University of Southern Denmark · Syddansk Universitet

CONJOINT RELIABILITY INVESTIGATED AT THE INPUT AND OUTPUT LEVEL

By

Ulla Slothuus Skjoldborg , Ph.D., Jørgen Lauridsen, Ph.D.,
Peter Junker ,MD.

Health Economics Paper 2005:7

Foreword

Discrete choice analysis is one of the key research topics at the Health Economics research unit. The present study investigates the issue of conjoint reliability. The was initiated while Ulla Slothuus Skjoldborg was working as an associate professor at the Institute of Public Health.
Financial support by the Danish Medical Research Council is acknowledged.

Terkel Christiansen
Professor, Head of Health Economics research unit.

1. Abstract

Methodological issues in conjoint measurement still need to be solved. This paper investigates the issue of conjoint reliability. A discrete choice experiment was applied using scenarios that describe the effect of treating rheumatoid arthritis patients with biological drugs, i.e., TNF-alpha inhibitor treatment. A high level of conjoint reliability at both input and output levels was demonstrated.

KEY WORDS: conjoint measurement, discrete choice experiment, reliability, rheumatoid arthritis, TNF-alpha inhibitor treatment

2. Introduction

Conjoint analysis (CA) could be used to elicit preferences and willingness to pay (WTP) [1,2,3], by having the respondent rate, rank or choose between hypothetical alternatives. Still new to the health care field the method has increasingly been applied in the last number of years which makes it important to focus on the unresolved methodological issues related to CA. Among others such a methodological issue is reliability, which will be investigated in this paper.

CA is based on random utility theory, where the utility of the good being evaluated is considered as being composed of a deterministic part, interpreted as an indirect utility function and a random part, which is assumed to be composed of factors - not observable – which influence utility. This random element measures errors in the dependent variable and/or model specification errors.

Assume a linear additive utility function, where the utility of the good is a function of all characteristics:

$$U = V + \varepsilon = \sum \beta_i A_i + \varepsilon$$

V is the cumulative utility function and β_i (i=1,....,n) is a vector of coefficients related to the good characteristics A_i.

When applying CA hypothetical scenarios are presented to the respondent, each describing different levels of the attributes that characterize the good being evaluated. The respondents' preferences are measured by asking them to state which alternative they prefer. Its possible to determine the attributes relative importance (i.e., the marginal rate of substitution (MRS)) when giving the attributes different values. The total explained utility (V) for different combinations of attributes can further be estimated, thereby determining the combination the respondent prefers the most. Willingness to pay (WTP) estimates can be calculated by including a cost attribute.

This paper investigated the issue of reliability in CA. Reliability was defined as consistency of results, meaning that a reliable measure had no variation in the observed score due to random errors [4]. Reliability could be considered in different forms [4,5]. Here we focused on reliability over time, involving that a measurement would be repeated using the same instrument and the same respondents at least two times with separate time intervals. To the authors' knowledge this has only been tried once before in a health care setting [6]. The present study, however, used a data material with a much longer time interval between measurements and a set of hypothetical alternatives that seem more complex than in the before mentioned study.

In the paper, first we considered reliability from an input variable point of view, considering how consistent the respondents' answers to the CA-questions were over time.

Secondly, reliability was investigated by looking at the outcome/ result of the CA analysis, i.e, we determined the consistency of the different attribute weights over time. The scenarios applied to the conjoint measurements were in the field of rheumatoid arthritis (RA) therapy with particular respect to therapy with biological agents (TNF-alpha inhibitors) which have shown to be highly effective in patients which have not benefited from traditional disease modifying anti-rheumatic drugs [7]. Of note, however, this novel therapeutic modality is expensive, amounting to around 20.000 USD per year for each individual patient.

3. Method

3.1. Setting

A total of 325 patients diagnosed with rheumatoid arthritis (i.e., rheumatoid arthritis according to the 1987 ACR classification criteria), who received therapy at the outpatient clinic at Odense University Hospital, Section of Rheumatology, were asked to participate in the study. The outpatient clinic is a tertiary center, serving the County of Funen. The 325 patients were the total number of diagnosed RA patients who were registered in the outpatient clinics database as of July 2003, and were between 18 and 70 years of age. The reason for limiting the age range was that the interviews would involve quite abstract concepts. All patients received a letter of introduction in which the study was described. Subsequently the patients were contacted by phone regarding participation. 178 agreed to participate.

3.2. Choice of attributes

The attributes included in the description of the treatment scenarios are presented in Table 1. Selection of attributes was based on the criteria that each attribute described a relevant issue of the effect of TNF-alpha inhibitor treatment. Levels of attributes have been taken from the literature and been validated through interviews with rheumatologists. Concerning the cost attribute, levels from an earlier study investigating RA patients' willingness to pay for RA-treatment [8] was applied since this chosen range showed that the maximum levels of payment were adequately high to ensure that maximum willingness-to-pay estimates were derived.

Tabel 1.: The different attribute values used in the hospital model

Attribute no.	Attributes	*Hospital A*	*Hospital B*
1	Duration of morning stiffness	0;5;30;60;90;120	0;5;30;60;90;120
2	Pain level	0;2;4;6;8;10	0;2;4;6;8;10
3	Number of swollen joints	0;5;10;15;20;25	0;5;10;15;20;25
4	Feeling of being tired	reduced(0); unchanged (1)	reduced(0); unchanged (1)
5	Slightly higher risk of a minor	yes (1); no (0)	yes (1); no (0)
6	Out–of-pocket payment per month in excess of present expenditure for arthritis medication (DKK)	0,50,100,200,450,575, 800,900,1075,1150,12 50,1500,2150,2300,25 00,3000,4300,5000	0,50,100,200,450,575, 800,900,1075,1150,12 50,1500,2150, 2300, 2500,3000,4300,5000

Attributes listed in Table 1 were included as explanatory variables in a random effect logit model.

3.3. Selection of scenarios

Given the number of attributes and the number of possible outcomes per attribute, the total number of possible combinations was exceedingly high, necessitating a systematic reduction in number of scenarios applied. Such a reduction in the number of scenarios was accomplished by establishing a 'fractional factorial design', assuming interactions among attributes to be insignificant. The PLAN procedure from the computer package SAS was used for this purpose.

To reduce the number of choices faced by the respondent, a block design was used. There were 8 subgroups, which were tested to be homogenous with respect to age, gender, and duration of illness of the respondent (please see Appendix A for details).

