributes, however, is cognitively too burdensome for respondents to evaluate [1,13]. For this reason, experiment A centers on Managed Care-related attributes (PHYSLIST, HOSPITAL, LTCARE, MINOR), while experiment B emphasizes more conventional parameters of health insurance (DEDUCTIBLE, COPAYMENT, ALTMED).

The attributes present in both experiments are delayed access to innovation (INNOVATION), restricted drug benefit (GENERICS), and the change in the monthly premium (PREMIUM). A price attribute is necessary to derive money WTA values.

Even so, the total number of attributes and their levels combine for a very large number of scenarios, which would cause interviews of excessive lengths. In an environmental application, Hanley et al. [2] found that increasing the number of choices influences parameter estimates. In most studies in health care, the number of choices ranges from 9 to 16 per respondent [1]. Using statistical design theory [36-38], the number of alternatives was reduced to obtain a fractional design that makes estimation of main effects and two-way interaction effects possible (so-called resolution 5 orthogonal design [39]).

For experiment A, 40 alternatives were selected, for experiment B, 27. These alternatives were randomly assigned to 4 and 3 split samples, respectively. To obtain a set of 10 choices per person in each split sample, one choice was included twice in each sequence of experiment B. This allows the answers of a given individual to be tested for consistency. These 10 choices were presented in a random ordering to avoid responses being affected by learning and fatigue effects. Each alternative had to be evaluated against the status-quo insurance contract. No opting-out possibility was provided in view of the fact that health insurance is mandatory in Switzerland.

The organization of Swiss health insurance facilitates conducting a choice experiment of this degree of complexity. Several elements of choice were introduced in 1996 as part of a reform. In the status quo of 2003, the insured could already choose between different levels of annual deductibles, with CHF 230 (US$ 184 at 2004 exchange rates) being the minimum, and between conventional fee-for-service and Managed Care alternatives. In addition, consumers can change their insurer every year, basically without bearing transaction costs. Insurance premiums differ between competing insurers and regions but are otherwise uniform across sex and age groups. About 80 percent of consumers have some kind of supplementary private insurance, which, however, must not cover legally prescribed cost sharing (viz., the CHF 230 deductible plus 10 percent copayment on health care expenditure with an annual cap at CHF 600). The Swiss are therefore familiar with choice options in their health insurance, which should make the experiments less hypothetical.

Realism is important because complexity and experience seem to influence preferences stated in a DCE [23,29,40-41]. If the elicitation task is too demanding, or if people are unfamiliar with the topic, preferences might be incomplete or unstable, being formed and adjusted in the course of the experiment [30,42]. Other studies suggest that the choice of the payment vehicle (e.g. tax or insurance premium) to represent the price attribute of a DCE may be critical [7]. The choice of an appropriate price variable and its range and levels to induce tradeoffs has also been discussed [12]. Moreover, recent research [43] finds that in a health-related context, it makes a difference whether or not respondents are reminded that the price is to be paid out of pocket rather than by the insurer. These issues, however, do not seem to be of much relevance to the present study because the two experiments deal with insurance contracts, with Swiss consumers paying different premiums according to type of plan out of pocket. Moreover, plans impose copayments throughout.

The two experiments were developed and implemented in a coordinated way in order to allow for a joint analysis of the data. The documentation materials accompanying the DCE were identical. Two representative telephone surveys with 1,000 persons aged over 25 years[2] were conducted independently in the German and French parts of Switzerland during September 2003. The procedure was in two steps due to the special character and information requirements of a DCE. In a first telephone contact, people were asked if they would be willing to take part in the study. Those agreeing to participate received a package containing documentation materials to make sure that all respondents had the same information about the Swiss health care system and knew the deductible level and premium of their health insurance plan. In this way, respondents were given time to reflect, which may result in more consistent choices during the experiment [44].

For the actual DCE, each respondent received 11 decision cards. One (blue) card described the status quo with regard to the attributes to be varied in the experiment. The remaining 10 (yellow) cards described the 10 alternative insurance contracts respondents had to opt for or against. Attributes were described in detail including a glossary.[3] The experiments themselves were conducted during a later telephone contact. Respondents also answered additional questions concerning their utilization of health care services, overall satisfaction with the health care system, insurer and insurance policy, and their attitudes towards innovation in health

[2] Below age 26, reduced premiums for young adults and children apply.

[3] See also [31] on the importance of a-priori information for consistency of choices in DCE. The authors propose a summary sheet describing attributes and their levels. Such a sheet was provided in both experiments A and B.

insurance. Socioeconomic variables included age, sex, education, total household income, place of residence, occupation, and household size.

5. Results

5.1 Descriptive statistics

Table 2 provides information on the socioeconomic characteristics of the two samples, showing that they are very similar. In both samples, one-half of respondents are male, and mean age is 48 years. Monthly per capita income is about CHF 3,000 (1 CHF=0.8 US$ at 2004 exchange rates). Respondents live in a household averaging 2.5 persons and pay a monthly premium of CHF 222 (sample A) and CHF 240 (sample B), respectively. The mean annual deductible is about CHF 650 for both samples, with an overrepresentation of those individuals having chosen the highest possible deductible (CHF 1,500), compared to official statistics.

Table 2 Descriptive statistics of the two samples

	Sample A			Sample B		
	Mean	Std.dev.	Median	Mean	Std.dev.	Median
Sex (Dummy, male=1)	0.51	0.50	1	0.49	0.50	0
Age	48.27	16.13	45	48.85	15.27	47
25-40 (Dummy)	0.39	0.489	0	0.36	0.48	0
Over 62 (Dummy)	0.24	0.428	0	0.21	0.41	0
Language (Dummy, French=1)	0.29	0.455	0	0.30	0.46	0
Monthly income (CHF p.c.)	2938	1783	2400	2952	1842	2400
Household size	2.54	1.33	2	2.59	1.36	2
Insurance premium (CHF)	221	65	217	240	69	218
Deductible (CHF)	635	510	400	656	510	400
Hospital stay (Dummy, yes=1)[1)]	0.16	0.36	0	0.11	0.32	0
Physician visit (Dummy yes=1) [1)]	0.57	0.50	1	0.49	0.50	0

[1)] Previous 12 months (A) and 6 months (B), respectively.

