

Schmidt, Ulrich; Stolpe, Michael

**Working Paper**

## Transitivity in health utility measurement: An experimental analysis

Kiel Working Paper, No. 1412

**Provided in Cooperation with:**

Kiel Institute for the World Economy (IfW)

*Suggested Citation:* Schmidt, Ulrich; Stolpe, Michael (2008) : Transitivity in health utility measurement: An experimental analysis, Kiel Working Paper, No. 1412, Kiel Institute for the World Economy (IfW), Kiel

This Version is available at:

<http://hdl.handle.net/10419/4249>

**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.*



*Kiel*

## **Working Papers**

**Kiel Institute  
for the World Economy**



### **Transitivity in Health Utility Measurement: An Experimental Analysis**

**by Ulrich Schmidt and Michael Stolpe**

**No. 1412 | March 2008**

**Web: [www.ifw-kiel.de](http://www.ifw-kiel.de)**

Kiel Working Paper No. 1412 | March 2008

## **Transitivity in Health Utility Measurement: An Experimental Analysis**

Ulrich Schmidt and Michael Stolpe

### Abstract:

Several experimental studies observed substantial violations of transitivity in decisions between risky lotteries over monetary outcomes. The goal of our experiment is to test whether these violations also affect the evaluation of health states. A particular feature of our experimental design is that it takes into account the possible role of decision errors in generating violations of transitivity. Since we find neither substantial nor systematic deviations from transitive choice behaviour, we can conclude that previously reported violations do not seem to bias health utility measurement.

Keywords: Transitivity; Health utility; Errors

JEL classification: D81, I10, C91

### **Ulrich Schmidt**

Department of Economics  
Christian-Albrechts-Universität  
Olshausenstr. 40, 24118 Kiel, Germany  
Telephone: ++49 431 880-1400  
E-mail: uschmidt@bwl.uni-kiel.de

### **Michael Stolpe**

Kiel Institute for the World Economy  
24100 Kiel, Germany  
Telephone: ++49 431 8814-246  
E-mail: michael.stolpe@ifw-kiel.de

---

*The responsibility for the contents of the working papers rests with the author, not the Institute. Since working papers are of a preliminary nature, it may be useful to contact the author of a particular working paper about results or caveats before referring to, or quoting, a paper. Any comments on working papers should be sent directly to the author.*

*Coverphoto: uni\_com on photocase.com*

## 1 Introduction

Health utility measurement plays an important role in medical decision making, in particular in cost-effectiveness analyses of alternative treatments. A central assumption in health utility measurement is that preferences satisfy transitivity. Transitivity demands that whenever option A is preferred to option B and B is preferred to C, then also A has to be preferred to C. In the absence of transitivity a well-defined utility function  $U$  (i.e. for all options A and B we have  $U(A) \geq U(B)$  if and only if A is weakly preferred to B) does not exist. Consequently, standard methods in health utility measurement, such as the time tradeoff method or quality adjusted life years, cannot be meaningfully applied in the absence of transitivity. Several empirical studies observed substantial violations of transitivity, in particular for choice between risky options (e.g. Tversky, 1969, Loomes, Starmer, and Sugden, 1989, 1991, Loomes and Taylor, 1992, Starmer and Sugden, 1998, Starmer, 1999, Birnbaum, Patton, and Lott, 1999, and Humphrey, 2001). The validity of health utility measurement would be seriously challenged if these violations carry over to the evaluation of health states. However, since the evaluation of health states involves consequences composed of several attributes (i.e. at least health status and life duration), behaviour may be fundamentally different than in the above mentioned studies which are based on monetary consequences.

To the best of our knowledge, it has not been analyzed before whether transitivity empirically holds for the evaluation of health states. However, several studies observed preference reversals in health utility measurement (e.g. Stalmeier, Wakker, and Bezembinder, 1997, Robinson, Dolan, and Williams, 1997, and Bleichrodt and Pinto, 2007) which are closely related to violations of transitivity. A preference reversal occurs if a risky option  $p$  is preferred to another risky option  $q$  (i.e.  $p \succ q$ ) in a straight choice but a higher certainty equivalent (CE) is assigned to  $q$  (i.e.  $CE(q) > CE(p)$ ). Assume such a reversal and consider an amount  $z$  with  $CE(q) > z > CE(p)$ . Since the certainty equivalent of an option should be indifferent to that option we get  $q \sim CE(q) \succ z$ ,  $z \succ CE(p) \sim p$ , and  $p \succ q$  from choice. Hence we have an intransitive cycle  $p \succ q \succ z \succ p$ , i.e. an intransitive preference cycle. Precisely this result has been utilized by many studies mentioned above (in particular Loomes, Starmer, and Sugden, 1989, 1991, Starmer and Sugden, 1998, and Humphrey, 2001) to derive from the preference reversal literature experimental designs where substantial violation rates of transitivity have been observed. Consequently, the evidence of preference reversals in health utility measurement seems to indicate that also violations of transitivity may occur in the evaluation of health states.