Studies [9] have shown that respondents are capable of handling up to 13 discrete choice

questions per interview. In this study, each interview comprised eight such questions.

3.4. Analytical model

A linear additive utility function was assumed, i.e., a rise in the value of one attribute would give a proportional rise or proportional fall in total utility. Further, it was assumed that the utility associated with one attribute was not affected by the utility experienced from another attribute. A basic model describing the utility associated with the effect of a given TNF-alpha inhibitor treatment relative to an alternative option was therefore described as:

$$\Delta U = \beta_1 * \Delta x_1 + \beta_2 * \Delta x_2 + \beta_3 * \Delta x_3 + \beta_4 * \Delta x_4 + \beta_5 * \Delta x_5 + \beta_6 * \Delta x_6 + \varepsilon + \mu$$

where six attributes were included as explanatory variables. $\Delta x_1, \ldots, \Delta x_6$ represents the differences in attribute values between alternative A and alternative B, $\beta_1, \ldots, \beta_6$ are the attribute specific weights, and ΔU the change in utility as a result of choosing alternative B instead of alternative A. The error term is the random error term, including random variation across discrete choices, and $\mu\bullet$ is the random variation across respondents.

In the equation above the utility of alternative A was defined to be zero, which implied that $\Delta U > 0$ if B generated higher utility than A, and $\Delta U < 0$ if B generated lower utility. It was assumed that the individual would choose alternative B only if $\Delta U > 0$.

A discrete choice experiment (DCE) was applied. Respondents were asked to perform eight pair wise choices. Since correlation might exist across the discrete choices made by one individual, we controlled for random individual effects. The analysis was based on a random effect logit model. Interaction variables between survey time and respondent characteristics were created and included in the analysis. Non-linear second degree on all the continuity variables were further tested to make sure the model would be complete.

Variances were calculated using the Krinsky-Robb method. 95% confident intervals for WTP for attribute 1-6 were created as a function of individual characteristics. This was done by having the individual characteristic vary over the sample range (against the remaining characteristics on a sample average). For each value of the individual characteristic 10,000 Krinsky-Robb replications were made.

3.5. The interviews

The respondents participated in three face-to-face interviews in which they were asked to select between various effects of treatments with TNF-alpha inhibitors. The time interval between each interview was 4 months. In the first interview in addition to the clinical questions and the DCE questions, questions concerning background information (socioeconomic variables) about the respondents were asked. These questions were asked at the end of the interview to ensure that there were no differences in the three interviews concerning the timing of the DCE questions.

4. Results

178 respondents participated in the first survey. A response rate of 55% was attained. Out of the 178 respondents, 145 and 130 participated in the second and third survey, respectively.

The attributes and their relative weights for the three surveys can be seen in the Appendix B, Table B1. A few socio-demographic variables, the EuroQol estimate (TTO for future reference) and variables describing the extend of inconvenience associated with having arthritis, were in addition to the income variable included in the model. As seen tin Table B1 most variables were significant, and appeared to have influence on the choice of card A or B.

4.1. Reliability at the input level

To asses the reliability at the input data level, the consistency of matches made by respondents to the DCE question between replications was determined.

Of the 1661 choices made in survey 1, 1316 were repeated in survey 2. Table 2 presents a tabulation of these. The observed number of consistently repeated choices was (366+632) = 998, which was equivalent to (998/1316)*100% = 75.8%. The expected number by chance was (209+475) = 684, which was equivalent to (684/1316)*100% = 52.0%, thus a good correspondence between the choices in the two surveys was found. This was further confirmed by the highly significant chi-square statistics.

Tabel 2.: Tabulation of repeated choices in survey 1 and 2

		Survey 2		Total
		B	A	
Survey 1	B	366 (209)	164 (321)	530
	A	154 (311)	632 (476)	786
Total		520	796	1316
Chi-square= 324.0, prob<0.0001				

Note: Numbers in parentheses are expected number by chance

Of the 1661 choices made in survey 1, 1152 were repeated in survey 3. Table 3 presents a tabulation of these. The observed number of consistently repeated choices was 898, which was equivalent to 78.0%, while the expected number by chance was only 596 or 52.0%, indicating a good correspondence between the two surveys.

Table 3.: Tabulation of repeated choices in survey 1 and 3

		Survey 3		Total
		B	A	
Survey 1	B	341 (190)	119 (270)	460
	A	135 (286)	537 (406)	692
Total		476	676	1152
Chi-square= 340.0, prob<0.0001				

Note: Numbers in parentheses are expected number by chance

Of the 1429 choices made in survey 2, 1139 were repeated in survey 3. A tabulation of these were seen in Table 4. The observed number of consistently repeated choices was 892, which was equivalent to 78.3%, while the expected number by chance was only 594 or 52.2%, demonstrating a good correspondence between the two surveys.

Tabel 4.: Tabulation of repeated choices in survey 2 and 3

		Survey 3		Total
		B	A	
Survey 1	B	328 (179)	121 (270)	449
	A	126 (275)	564 (415)	690
Total		454	685	1139
Chi-square= 340.6, prob<0.0001				

Note: Numbers in parentheses are expected number by chance

Finally, of the 998 consistently repeated choices from survey 1 to survey 2, 818 were repeated in survey 3 (Table 5). The observed number of consistently repeated choices was 713, which was equivalent to 87.2%, while the expected number by chance was 437 or 53.4%, thus indicating a good correspondence between a consistent choice in survey 1/2 and the succeeding choice in survey 3.