Another point of interest for the present study is actual willingness to change health insurer or type of contract. No less than 77 percent of the respondents in both samples stated that they had not changed their insurer during the past 5 years, and 64 (B: 66) percent had not undertaken a change of their insurance contract, such as switching to a different deductible or to a Managed Care option. Therefore, a preference for the status quo is expected to characterize both experiments. So-called status quo bias is also likely in view of the rather short time since

the introduction of choice elements and the uncertainty surrounding future health care utilization [45].

The overall results from the DCEs do point to limited flexibility in choice behavior. Out of the 10,000 possible choices per experiment, only 21 (experiment A) and 18 percent (experiment B) were made in favor of the alternative. Moreover, these figures might mask the fact that few individuals make up for them. However, the evidence does not support this suspicion since no less than 65 percent (A) and 60 percent (B) of the respondents deviated from their status quo at least once.

On the other hand, respondents of higher age, female sex, with a lower education level, and with a lower initial premium are less likely to prefer the alternative at least once. To avoid selection bias in the results, these 'non-traders' (making up 35 and 40 percent of the samples, respectively) are not dropped from the analysis [5]. There are at least three explanations for the 'non-traders' phenomenon. First, the levels of the attributes offered in the experiment may not have been extreme enough to induce a trade-off between attributes; second, the attributes may not have been sufficiently valued by respondents; and finally, respondents may have simply made errors. However, this last explanation can be discarded on two counts. The consistency check of experiment B shows that only 13 out of 1,000 individuals made 'incorrect' and inconsistent decisions, and 81 percent of respondents in sample A (88 percent in B) stated that they found the experiment easy or rather easy to accomplish.

5.2 Estimation Results

5.2.1 Estimation of a simple linear model

To begin with, a simple model (Model 1) is estimated for both experiments. Here, the utility function is assumed to be the same linear one for all individuals, with the attributes of the health insurance contract as described in Table 1 as its sole arguments. For the deductible, a quadratic term was included to account for a decreasing marginal utility of income. For the premium, the same argument applies in principle; however, preliminary tests showed the squared value of PREMIUM to be statistically insignificant.

Model 1 provides a first benchmark since almost every application of DCEs in the health care field uses this specification. The estimation results for the two scenarios are shown in the appendix. With the exception of the two attributes describing a restricted access to drugs in experiment A (GENERICS and MINOR), all coefficients are statistically significant and have expected signs. Since the two overlapping attributes amount to restrictions compared to the status quo, their valuation is given by willingness-to-accept (WTA), or compensation de-

manded, rather than WTP values.[4] Therefore, the WTA values in Table 3 indicate the money amount of compensation that is necessary on average for respondents to accept a less generous plan.

In experiment A, accepting a physician list based on a cost criterion (PHYSCOST) requires the highest compensation of CHF 103, more than one-third of the average monthly premium of CHF 270 (as of 2003, according to official statistics). By way of contrast, the exclusion of medications of minor ailments from reimbursement (MINOR) might even meet with a negative WTA, i.e. a positive WTP value. This may well be due to a 'warm-glow' effect [48-49], which occurs when respondents believe that a particular alternative meets with approval by society.

Nonetheless, the difference between the WTA values for PHYSCOST, PHYSQUAL, PHYSEFF, and HOSPITAL on the one hand and MINOR on the other constitutes a first piece of evidence suggesting theoretical validity. After all, the restriction of choice implied by MINOR, being far less important then the restrictions imposed by PHYSCOST, PHYSQUAL, PHYSEFF, and HOSPITAL, should be associated with a smaller (and possibly even zero) WTA. The mean values and confidence intervals (which are disjoint, pointing to significant differences) within experiment A confirm this prediction. Second, MINOR and GENERICS both concern drug use only and should therefore have similar WTA values given validity. This prediction is borne out as well.

In experiment B, both a higher deductible (DEDUCTIBLE) and an increase of the rate of copayment from 10 to 20 percent (COPAYMENT) clearly require compensation to be accepted. Since the first change is defined in terms of CHF 1, it entails a minimal increase in financial risk and therefore should be associated with a much smaller WTA value than the latter. This is confirmed, providing evidence for theoretical validity.

[4] Typically, WTA values for restrictions from an existing level are much higher than WTP values for a corresponding improvement from a lower level (see e.g. [46]Horowitz and McConnell, 2002 and [47]Zweifel et al., 2006).

Table 3 WTA derived from Model 1 (attributes only), in CHF per month

	WTA	Std.err. [1]	z value	95% confidence interval	
Experiment A					
PHYSCOST	103.28	13.16	7.85	77.49	129.06
PHYSQUAL	53.33	8.85	6.03	35.98	70.67
PHYSEFF	41.96	7.78	5.39	26.71	57.21
HOSPITAL	37.30	5.67	6.58	26.18	48.42
LTCARE	24.90	4.76	5.24	15.57	34.22
MINOR	-6.47	5.33	-1.21	-16.92	3.97
GENERICS	2.67	5.49	0.49	-8.08	13.43
INNOVATION	64.64	7.88	8.20	49.19	80.09
Experiment B					
DEDUCTIBLE [2]	0.0320	0.0017	20.60	0.0314	0.038
COPAYMENT	18.91	2.98	6.34	13.06	24.75
ALTMED [3]	-24.71	3.11	-7.96	-30.80	-18.63
GENERICS	13.77	3.06	4.50	7.77	19.77
INNOVATION	38.39	3.33	11.54	31.87	44.91

[1] Standard errors computed by the delta method.

[2] Compensation required for a CHF 1 increase in the annual deductible for the mean individual with a deductible of CHF 656.

[3] Expanded coverage of alternative medicine (0=no inclusion).