There is, however, some debate concerning the evidence of violations of transitivity (Iverson and Falmagne, 1985, Stevenson, Busemeyer, and Naylor, 1991, Sopher and Gigliotti, 1993, Luce, 2000, Iverson, Myung, and Karabatsos, 2006, Regenwetter and Stober, 2006, and Birnbaum and Schmidt, 2007, 2008). In particular, it has been argued that the observed violations are not “true” violations since they may simply be caused by random errors. Consequently, we develop an experimental design which allows us to discriminate between true violations and violations caused by random error. Our design is presented in the next section where we also discuss the role of errors as potential causes of transitivity violations. Section 3 reports the results of our experiment and Section 4 offers some concluding observations.

## 2 Experimental Design and the Role of Errors

As most other studies on transitivity, our experiment was conducted as a classroom experiment. We ran two studies with undergraduate economics students at the University of Kiel, a pretest with 40 subjects and a main test with 98 subjects. In both studies subjects had to make six pairwise choices which were composed of two series with three alternatives each. Alternatives were described as in Figure 1 which presents the choice between alternatives A and B of series I.

*Suppose you have a serious disease which leads to immediate death if uncured. Two alternative treatments are available:*

*Under treatment A you live for a further 15 years and have serious migraine on one day per week.*

*Under treatment B you die immediately with a probability of 33%. With a probability of 67% you live for a further 20 years and have no migraine.*

*Which treatment do you choose?*

**Figure 1:** Presentation of alternatives

Each alternative is characterized by three attributes: (i) the probability of immediate death, (ii) the number of additional life years in case of survival, and (iii) the number of days with migraine per week. The corresponding values of our design are presented in Table 1. In the main test, we revised the attribute values of some alternatives (i.e. B and C in Series I and C in Series II) used in the pretest. In these cases, the attribute values of the pretest appear in

parentheses. In both tests, attribute values of the alternatives were chosen such that the majority rule implies intransitive cycles. The majority rule implies that an alternative A is preferred to another alternative, B, if the number of attributes for which A is better than B exceeds the number of attributes for which A is worse, i.e.

$$(1) \quad A \succ B \Leftrightarrow \#\{i: A_i \succ B_i\} > \#\{i: B_i \succ A_i\},$$

where  $A_i$  and  $B_i$  denote the attribute values of A and B. It is easy to see that for the values in Table 1 we have for Series I according to the majority rule  $B \succ A$  (more years and less migraine),  $C \succ B$  (more years and lower probability of death) and  $A \succ C$  (lower probability of death and less migraine). Also for Series II the majority rule implies the cycle  $B \succ A \succ C \succ B$ .

**Table 1:** Experimental Design of Series I (left panel) and Series II (right panel)

	Prob (Death)	Years	Migraine		Prob (Death)	Years	Migraine
A	0	15	1	A	15	30	3
B	33 (35)	20	0	B	25	35	2
C	13 (10)	30	2	C	22 (20)	24 (25)	1

In the pretest all subjects received an identical booklet with all six choices, each choice on a separate sheet. In the main test there were two booklets, one with three choices (two from one series and one from the other series) at the beginning of class and a second booklet with the remaining three choices at the end of class. This procedure ruled out that subjects could make consistency checks. The main test also divided subjects into two groups. Compared to the first group, the order of the two booklets and the order of alternatives in the booklets were reversed in the second group. Since there are no significant differences in the results for both groups we do not distinguish them in the following. Finally, in order to motivate subjects for careful consideration of the questions, 12 subjects drawn randomly received a flat payment of 20 Euros in the main test.