Tabel 5.: Tabulation of repeated choices in survey 1-2 (those who were consistent and 3)

		Survey 3		Total
		B	A	
Survey 1	B	249 (111)	44 (182)	293
	A	61 (199)	464 (326)	525
Total		310	508	818
Chi-square= 430.1, prob<0.0001				

Note: Numbers in parentheses are expected number by chance

4.2. Reliability at the output level

To investigate reliability at the output/result level it was necessary to estimate the parameters in the conjoint model and performing Wald tests (as shown in Table 6). The Wald tests supplements the results in Table B1. Table B1 showed which interactions were significant, where the Wald tests shows whether there would be a considerable amount of interaction with the different attributes. The results were used to investigate the reliability of DCE on the output level:

Tabel 6.: Wald tests for significance of interactions

Variable	DF	Wald	Prob.
att1 interacts	26	76.08	<.0001
att2 interacts	26	109.34	<.0001
att3 interacts	26	61.81	<.0001
att4 interacts	26	49.53	0.0036
att5 interacts	26	34.03	0.1342
att6 interacts	26	52.34	0.0016
Survey 2	6	9.51	0.1471
Survey 3	6	5.36	0.4988
Survey 2 & survey 3	12	13.80	0.3136
Duration of illness	6	9.18	0.1639
Duration of illness2	6	12.76	0.0470
Duration of illness & Duration of illness2	12	42.31	<.0001
Reported degree of morning stiffness	6	10.02	0.1240
Reported degree of morning stiffness2	6	13.25	0.0393
Reported degree of morning stiffness & reported degree of morning stiffness2	12	45.47	<.0001
Reported degree of pain	6	5.21	0.5167
Reported degree of pain	6	7.39	0.2863
Reported degree of pain & reported degree of pain2	12	13.00	0.3692
Reported degree of swollen joints	6	4.38	0.6251
Reported degree of swollen joints2	6	5.26	0.5111
Reported degree of swollen joints & reported degree of swollen joints2	12	8.31	0.7604
Reported degree of tiredness	6	6.13	0.4089
Reported degree of tiredness2	6	4.62	0.5940
Reported degree of tiredness & reported degree of tiredness2	12	11.94	0.4508
Reported degree of adverse effects	6	5.77	0.4489
Reported degree of adverse effects2	6	7.88	0.2469
Reported degree of adverse effects & reported degree of adverse effects2	12	24.35	0.0182
Prescriptive drug	4	13.99	0.0073
TTO	6	12.20	0.0578
TTO2	6	9.48	0.1483
TTO and TTO2	12	15.49	0.2158
Birth cohort	6	13.79	0.0321
Birth cohort2	6	16.33	0.0121
Birth cohort & birth cohort2	12	25.20	0.0139

Gender	6	4.09	0.6650
Civil status	6	8.47	0.2055
Occupation_2	6	18.55	0.0050
Occupation_3	6	10.42	0.1081
Occupation_4	6	14.13	0.0282
Occupation (1,2 and 3)	18	49.34	<.0001
Income	6	9.44	0.1503
Income2	6	10.80	0.0947
Income & income2	12	17.19	0.1426

att1 refers to attribute no.1 in Table 1, att2 refers to attribute no. 2, etc.
duration of illness = the length of time the respondent has been diagnosed with arthritis
reported degree of morning stiffness = the respondents own valuation of experiencing morning stiffness on a scale from 0 to 10 (10 being the worst)
reported degree of pain= the respondents own valuation of experiencing pain on a scale from 0 to 10 (10 being the worst)
reported degree of swollen joints= the respondents own valuation of experiencing swollen joints on a scale from 0 to 10 (10 being the worst)
reported degree of tiredness= the respondents own valuation of experiencing tiredness on a scale from 0 to 10 (10 being the worst)
reported degree of adverse effects= the respondents own valuation of experiencing adverse effects on a scale from 0 to 10 (10 being the worst)
prescriptive drug= does the respondents have a monthly expenditure for prescriptive drugs (yes,no)
TTO= the EURO Qol estimate (Danish weights [10] have been used for calculation purposes)
birth cohort=the respondents year of birth
gender= the respondents gender (male=0; female=1)
civil status=the respondents civil status, (singel=0, married/cohab =1)
occupation=self-employed=1:public/private employed=2:retired=3:other non-employed=4
income= the respondents yearly income before tax (in 1000 DKK)

The Wald test, in Table 6, for 'Survey 2' was a test for the composed hypothesis (where att1 refers to attribute no.1 in Table 1)

H0 : ′att1survey2′ =0, ′att1survey3′=0, ′att1duration of illness′=0, ..., ′att1income2′=0.

The test result in Table 6 showed, that survey 2 did not differ from survey 1 with regard to the coefficients in the logistic regression. Looking at Table B1, except for att5survey2 which just kept its place on a 10 percent significant level, most of the interactions were not significant. Thus, with confidence we conclude that survey 2 did not differ from survey 1. The same conclusion appeared when looking at survey 1 and survey 3.

However, since WTP was a non-linear function of parameters, it was necessary to take a closer look at the confidence intervals before making any final conclusions.

Table 7.: WTP (1000 DDK) by survey, with Krinsky-Robb 95 Percent CI

Attribute	Survey	WTP	Lower	Upper
Morning Stiffness	1	0.00754	0.00380	0.01127
Morning Stiffness	2	0.00726	0.00279	0.01176
Morning Stiffness	3	0.00398	-0.00086	0.00881
Pain	1	0.22603	0.16495	0.28711
Pain	2	0.22219	0.15960	0.28477
Pain	3	0.23385	0.17047	0.29722
Swollen Joints	1	0.02872	0.00757	0.04986
Swollen Joints	2	0.01404	-0.00918	0.03727
Swollen Joints	3	0.01477	-0.00921	0.03876
Tiredness	1	0.82046	0.47343	1.16749
Tiredness	2	0.54162	0.14309	0.94016
Tiredness	3	0.34620	-0.01981	0.71220
Adverse Effects	1	0.69515	0.36738	1.02291
Adverse Effects	2	0.41129	0.07792	0.74465
Adverse Effects	3	0.49293	0.11934	0.86653

Table 7 shows that the confident intervals for the calculated WTP at survey time 1 and 2 did overlap, i.e., the two WTP values lied in each others confident intervals, which means, that the WTP value was the same at survey time 1 and 2. Looking at survey 3 we can further conclude that there was no difference between survey 1 and 3. Hence, overall, the results were constant over surveys 1, 2 and 3, i.e., the hypothesis (where att1 refers to attribute no.1 in Table 1, att2 refers to attribute no. 2, etc.)

H0: att1survey2=0, att2survey2=0, ..., att6survey2=0, att1survey3=0, att2survey3=0, ..., att6survey3=0

was accepted. In other words, the DCE was reliable on the output level.

4.3. Selected variables influence on the WTP value

Since we considered reliability at the output level, it was of interest to explore how the output, represented by the WTP value, was influenced by the variables selected in the DCE model.

The respondent's willingness to pay for a reduction in morning stiffness was higher if the respondent already has a monthly expenditure for prescriptive drugs (Table 8). The same was the case for reduced tiredness. For reduction in pain level and adverse effects it appeared that those respondents who did not have a monthly expenditure for prescriptive drugs had a higher WTP for a reduction in these attributes, than respondents with a monthly expenditure for prescriptive drugs. Regarding the attribute swollen joints, there was almost no difference between the two groups of respondents.