1 CHF=0.8 US$ at 2004 exchange rates

Turning to the two overlapping attributes INNOVATION and GENERICS, their WTA values are ordered in a sensible way in both experiments. After all, delaying access to medical innovation by two years entails a larger risk than settling for generics (which are supposed to be chemically equivalent to original products). However, absolute WTA values differ substantially between experiments (see Figure 1). In experiment A, the WTA value for GENERICS is not distinguishable from zero, while in experiment B, it is significantly positive (CHF 14) . Yet the two 95% confidence intervals overlap, suggesting that the hypothesis of equality between the two experiments need not be rejected.

On the other hand, WTA values for INNOVATION are clearly positive in both experiments. But they differ, the 95% confidence intervals being disjunct. This is a first indication of lacking validity. Finally, Figure 1 reveals that the WTA values of experiment B are much more precisely estimated than those of experiment A, especially for INNOVATION. As there is no reason for the variance of random errors [Var(e_A), Var(e_B)] to systematically differ between experiments, this points to the possibility that $\mu_A \neq \mu_B$, a systematic error introduced by the

presence of strongly differing other attributes in the two experiments [see eq. (6), with subscripts 1 and 2 replaced by A and B].

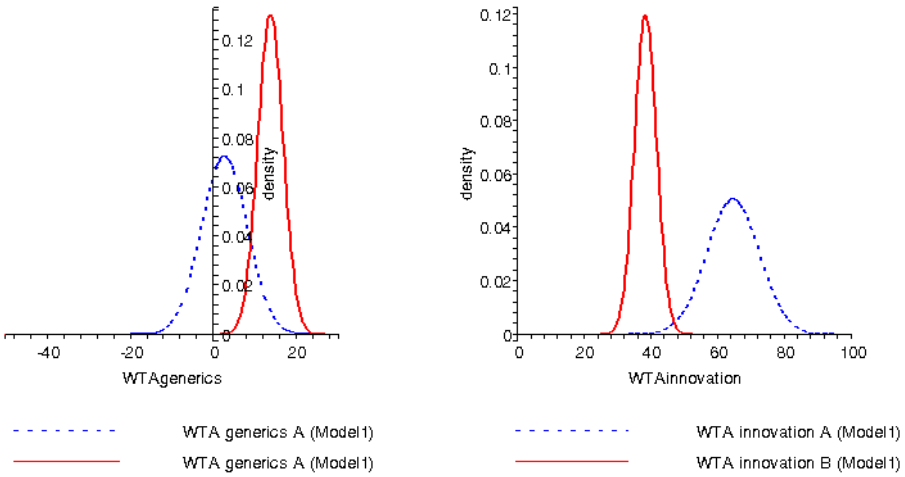

Figure 1 Distribution of WTA for GENERICS (left) and INNOVATION (right) in experiments A and B (Model 1)

5.2.2 Estimation of more comprehensive models

A first generalization of Model 1 is to allow interactions between attributes, relaxing the assumption of an additively separable utility function. This implies that WTP values become conditional on the values assumed by the other attributes in the respective experiment. Therefore, a comparison of WTA values pertaining to GENERICS and INNOVATION are only possible in the status quo. While significant interaction terms popped up here and there, it proved impossible to find a common specification for both experiments due to collinearity problems. Moreover, in experiment-specific estimations, WTP values of attributes turned out to be similar to those of Model 1 when evaluated at the status quo. For these reasons, this generalization was not pursued any further.

Next, the assumption that all respondents have the same utility function needs to be relaxed. Specifically, marginal utilities of attributes are now permitted to vary with socioeconomic characteristics of the respondents. This calls for introducing interaction terms in the econometric estimation. Two interaction models were estimated. The first one (Model 2) follows Johnson and Desvouges [50] by interacting the price attribute (here: PREMIUM) with socioeconomic characteristics, thus allowing for different marginal utilities of income between subgroups. The second interaction model (Model 3) goes one step further by letting the marginal utility not only of the price attribute but of all product attributes differ between socioeconomic groups. This comprehensive specification is best capable of capturing preference

heterogeneities. The socioeconomic characteristics included in Models 2 and 3 are age, gender, region (German- or French-speaking part of Switzerland), income, household size, health status, and initial level of premium paid.

For Model 2, WTA values relating to the two overlapping attributes are shown in Table 4. WTA values for the mean individual are close to those of Model 1 for both experiments, which was to be expected since a predominantly linear model works best for mean values.

Table 4 WTA derived from Model 2 (interactions with PREMIUM only), in CHF per month, evaluated at the mean individual of the estimation sample

	WTA	Std.err. [1)]	z value	95% confidence interval	
GENERICS					
Experiment A	3.28	5.90	0.56	-8.28	14.84
Experiment B	10.94	3.46	3.17	4.17	17.72
INNOVATION					
Experiment A	69.21	8.97	7.72	51.64	86.78
Experiment B	31.80	7.51	4.24	17.08	46.51

[1)] Standard errors computed by the delta method.

1 CHF=0.8 US$ at 2004 exchange rates

Again, the hypothesis of equal WTA with regard to GENERICS cannot be rejected, the overlap between the two distributions being even more marked than in Model 1 (compare Figures 2 and 1, left-hand side). By way of contrast, Model 2 results in a WTA value for INNOVATION that is clearly larger in experiment A than in experiment B (see the divergent distributions in Figure 2, right-hand side). This is still compatible with theoretical validity in view of consistently higher WTA values for INNOVATION than GENERICS in both experiments. The fact that this relative ordering of WTA values results from both Model 1 and Model 2 points to robustness of results. Their divergence, however, could be a sign of systematic error in either one of the experiments or in both [$\mu_1 \neq \mu_2 \neq 0$ in eq. (6)].