In the studies of Loomes, Starmer, and Sugden (1989), (1991), Starmer and Sugden (1998), and Humphrey (2001) intransitivities are tested by testing for the so-called cycling asymmetry. With three choices, there exist two possible intransitive response cycles and the studies conclude that observed intransitivities are not simply noise if one cycle occurs with a significantly higher frequency than the other. The validity of this argument has been questioned by Sopher and Gigliotti (1993). Consider for instance the result for Case 2 of Starmer and Sugden (1998) presented in Table 2. The entries in the first row give the possible

choice patterns with which a subject could respond to the three questions. For example, ABA means that a subject chose A over B in the first choice, B over C in the second, and A over C in the third. Altogether, there are eight different possible response patterns out of which the last two (ABC and BCA) are intransitive. Suppose that all subjects have transitive preferences but they sometimes make errors when choosing between lotteries. It is easy to see that the cycle ABC can be caused by only one error if the true preferences are ABA, BBC, or ACC. Analogously, the cycle BCA is caused by one error if the true preferences are ACA, BBA, or BCC. According to Table 2, 18 subjects chose patterns ABA, BBC, or ACC while 69 subjects chose ACA, BBA, or BCC, i.e. the latter three patterns occur much more frequently. As the simplest case, we can assume that the probability of errors is the same for all subjects and independent of the underlying true preference pattern. This implies that errors will lead much more frequently to the cycle BCA than to the cycle ABC given that the true preferences are always transitive. The observed differences in those frequencies thus may not be taken as evidence of intransitivities.

**Table 2:** Results of Starmer and Sugden (1998)

	ABA	ACA	BBA	BBC	ACC	BCC	ABC	BCA
Case 2	5	2	12	9	4	45	3 (3.3 %)	10 (11.1 %)

To control for this problem, Birnbaum and Schmidt (2007, 2008) estimate an explicit error model using data from an identically repeated experiment. However, this procedure requires additional assumptions on the structure of the error model. We therefore follow a different route in this article: We simply adjust the attributes of the alternatives such that the number of subjects choosing patterns ABA, BBC, or ACC roughly equals the number of subjects choosing the other three transitive patterns ACA, BBA, or BCC. In this case, random errors would imply an approximately equal frequency of the two intransitive patterns so that we can use the cycling asymmetry to test for intransitivities.

### 3 Results

Table 3 presents the results of the pretest. The table shows that there are no intransitivities for Series II whereas 10% of subjects responded with intransitive cycles for Series I; yet both cycles occur with equal frequency and no cycling asymmetry can be observed. However, the main goal of the pretest was not to test for transitivity, but to help adjust the alternatives for the main test as outlined in the preceding section. The frequency of

patterns which can lead to the cycle ABC by one error, i.e. ABA, ACC, and BBC, is 16 for Series I and 9 for Series II. The frequency of the opposite patterns (BCC, BBA, and ACA) is given by 20 for Series I and 31 for Series II. These observations show that in particular Series II was unbalanced in the pretest, mainly due to the fact that alternative C is preferred too frequently. We therefore made the attributes of this alternative less attractive in the main test. For Series I, we also made alternative B slightly more attractive as indicated in Table 1.

**Table 3:** Results of the Pretest

	ABA	ACA	BBA	BBC	ACC	BCC	ABC	BCA
Series I	0	11	5	2	14	4	2 (5.0 %)	2 (5.0 %)
Series II	2	8	2	5	2	21	0	0

The results of the main test are presented in Table 4. We can verify that our alternatives are rather balanced as the frequency of patterns BCC, BBA, and ACA (52 for Series I and 47 for Series II) is relatively close to the frequency of patterns BCC, BBA, and ACA (40 for Series I and 43 for Series II). The incidence of intransitivities is rather limited compared to previous studies which reported substantial violations of transitivity. In our study only 6.1% of subjects violated transitivity in Series I and 8.1% of subjects in Series II.

**Table 4:** Results of the Main Test

	ABA	ACA	BBA	BBC	ACC	BCC	ABC	BCA
Series I	9	15	12	14	29	13	5 (5.1 %)	1 (1.0 %)
Series II	20	6	7	20	7	30	6 (6.1 %)	2 (2.0 %)

In both series, the cycle ABC occurs more frequently than the cycle BCA. However, according to a one-tailed binominal test (which was also used by Starmer and Sugden, 1998), this difference is not statistically significant, so we can not conclude that a cycling asymmetry exists. Since the rarely observed cycle BCA is the one implied by the majority rule for our design, this rule appears to perform rather poorly in our experiment.