Table 8.: WTP (1000 DDK) by prescribtive drug, with Krinsky-Robb 95 percent CI

Attribute	Prescribtive Drug	WTP	Lower	Upper
MorningStiffness	No	0.00601	0.00287	0.00914
MorningStiffness	Yes	0.00757	-0.00123	0.01636
Pain	No	0.24515	0.20465	0.28564
Pain	Yes	0.03775	-0.06455	0.14005
Swollen Joints	No	0.01894	0.00441	0.03348
Swollen Joints	Yes	0.01844	-0.03202	0.06890
Tiredness	No	0.55731	0.30037	0.81425
Tiredness	Yes	0.59938	-0.11872	1.31748
Adverse Effects	No	0.54289	0.31363	0.77215
Adverse Effects	Yes	0.41837	-0.24591	1.08264

Note to table: prescriptive drug= does the respondents have a monthly expenditure for prescriptive drugs (yes,no)

Men tend to have a higher WTP for reduction in morning stiffness and pain level than women, whereas the opposite was the case when considering the attributes swollen joints, tiredness and adverse effects (Table 9). Interestingly respondents who were single appeared to have a lower WTP for every one of the attributes, compared to respondents who were married or living with someone (table 10).

Table 9.: WTP (1000 DDK) by gender, with Krinsky-Robb 95 percent CI

Attribute	Gender	WTP	Lo Lower	Upper
MorningStiffness	Male	0.00861	0.00387	0.01336
MorningStiffness	Female	0.00485	0.00172	0.00799
Pain	Male	0.23769	0.17899	0.29639
Pain	Female	0.22179	0.17290	0.27068
Swollen Joints	Male	0.01719	-0.00533	0.03972
Swollen Joints	Female	0.01991	0.00325	0.03658
Tiredness	Male	0.46490	0.11897	0.81084
Tiredness	Female	0.61272	0.34898	0.87646
Adverse Effects	Male	0.39181	0.03199	0.75164
Adverse Effects	Female	0.60900	0.34370	0.87431

Table 10.: WTP (1000 DDK) by civil status, with Krinsky-Robb 95 percent CI

	Civil status	WTP	Lower	Upper
Morning Stiffness	Single	0.00351	-0.00138	0.00841
Morning Stiffness	Married/ Cohab	0.00721	0.00336	0.01106
Pain	Single	0.17093	0.09374	0.24811
Pain	Married/ Cohab	0.24995	0.20140	0.29850
Swollen Joints	Single	0.02762	0.00180	0.05343
Swollen Joints	Married/ Cohab	0.01551	-0.00271	0.03374
Tiredness	Single	0.49286	0.04280	0.94291
Tiredness	Married/ Cohab	0.58719	0.30045	0.87393
Adverse Effects	Single	0.17030	-0.24637	0.58697
Adverse Effects	Married/ Cohab	0.67725	0.41346	0.94104

Concerning the individual characteristic occupation, compared to respondents who were self-employed, public or private employed or not employed, respondents who were retired had the highest WTP for a reduction in morning stiffness and swollen joints (table 11). Self-employed respondents had the highest WTP for reduction in pain and adverse effects, whereas not employed respondents willingness to pay for reduction in tiredness appeared to be the highest among occupation groups.

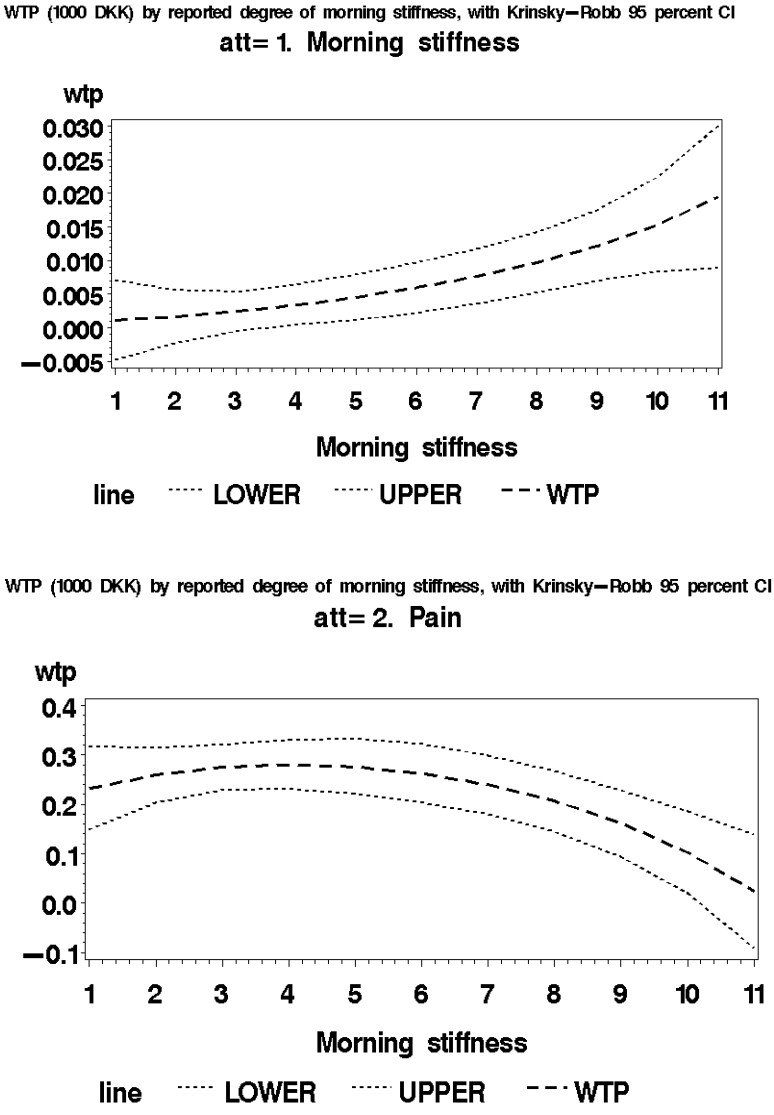
Table 11 WTP (1000 DKK) by occupation, with Krinsky-Robb 95 percent CI