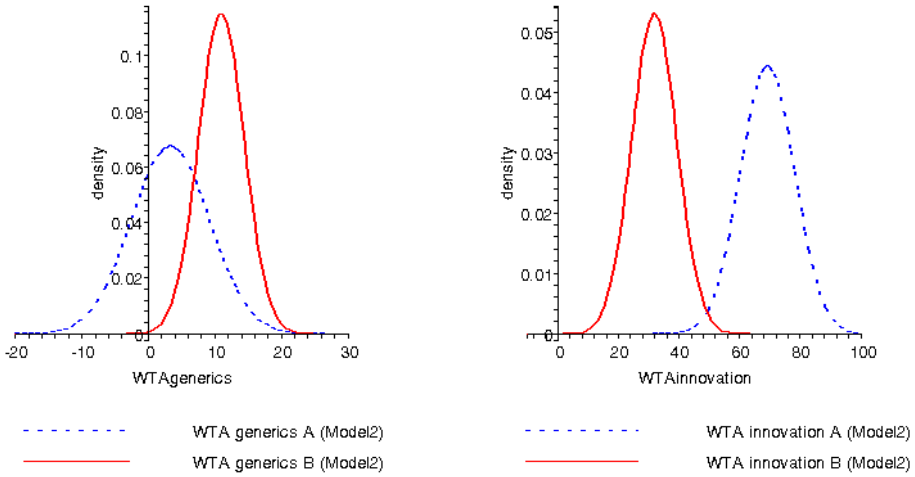

Figure 2 Distribution of WTA for GENERICS (left) and INNOVATION (right) in experiments A and B (Model 2)

As to Model 3 with its full set of interaction terms, INNOVATION continues to command a higher WTA than GENERICS in both experiments, suggesting theoretical validity. There is again an indication that WTA values might diverge between the two samples. However, the 95% confidence intervals are [-9.37; 13.68] for GENERICS in experiment A and [-2.61; 43.25] in experiment B, suggesting that the equality hypothesis need not be rejected (see also Figure 3, left-hand side). This outcome is due to a marked increase of estimated standard errors in Model 3, especially for experiment B (compare Figures 2 and 3), likely caused by multicollinearity between the many interaction terms. In the case of INNOVATION, intervals are [50.32; 84.32] in A and [4.65; 41.49] in B. Here, intervals do not overlap, suggesting rejection of the equality hypothesis.

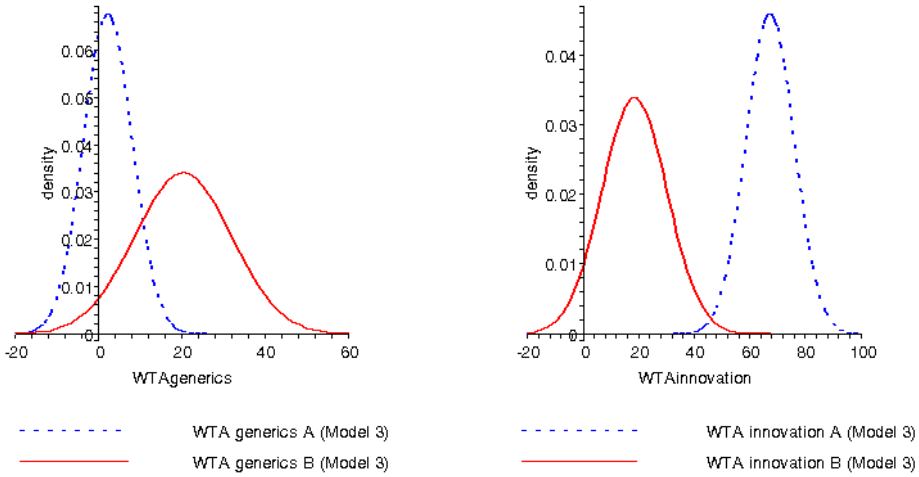

Figure 3 Distribution of WTA for GENERICS (left) and INNOVATION (right) in experiments A and B (Model 3)

On the whole, large standard errors limit the usefulness of Model 3 for an assessment of reliability but also of theoretical validity because WTA values tend not to differ according to socioeconomic characteristics in a significant way. Therefore, the following tests of reliability and theoretical validity are based on Model 2 (interaction terms with PREMIUM only), with results displayed in Table 5. These tests revolve around differences between socioeconomic groups, using the concepts introduced in section 3. Recall that for theoretical validity, WTA values may be biased overall, yet reflect differences between socioeconomic groups adequately in both experiments, i.e. $\lambda_1 \neq \lambda_2$ and $\mu_1 = \mu_2 \neq 0$ in equation (6).

Indeed, WTA values can be shown to differ between socioeconomic groups in a way predicted by economic theory while diverging between experiments A and B. This contributes evidence in favor of theoretical validity but persistence of a systematic bias in at least one experiment.

Age: Demand for health insurance coverage may rise with age because the asset 'health' to be protected becomes more risky. However, the value of this asset decreases beyond the earnings peak in the life cycle [51-52]. Therefore, the relationship between age and WTA concerning restrictions in coverage is ambiguous, precluding a test for theoretical validity. In Table 5, the compensation required tends to increase with age in both experiments. With regard to systematic bias, experiment A consistently yields smaller WTA values for GENERICS than experiment B in all age categories. In the case of INNOVATION, it is the other way around, with WTA values in experiment A exceeding by far those in B. Again, confidence intervals clearly overlap for GENERICS; therefore the equality hypothesis need not be rejected. These find-

ings do not hold for INNOVATION, pointing to systematic bias [i.e. inappropriate influence from one or several *z* variables in eq. (6)] in the case of INNOVATION.