## 4 Conclusions

This paper has presented an experiment aimed at testing transitivity in health utility assessment. A particular feature of our design is the use of a balanced set of alternatives such that an asymmetric frequency of the two intransitive cycles is unlikely to result from random



errors. In contrast to many previous experimental studies, we find neither a substantial frequency of subjects violating transitivity nor a significant cycling asymmetry. Further research may be required to understand whether the absence of the intransitivity problem is mainly due to the control for errors in our design or due to the fact that our alternatives are not lotteries over monetary amounts as in previous studies but composed of health states. Altogether our results suggest that practical applications of health utility measurement may not be substantially biased by violations of transitivity. This conclusion is in line with other recent experimental studies (Birnbbaum and Gutierrez, 2007; Birnbbaum and Schmidt, 2007, 2008) that test for transitivity while controlling for the possible role of errors.

## References

- Birnbaum, M. H., and Gutierrez, R. J. (2007). Testing for intransitivity of preferences predicted by a lexicographic semiorder. Organizational Behavior and Human Decision Processes 104, 97-112.
- Birnbaum, M. H., Patton, J. N., and Lott, M. K. (1999). Evidence against rank-dependent utility theories: Violations of cumulative independence, interval independence, stochastic dominance, and transitivity. Organizational Behavior and Human Decision Processes 77, 44-83.
- Birnbaum, M. H., and Schmidt, U. (2007). Testing transitivity in choice under risk. Working paper, available upon request.
- Birnbaum, M., and Schmidt, U. (2008). An experimental investigation of violations of transitivity in choice under uncertainty. Kiel Working Paper, 1396, Institut für Weltwirtschaft, Kiel.
- Bleichrodt, H., and Pinto, J. L. (2007). A new preference reversal in health utility measurement. Working paper, University of Rotterdam.
- Humphrey, S. J. (2001). Non-transitive choice: Event-splitting effects or framing effects? Economica 68, 77-96.
- Iverson, G., and Falmagne, J.-C. (1985). Statistical issues in measurement. Mathematical Social Sciences 10, 131-153.
- Iverson, G. J., Myung, J. I., and Karabatsos, G. (2006). Intransitivity of preference: Revisited. Working paper, available from Geoffrey J. Iverson, Institute for Mathematical Behavioral Sciences, UCI, Irvine, CA 92697.
- Loomes, G., Starmer, C., and Sugden, R. (1989). Preference reversal: information-processing effect or rational non-transitive choice? Economic Journal 99, 140-151.
- Loomes, G., Starmer, C., and Sugden, R. (1991). Observing violations of transitivity by experimental methods. Econometrica 59, 425-440.
- Loomes, G., and Taylor, C. (1992). Non-transitive preferences over gains and losses. Economic Journal 102, 357-365.
- Luce, R. D. (2000). Utility of gains and losses: Measurement-theoretical and experimental approaches. Mahwah, NJ: Lawrence Erlbaum Associates.
- Myung, J. I., Karabatsos, G., and Iverson, G. J. (2005). A Bayesian approach to testing decision making axioms. Journal of Mathematical Psychology 49, 205-225.
- Regenwetter, M., and Stober, C. (2006). Testing transitivity by testing the triangle inequality instead of weak stochastic transitivity. Edwards Bayesian Research Conference,

Fullerton, CA, January, 2006.

- Robinson, A., Dolan P., Williams A. (1997). Valuing health states using VAS and TTO: What lies behind the numbers? Social Science and Medicine 45, 1289-1297.
- Sopher, B., and Gigliotti, G. (1993). Intransitive cycles: Rational Choice or random error? An answer based on estimation of error rates with experimental data. Theory and Decision 35, 311-336.
- Stalmeier, P. F. M., Wakker P. P., and Bezembinder T. G. G. (1997). Preference reversals: Violations of unidimensional procedure invariance. Journal of Experimental Psychology: Human Perception and Performance 23, 1196-1205.
- Starmer, C. (1999). Cycling with rules of thumb: An experimental test for a new form of non-transitive behaviour. Theory and Decision 46, 141-158.
- Starmer, C., and Sugden, R. (1998). Testing alternative explanations of cyclical choices. Economica 65, 347-361.
- Stevenson, M. K., Busemeyer, J. R., and Naylor, J. C. (1991). Judgment and decision-making theory. In M. Dunnette & L. M. Hough (Eds.), New handbook of industrial-organizational psychology (pp. 283-374). Palo Alto, CA: Consulting Psychologist Press.
- Tversky, A. (1969). Intransitivity of preferences. Psychological Review 76, 31-48.