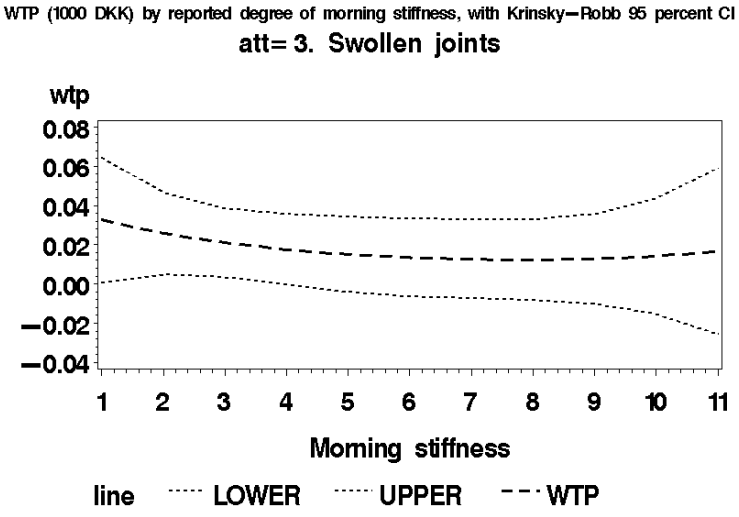
Attribute	Occupation	WTP	Lower	Upper
MorningStiffness	Public/private employed	-0.00003	-0.00626	0.00619
MorningStiffness	Self-employed	0.00624	-0.00628	0.01877
MorningStiffness	Retired	0.01026	0.00531	0.01520
MorningStiffness	Other non-employed	0.00139	-0.00532	0.00810
Pain	Public/private employed	0.32058	0.23223	0.40893
Pain	Self-employed	0.57785	0.35569	0.80002
Pain	Retired	0.15660	0.09751	0.21569
Pain	Other non-employed	0.14818	0.06713	0.22923
Swollen Joints	Public/private employed	-0.00594	-0.03619	0.02431
Swollen Joints	Self-employed	-0.01439	-0.07510	0.04633
Swollen Joints	Retired	0.04049	0.01031	0.07067
Swollen Joints	Other non-employed	-0.00131	-0.03088	0.02826
Tiredness	Public/private employed	0.39604	-0.14557	0.93764
Tiredness	Self-employed	0.42624	-0.62885	1.48133
Tiredness	Retired	0.60842	0.23798	0.97885
Tiredness	Other non-employed	0.80338	0.27150	1.33525
Adverse Effects	Public/private employed	0.55295	0.12058	0.98532
Adverse Effects	Self-employed	0.77003	-0.11092	1.65097
Adverse Effects	Retired	0.50084	0.12596	0.87572
Adverse Effects	Other non-employed	0.53390	0.00509	1.06272

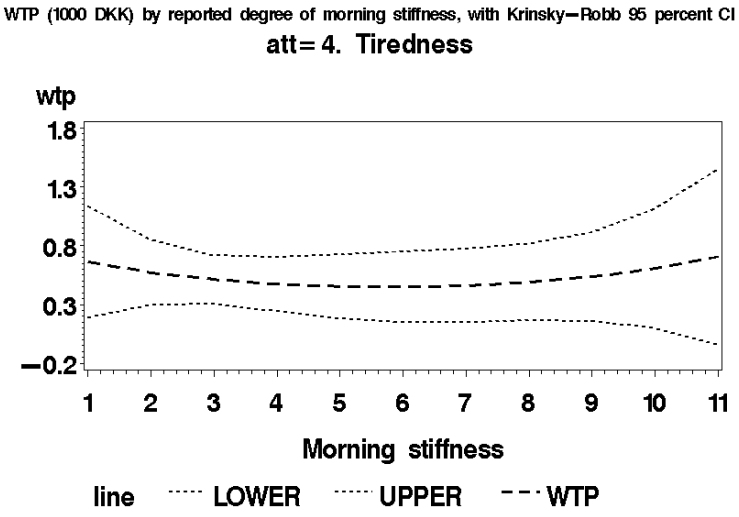
Figure 1-9 illustrates continuously variables selected in the DCE model and their influence on respondent's WTP.

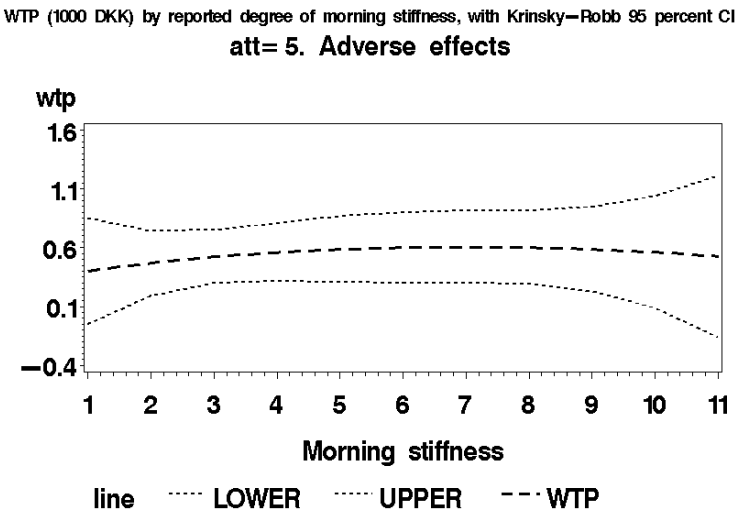
Figure 1 shows that the longer the respondent's duration of morning stiffness, the higher was the WTP for a reduction in morning stiffness and the lower was the WTP for a reduction in pain. WTP for swollen joints appeared also to decrease a little with increasing morning stiffness, whereas WTP for a reduction in tiredness or a reduction in adverse effects only slightly increased.