Table 5 WTA derived from Model 2 (interactions with PREMIUM only), in CHF per month evaluated at the mean individual of the socioeconomic group

	WTA GENERICS				WTA INNOVATION			
	Experiment A		Experiment B		Experiment A		Experiment B	
	Mean	95% CI [1)]	Mean	95% CI [1)]	Mean	95% CI [1)]	Mean	95% CI [1)]
Age								
25-39	-2.24	-5.64, 10.11	11.02	5.99; 16.04	47.21	36.40; 58.02	32.01	26.03; 38.00
40-62	3.57	-9.05, 16.18	13.27	7.14; 19.41	75.29	48.47; 102.11	38.57	31.08; 46.05
63+	10.21	-28.04, 48.46	18.27	8.92; 27.62	215.53	44.94; 475.99	53.09	38.22; 67.96
Sex								
Male	4.50	-11.45, 20.44	13.57	7.29; 19.85	94.94	56.45; 133.42	39.44	31.94; 46.94
Female	2.60	-6.56; 11.75	12.62	6.85; 18.39	54.84	41.48; 68.24	36.68	29.86; 43.50
Region								
French	4.84	-12.40; 22.07	15.23	7.99; 22.46	102.14	49.23; 155.04	44.25	24.44; 54.05
German	2.93	-7.34; 13.26	12.34	6.71; 17.94	61.87	46.48; 77.25	35.83	29.46; 42.19
Health								
Ill	5.06	-12.97; 23.09	14.55	7.82; 21.29	106.78	49.93; 163.63	42.29	34.09; 50.50
Healthy	2.85	-7.20; 12.90	11.71	6.36; 17.06	60.18	45.51; 74.84	34.04	27.82; 40.25

[1)] 95% confidence intervals (standard errors computed using the delta method).

1 CHF=0.8 US$ at 2004 exchange rates

Sex: Women have lower levels of wealth, implying a lower effective demand for safety ceteris paribus.[5] Their WTA for restrictions of coverage in health insurance should therefore be lower. This prediction tends to be borne out; however confidence intervals especially in experiment A are too wide for statistical significance and hence support of theoretical validity.

Region: Cultural differences between the French- and German-speaking regions are likely to swamp any prediction that could be derived e.g. from income differences. A test for theoretical validity thus does not seem possible. Table 5 indeed shows WTA values for both GENERICS and INNOVATION to be higher in the French-speaking part of Switzerland. This effect is consistent across experiments. Again, confidence intervals overlap for GENERICS while this is hardly the case for INNOVATION, indicating presence of systematic error.

[5] Due to imprecise and partially missing income data of the respondents it was not possible to adequately derive WTA w.r.t. income.

Health: Certainly, the asset 'health' must have been more at risk among those reporting an illness during the past 12 months compared to the others, and possibly financial wealth as well to the extent that having consulted a physician or visited a hospital during the previous months entailed a copayment. Those having experienced illness are therefore predicted to demand a higher compensation for restrictions of health insurance coverage, especially concerning access to medical innovation. This prediction tends to be confirmed by the entries of Table 5, lending some support to theoretical validity. There are again strong indications of differing WTA values for INNOVATION between experiments.

Overall, the attribute INNOVATION seems to be affected by systematic bias across all socioeconomic groups. The attribute GENERICS displays results that are sufficiently similar between the two experiments to conclude that the DCEs exhibit theoretical validity and reliability.

6 Discussion and Conclusion

This paper reports on the results of two independent discrete-choice experiments (DCEs) seeking to measure the willingness-to-accept (WTA) values concerning restrictions in Swiss health insurance and health care, based on 1,000 interviews each conducted in 2003. Experiment A presented respondents with alternative health insurance contracts emphasizing Managed Care elements (physician lists, restricted choice of hospital) and compulsory long-term care insurance (see Table 1). Experiment B focused on more conventional and familiar health insurance parameters such as the annual deductible and the rate of copayment. The distinguishing feature of this study is the fact that the two experiments have three attributes in common, viz., a drug benefit limited to generics (GENERICS), delayed access to medical innovation (INNOVATION), and the price attribute (PREMIUM). This setup permits testing for reliability and validity.

With regard to (theoretical) validity, results yielded negative estimated marginal utilities for all restrictions in both experiments, as predicted. However, some of the implied WTA values do not differ from zero. Among the overlapping attributes, a reasonable prediction is that WTA for INNOVATION should be higher than for GENERICS because the first restriction may be binding in a situation when survival is at stake. This prediction tends to be borne out in both experiments, although the difference fails to be statistically significant. Additional testing becomes possible in a more general model allowing for WTA values to depend on socioeconomic characteristics (Model 2). Specifically, men (having more wealth) and the ill (being exposed to higher risk with regard to health and possibly wealth) should exhibit higher

WTA values than women and the healthy, respectively. Also, the ill/healthy differential should be more marked for INNVATION than for GENERICS. Whereas these predictions receive some empirical support in both experiments, conventional statistical significance levels fail to be attained.

Turning to reliability, the evidence comes from comparisons between experiments A and B. For GENERICS, experiment A yields lower estimated WTA values than experiment B; for INNOVATION, the reverse is true. This pattern holds also within different socioeconomic groups (Model 2). However, the 95 percent confidence intervals are too wide to cause rejection of the hypothesis that the WTA values for GENERICS are in fact equal, i.e. that they seem to be reliable. This does not hold with regard to the WTA values for INNOVATION, which seem to be distinct across the two experiments, likely indicating presence of a systematic bias.

The explanation may be that experiment A involved alternatives that are more hypothetical than in experiment B. Indeed, the attributes of experiment A are mostly of a general nature, with in the degree of restriction little specified (PHYSLIST, HOSPITAL, LTCARE; see Table 1). Experiment B on the other hand contained mostly attributes that respondents were familiar with; moreover, the restrictions, being in terms of money, were well defined (DEDUCTIBLE, COPAYMENT). However, Lack of concreteness in alternatives presented has been found to lead to a loss of precision in parameter estimates [13,53]. Indeed, the WTA values shown in Table 5 do exhibit larger confidence intervals for INNOVATION than for GENERICS in both experiments.

This explanation is compatible with the following observation. In 2003, respondents were already familiar with generics. In Switzerland, in intensive debate had revolved around the drug bill burdening social health insurance for quite a while. Opinion polls show increasing approval of the proposal to substitute cheaper generics for original branded drugs. By way of contrast, access to medical innovation does not refer to specific new treatments or new drugs. Furthermore, medical innovations are about future options, which are uncertain. For this reason, experiment A may have produced not only biased but also less reliable WTA estimates for Managed Care-type attributes and INNOVATION in particular. A full 19 percent of the respondents in experiment A had some difficulties in making their choices, compared to 11 percent in experiment B. It is therefore possible that in experiment A (despite the prior information on attributes provided), respondents' preferences were incomplete and may have been formed or adjusted during the course of the experiment. In this case, WTA (or WTP) values depend on the attribute set chosen, undermining validity.