Figure 1.: WTP (1000 DKK) for attributes 1-5, by reported degree of morning stiffness (v16)




WTP (1000 DKK) by reported degree of morning stiffness, with Krinsky−Robb 95 percent CI
att= 3. Swollen joints
wtp
0.08
0.06
0.04
0.02
0.00
−0.02
−0.04
1 2 3 4 5 6 7 8 9 10 11
Morning stiffness
line LOWER UPPER WTP

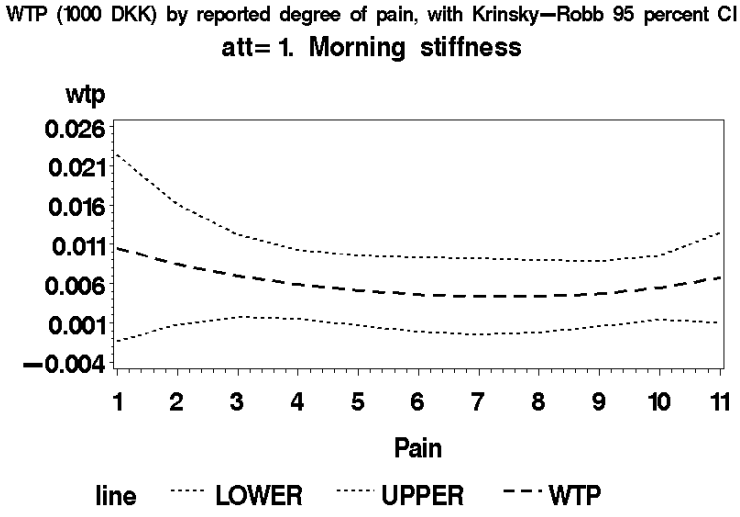

WTP (1000 DKK) by reported degree of morning stiffness, with Krinsky−Robb 95 percent CI
att= 4. Tiredness
wtp
1.8
1.3
0.8
0.3
−0.2
1 2 3 4 5 6 7 8 9 10 11
Morning stiffness
line LOWER UPPER WTP

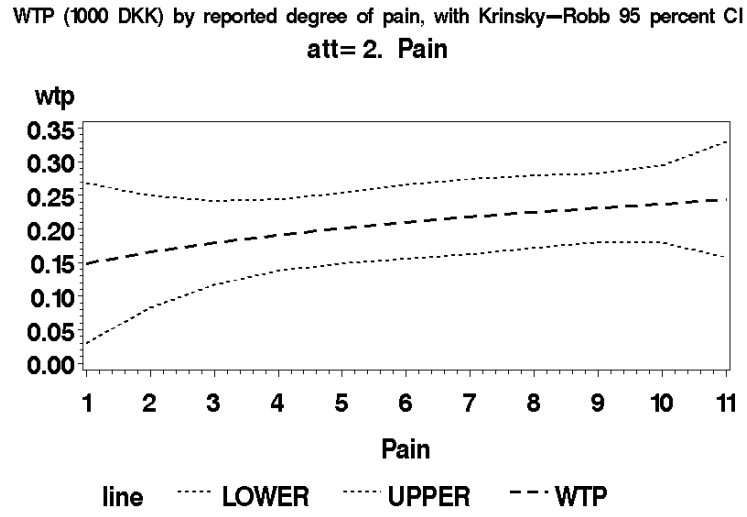


The more pain the respondent experienced the less he was willing to pay for a reduction in morning stiffness, however if the reported pain level was quite high the respondent's WTP increased again (figure 2).

More pain was associated with a higher WTP for pain reduction and also a slightly higher WTP for reduction in tiredness. Looking at the WTP for a reduction in swollen joints, up to a certain experienced pain level (about 6) the respondent's WTP increased, but decreased again after that point. The level of experienced pain seemed to have almost no influence on the WTP for reduction in adverse effects.

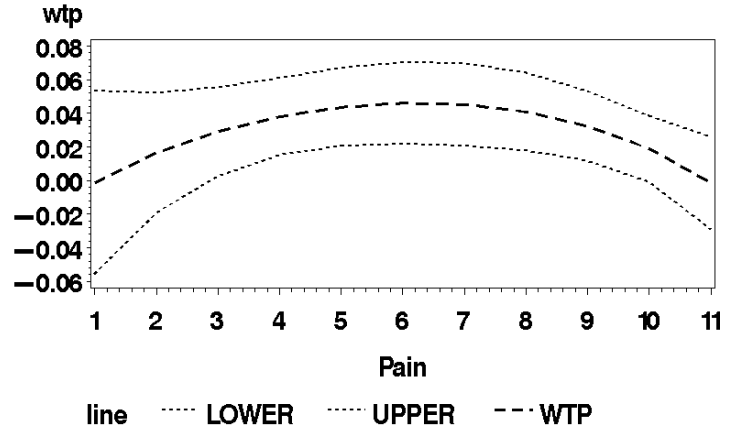
Figure 2.: WTP (1000 DKK) for attributes 1-5, by reported degree of pain (v17)





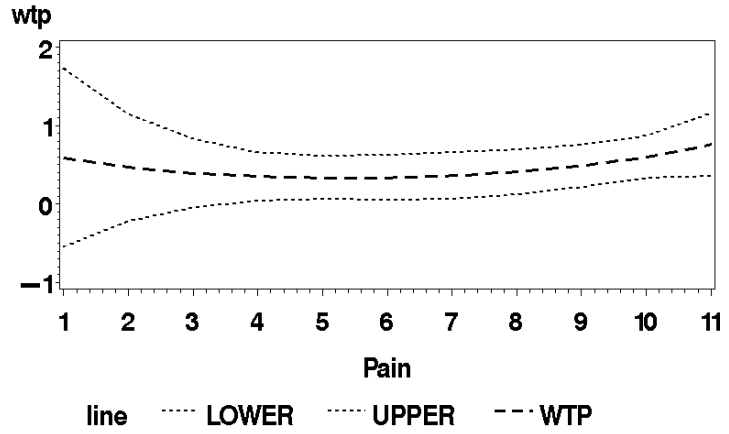
WTP (1000 DKK) by reported degree of pain, with Krinsky—Robb 95 percent CI