In conclusion, theoretical validity tends to receive empirical support in both experiments in all cases where economic theory makes predictions concerning differences between socioeconomic groups. However, in a comparison between the two experiments, validity must be rejected in one of two cases. A systematic inappropriate influence on measured WTA seems to be present in at least one experiment. This is likely experiment A, where respondents were far less familiar with proposed alternatives than in experiment B. In conclusion, measuring preferences for major, little-known innovations in a reliable way seems to present particular challenges for experimental research.

7 Appendix

Table 6 Estimation results for the linear model, Experiment A

Variable	Coefficient		Std. err.	z value	P > |z|
PHYSCOST	-0.9085349	***	0.054660	-16.62	0.000
PHYSQUAL	-0.4691062	***	0.052087	-9.01	0.000
PHYSEFF	-0.3691041	***	0.053688	-6.87	0.000
INNOVATION	-0.5686612	***	0.038207	-14.88	0.000
GENERICS	-0.0235289		0.048101	-0.49	0.625
MINOR	0.0569549		0.046406	1.23	0.220
NURSING	-0.219003	***	0.038198	-5.73	0.000
HOSPITAL	-0.3281199	***	0.037891	-8.66	0.000
PREMIUM	-0.008797	***	0.000983	-8.95	0.000
CONSTANT	-0.5124295	***	0.079204	-6.47	0.000
σ_u	1.052211		0.039998		
ρ	0.5254248	***	0.018957		
Number of observations = 9850					
χ^2 (9) = 573.65; Prob > χ^2 = 0.0000					
Likelihood ratio test of ρ = 0: χ^2 (1) = 1487.86; Prob > = χ^2 = 0.000					

(**, ***) Coefficients different from zero with error probabilities of 5% (1%, 0.1%).

Table 7 Estimation results for the linear model, Experiment B

Variable	Coefficient		Std. err.	z value	P > |z|
DEDUCT [1)]	-0.000565	***	0.0000303	-18.64	0.000
DEDUCT2	3.80e-08	***	7.88e-09	4.82	0.000
COPAYMENT	-0.270581	***	0.0428204	-6.32	0.000
ALTMED [2)]	0.353708	***	0.0419784	8.43	0.000
GENERICS	-0.197095	***	0.0421067	-4.68	0.000
INNOVATION	-0.549415	***	0.0480711	-11.43	0.000
PREMIUM	-0.014312	***	0.0006055	-23.63	0.000
CONSTANT	-0.000565	***	0.0000303	-18.64	0.000
σ_u	0.902244		0.040346		
ρ	0.448746	***	0.022124		

Number of observations = 9569

χ^2 (7) = 1296.2 > χ^2 = 0.0000

Likelihood ratio test of ρ = 0: χ^2 (1) = 745.05 > = χ^2 = 0.000

(**, ***) Coefficients different from zero with with error probabilities of 5% (1%, 0.1%).

[1)] Compensation in Swiss francs required for a 1CHF increase in the annual deductible.

[2)] Expanded coverage of alternative medicine (status quo=no inclusion).

8 References

1. Ryan M, Gerard K. Using discrete choice experiments to value health care programmes: current practice and future reflections. Appl Health Econ Health Policy 2003; 2(1): 55–64.

2. Hanley N, Ryan M, Wright R. Estimating the monetary value of health care: lessons from environmental economics. Health Econ 2003; 12: 3–16.

3. Scanlon DP, Chernew ME, Lave JR. Consumer health plan choice. Annu Rev Public Health 1997; 18: 507–28.

4. Ryan M. A comparison of stated preference methods for estimating monetary values. Health Econ 2004; 13: 291–296.

5. Gyrd-Hansen D, Søgaard J. Analysing public preferences for cancer screening programmes. Health Econ 2001; 10: 617–634.

6. Merino-Castellò A. Demand for pharmaceutical drugs: a choice modelling experiment. Working Paper. University of Barcelona; 2003.

7. Ryan M, Wordsworth S. Sensitivity of willingness to pay estimates to the level of attributes in discrete choice experiments. Scott J Polit Econ 2000; 47: 504–524.

8. San Miguel F, Ryan M, McIntosh E. Applying conjoint analysis in economic evaluations: an application to menorrhagia. Appl Econ 2000; 32: 823–833.

9. Ryan M, McIntosh E, Shackley P. Methodological issues in the application of conjoint analysis in health care. Health Econ 1998; 7: 373–378.

10. Ryan M, Hughes J. Using conjoint analysis to assess women's preferences for miscarriage management. Health Econ 1997; 6: 261–273.

11. Scott A, Vick S. Patients, doctors and contracts: an application of principal-agent theory to the doctor-patient relationship. Scott J Polit Econ 1999; 46: 111–134.

12. Slothuus-Skoldborg U, Gyrd-Hansen D. Conjoint analysis – the cost variable: an Achilles' heel? Health Econ 2003; 12(6): 479–497.

13. Louviere JL, Hensher DA, Swait J. Stated choice methods: analysis and applications. Cambridge, Mass.: Cambridge University Press; 2000.

14. DeShazo JR, Fermo G. Designing choice sets for stated preference methods: the effects of complexity on choice consistency. J Environ Econ Manage 2002; 44: 123–143.

15. Luce DR. Individual Choice Behavior. New York: Wiley and Sons; 1959.

16. Manski C, Lerman SR. The estimation of choice probabilities from choice based samples, Econometrica 1977; 45(8): 1977–1988.

17. McFadden D. Economic choices. Am Econ Rev 2001; 91(3): 351–378.

18. Lancaster K. Consumer demand: a new approach. New York: Columbia University Press; 1971.

19. Ben-Akiva M, Lerman SR. Discrete choice analysis, Cambridge, Mass.: The MIT Press; 1985.

20. Jöreskog K, Goldberger AS. Estimation of a model with multiple indicators and multiple causes of a single latent variable, J Am Stat Assoc 1975; 70: 631–39.