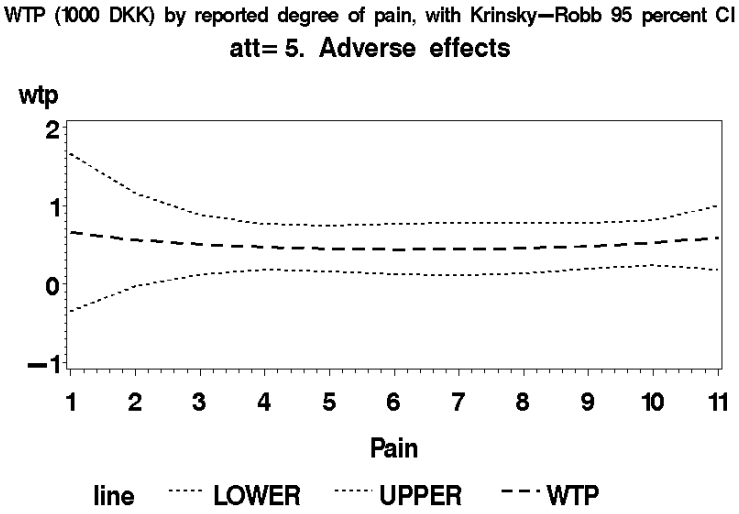
att= 3. Swollen joints



WTP (1000 DKK) by reported degree of pain, with Krinsky—Robb 95 percent CI

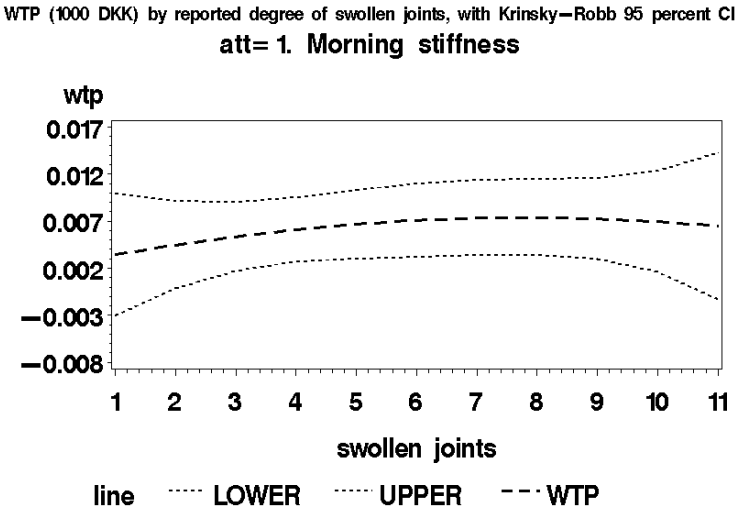
att= 4. Tiredness

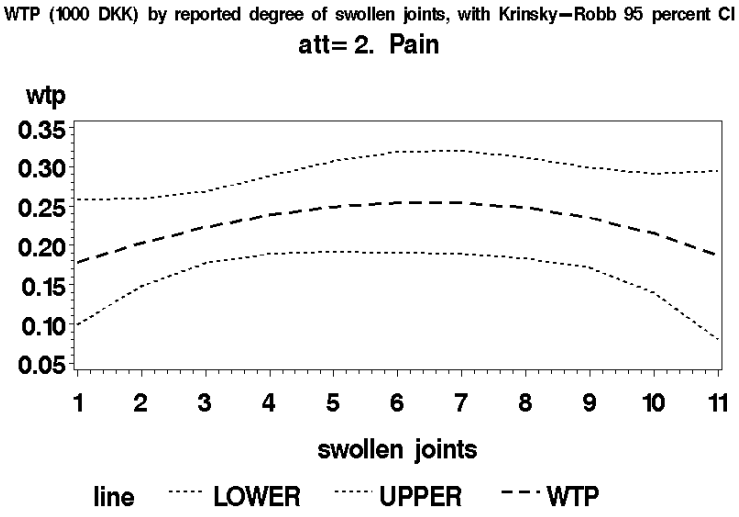




Regarding the influence of the number of swollen joints on the WTP value (figure 3), respondent's WTP increased slightly for a possible reduction in morning stiffness, swollen joints or tiredness the higher the number of swollen joints experienced. Looking at the WTP for a reduction in pain, up to a certain experienced number of swollen joints (about 7) the respondent's WTP increased, but decreased again after that point, the same pattern appeared for a reduction in adverse effects.

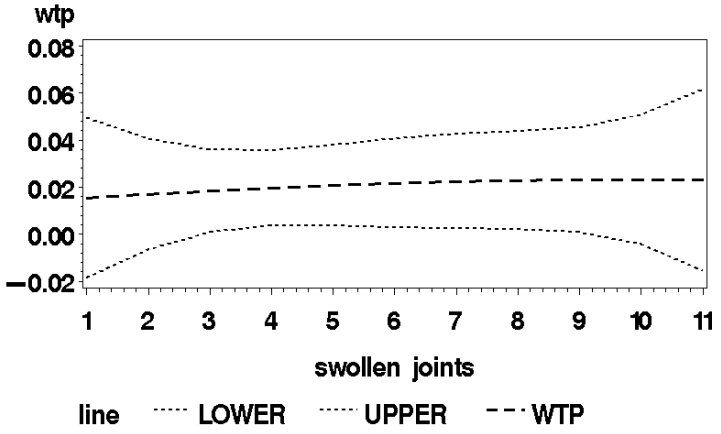
Figure 3.: WTP (1000 DDK) for attributes 1-5, by reported degree of swollen joint (v18)





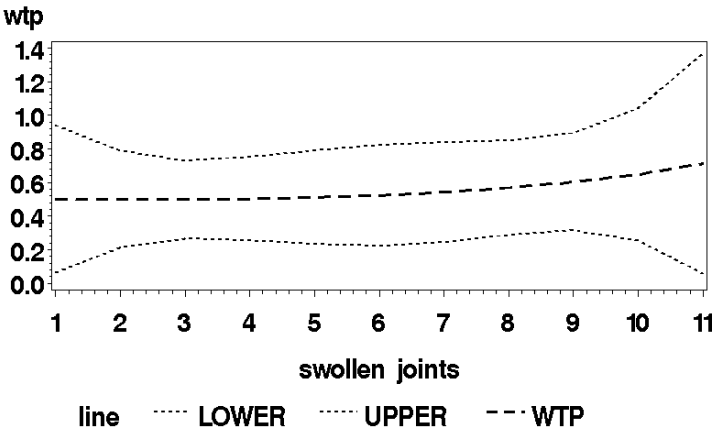
WTP (1000 DKK) by reported degree of swollen joints, with Krinsky−Robb 95 percent CI