21. Schoenberg R, Arminger G. Linear covariance structures version 2.0, user guide, RJS Software, Kensington, Maryland; 1989.

22. Litwin MS. How to Measure Survey Reliability and Validity, Thousand Oaks, London, New Delhi: SAGE Publications; 1995.

23. Bryan S, Gold L, Sheldon R, Buxton M. Preference measurement using conjoint methods: an empirical investigation of reliability, Health Econ 2000; 9: 385–395.

24. Cairns J, van der Pol M. Repeated follow-up as a method for reducing non-trading behaviour in discrete choice experiments. Soc Sci Med 2004; 58: 2211–2218.

25. Farrar S, Ryan M. Response-ordering effects: a methodological issue in conjoint analysis. Health Econ 1999; 8: 75–79.

26. Verlegh PWJ, Schifferstein HNJ, Wittink DR. Range and number-of-levels effects in derived and stated measures of attribute importance. Marketing Letters 2002; 13(1): 41–52.

27. Ryan M, Bate A. Testing the assumptions of rationality, continuity and symmetry when applying discrete choice experiments in health care. Appl Econ Letters 2001; 8: 59–63.

28. Scott A. Eliciting GP's preferences for pecuniary and non-pecuniary job characteristics. J Health Econ 2001; 20: 329–347.

29. Scott A. Identifying and analysing dominant preferences in discrete choice experiments: an application in health care. J Econ Psychol 2002; 23: 383–398.

30. Lloyd A. Threats to the estimation of benefit: are preference elicitation methods accurate? Health Econ 2003; 12: 393–402.

31. San Miguel F, Ryan M, Amaya-Amaya M. 'Irrational' stated preferences: a quantitative and qualitative investigation. Health Econ 2005; 14(3): 307–322.

32. Telser H, Zweifel P. Measuring willingness-to-pay for risk reduction: an application of conjoint analysis. Health Econ 2002; 11(3): 129–139.

33. Hausman JA; editor. Contingent Valuation – A Critical Assessment, Amsterdam, London, New York, Tokyo: North-Holland; 1993.

34. Nocera S, Bonato D, Telser H. The contingency of contingent valuation: what are people willing to pay against Alzheimer's Disease? Int J Health Care Finance Econ 2002; 2: 219–40.

35. Telser H, Zweifel P. Validity of Discrete-Choice Experiments: evidence for health risk reduction, Appl Econ 2007; 39(1): 69–78.

36. Kuhfeld WF, Tobias RD, Garratt M. Efficient experimental design with marketing research applications. J Mark Res 1994; 31: 545–557.

37. Hardin RH, Sloane NJA. A new approach to the construction of optimal designs. J Stat Plan Inference 1993; 37: 229–369.

38. Hardin RH, Sloane NJA. Operating manual for Gosset: a general purpose program for constructing experimental designs, 2nd ed. Murray Hill, NJ: AT&T Bell Laboratories; 1994.

39. Hedayat AS, Sloane NJA, Stufken J. Orthogonal srrays: theory and applications, New York, Berlin, Heidelberg: Springer; 1999.

40. Ryan M. Methodological issues in the application of conjoint analysis in health care. Health Econ 1998; 7: 373–378.

41. Ryan M, San Miguel F. Revisiting the axiom of completeness in health care. Health Econ 2003; 12: 295–307.

42. Maddala T, Phillips KA, Johnson RF. An experiment on simplifying conjoint analysis designs for measuring preferences. Health Econ 2003; 12: 1035–1047.

43. Johnson RF, Ozdemir S, Hauber AB. Motivating out-of-pocket treatment costs with cheap talk. Working Paper. Research Triangle Institute; 2007.

44. Cook J, Whittington D, Canh DG, Johnson RF, Nyamete A. Reliability of stated preferences for cholera and typhoid vaccines with time to think in Hue, Vietnam. Econ Inq 2007; 45(1): 100–114.

45. Samuelson W, Zeckhauser RJ. Status quo bias in decision making. J Risk Uncertain 1988; 1: 7–59.

46. Horowitz JK; McConnell KE. A review of WTA/WTP studies. J Environ Econ Manage 2002; 44: 426–447.

47. Zweifel P, Telser H, Vaterlaus S. Consumer resistance against regulation: the case of health care, Journal of Regulatory Economics 2006; 29(3): 319-332

48. Andreoni J. Giving with impure altruism: applications to charity and ricardian equivalence, J Polit Econ 1989; 97(6): 1447–1458.

49. Andreoni J. Warm-glow versus cold-prickle: the effects of positive and negative framing on cooperation in experiments, Q J Econ 1995; 110(1): 1–21.

50. Johnson RF, Desvouges WH. Estimating stated preferences with rated pair data: environmental, health, and employment effects of energy programs. J Environ Econ Manage 1997; 34: 79–99.

51. Shepard DS, Zeckhauser RJ. Survival and consumption. Manage Sci 1994; 30(4): 423–439.

52. Becker K, Zweifel P. Age and choice in health insurance: evidence from Switzerland. The Patient 2008; forthcoming.

53. Dellaert BGC, Brazell JD, Louviere JL. The effect of attribute variation on consumer choice. Marketing Letters 1999; 10(2): 139–147.

Working Papers of the Socioeconomic Institute at the University of Zurich

The Working Papers of the Socioeconomic Institute can be downloaded from http://www.soi.uzh.ch/research/wp_en.html

0801 Managed Care Konzepte und Lösungsansätze – Ein internationaler Vergleich aus schweizerischer Sicht, Johannes Schoder, Peter Zweifel, February 2008, 23 p.

0719 Why Bayes Rules: A Note on Bayesian vs. Classical Inference in Regime Switching Models, Dennis Gärtner, December 2007, 8 p.

0718 Monoplistic Screening under Learning by Doing, Dennis Gärtner, December 2007, 29 p.

0717 An analysis of the Swiss vote on the use of genetically modified crops, Felix Schläpfer, November 2007, 23 p.

0716 The relation between competition and innovation – Why is it such a mess? Armin Schmutzler, November 2007, 26 p.

0715 Contingent Valuation: A New Perspective, Felix Schläpfer, November 2007, 32 p.

0714 Competition and Innovation: An Experimental Investigation, Dario Sacco, October 2007, 36p.

0713 Hedonic Adaptation to Living Standards and the Hidden Cost of Parental Income, Stefan Boes, Kevin Staub, Rainer Winkelmann, October 2007, 18p.

0712 Competitive Politics, Simplified Heuristics, and Preferences for Public Goods, Felix Schläpfer, Marcel Schmitt, Anna Roschewitz, September 2007, 40p.

0711 Self-Reinforcing Market Dominance, Daniel Halbheer, Ernst Fehr, Lorenz Goette, Armin Schmutzler, August 2007, 34p.

0710 The Role of Landscape Amenities in Regional Development: A Survey of Migration, Regional Economic and Hedonic Pricing Studies, Fabian Waltert, Felix Schläpfer, August 2007, 34p.

0709 Nonparametric Analysis of Treatment Effects in Ordered Response Models, Stefan Boes, July 2007, 42p.

0708 Rationality on the Rise: Why Relative Risk Aversion Increases with Stake Size, Helga Fehr-Duda, Adrian Bruhin, Thomas F. Epper, Renate Schubert, July 2007, 30p.

0707 I'm not fat, just too short for my weight – Family Child Care and Obesity in Germany, Philippe Mahler, May 2007, 27p.

0706 Does Globalization Create Superstars?, Hans Gersbach, Armin Schmutzler, April 2007, 23p.

0705 Risk and Rationality: Uncovering Heterogeneity in Probability Distortion, Adrian Bruhin, Helga Fehr-Duda, and Thomas F. Epper, July 2007, 29p.

0704 Count Data Models with Unobserved Heterogeneity: An Empirical Likelihood Approach, Stefan Boes, March 2007, 26p.

0703 Risk and Rationality: The Effect of Incidental Mood on Probability Weighting, Helga Fehr, Thomas Epper, Adrian Bruhin, Renate Schubert, February 2007, 27p.

0702 Happiness Functions with Preference Interdependence and Heterogeneity: The Case of Altruism within the Family, Adrian Bruhin, Rainer Winkelmann, February 2007, 20p.

0701 On the Geographic and Cultural Determinants of Bankruptcy, Stefan Buehler, Christian Kaiser, Franz Jaeger, June 2007, 35p.

0610 A Product-Market Theory of Industry-Specific Training, Hans Gersbach, Armin Schmutzler , November 2006, 28p.

0609 Entry in liberalized railway markets: The German experience,
Rafael Lalive, Armin Schmutzler, April 2007, 20p.

0608 The Effects of Competition in Investment Games,
Dario Sacco, Armin Schmutzler, April 2007, 22p.

0607 Merger Negotiations and Ex-Post Regret,
Dennis Gärtner, Armin Schmutzler, September 2006, 28p.

0606 Foreign Direct Investment and R&D offshoring,
Hans Gersbach, Armin Schmutzler, June 2006, 34p.

0605 The Effect of Income on Positive and Negative Subjective Well-Being,
Stefan Boes, Rainer Winkelmann, May 2006, 23p.

0604 Correlated Risks: A Conflict of Interest Between Insurers and Consumers and Its Resolution,
Patrick Eugster, Peter Zweifel, April 2006, 23p.

0603 The Apple Falls Increasingly Far: Parent-Child Correlation in Schooling and the Growth of Post-Secondary Education in Switzerland,
Sandra Hanslin, Rainer Winkelmann, March 2006, 24p.

0602 Efficient Electricity Portfolios for Switzerland and the United States,
Boris Krey, Peter Zweifel, February 2006, 25p.

0601 Ain't no puzzle anymore: Comparative statics and experimental economics,
Armin Schmutzler, December 2006, 45p.

0514 Money Illusion Under Test,
Stefan Boes, Markus Lipp, Rainer Winkelmann, November 2005, 7p.

0513 Cost Sharing in Health Insurance: An Instrument for Risk Selection?
Karolin Becker, Peter Zweifel, November 2005, 45p.

0512 Single Motherhood and (Un)Equal EducationalOpportunities: Evidence for Germany,
Philippe Mahler, Rainer Winkelmann, September 2005, 23p.

0511 Exploring the Effects of Competition for Railway Markets,
Rafael Lalive, Armin Schmutzler, April 2007, 33p.

0510 The Impact of Aging on Future Healthcare Expenditure;
Lukas Steinmann, Harry Telser, Peter Zweifel, December 2006, 23p.

0509 The Purpose and Limits of Social Health Insurance;
Peter Zweifel, September 2005, 28p.

0508 Switching Costs, Firm Size, and Market Structure;
Simon Loertscher, Yves Schneider, August 2005, 29p.

0507 Ordered Response Models;
Stefan Boes, Rainer Winkelmann, March 2005, 21p.

0506 Merge or Fail? The Determinants of Mergers and Bankruptcies in Switzerland, 1995-2000; Stefan Buehler, Christian Kaiser, Franz Jaeger, March 2005, 18p.

0505 Consumer Resistance Against Regulation: The Case of Health Care
Peter Zweifel, Harry Telser, Stephan Vaterlaus, February 2005, 23p.

0504 A Structural Model of Demand for Apprentices
Samuel Mühlemann, Jürg Schweri, Rainer Winkelmann and Stefan C. Wolter, February 2005, 25p.

0503 What can happiness research tell us about altruism? Evidence from the German Socio-Economic Panel
Johannes Schwarze, Rainer Winkelmann, September 2005, 26p.

0502 Spatial Effects in Willingness-to-Pay: The Case of Two Nuclear Risks
Yves Schneider, Peter Zweifel, September 2007, 31p.

0501 On the Role of Access Charges Under Network Competition
Stefan Buehler, Armin Schmutzler, January 2005, 30p.