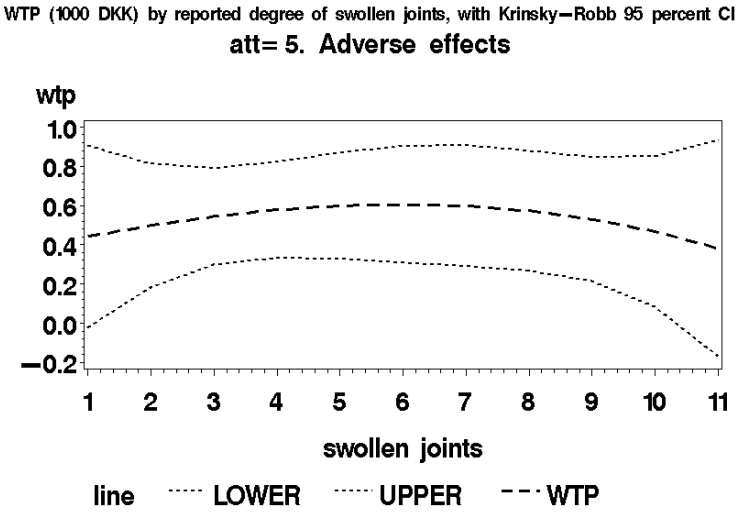
att= 3. Swollen joints



WTP (1000 DKK) by reported degree of swollen joints, with Krinsky−Robb 95 percent CI

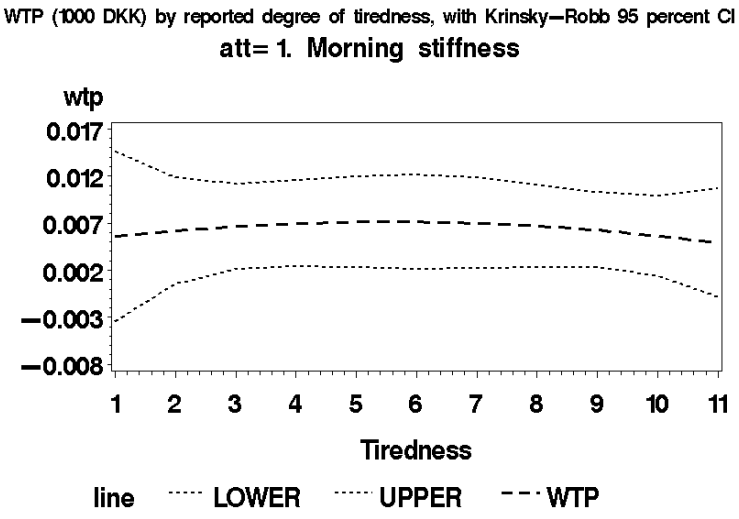
att= 4. Tiredness

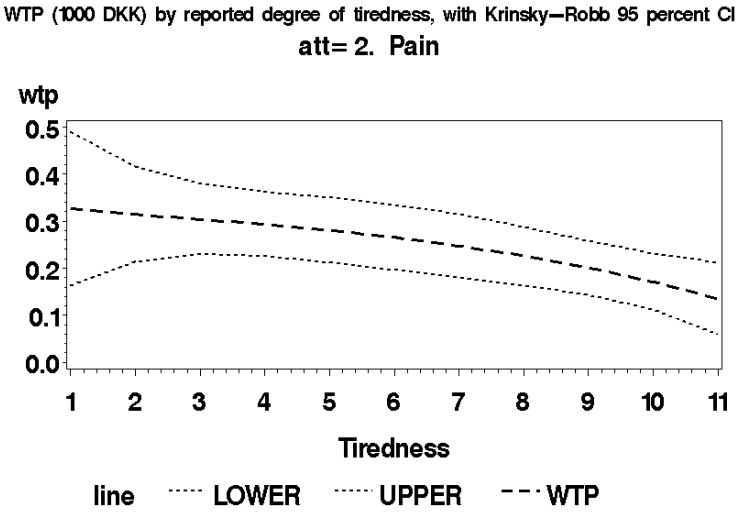




The WTP for a reduction in morning stiffness seemed almost unaffected by the level of tiredness the respondent experienced. Regarding adverse effects, there was a minor decrease in WTP when the respondent experienced low levels of tiredness, the WTP increased a little again with higher levels of experienced tiredness, but not much. The effect of a higher level of tiredness on the WTP for a pain reduction or a reduction in tiredness was however more dramatic. As illustrated in figure 4 the WTP for a reduction in pain decreased a lot as the level of experienced tiredness increased. The opposite was the case with WTP for a reduction in tiredness, which greatly increased with experienced tiredness.

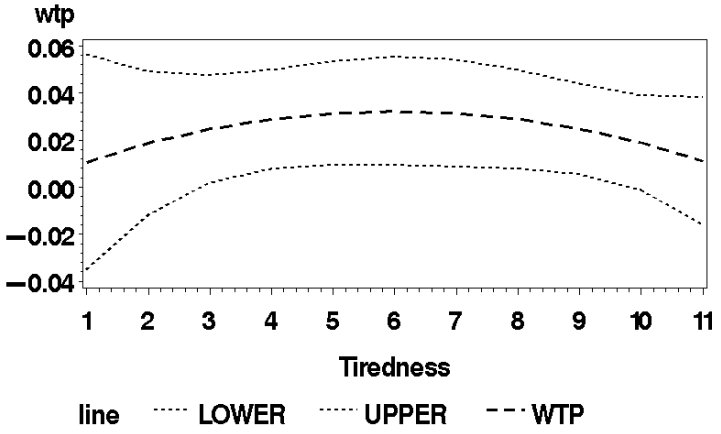
Figure 4.: WTP (1000 DDK) for attributes 1-5, by tiredness (v20)





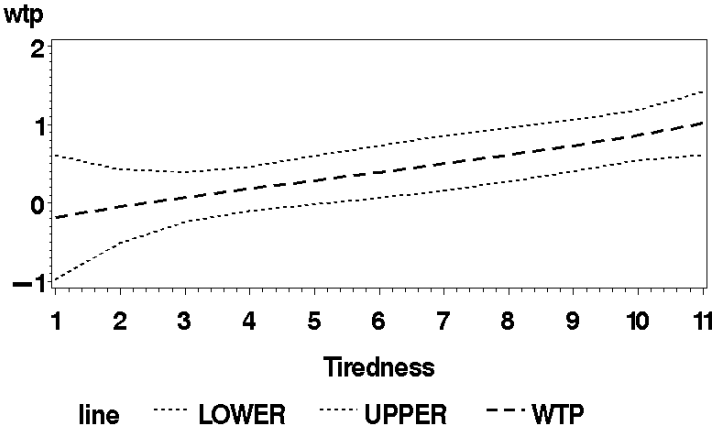
WTP (1000 DKK) by reported degree of tiredness, with Krinsky–Robb 95 percent CI

att= 3. Swollen joints



WTP (1000 DKK) by reported degree of tiredness, with Krinsky–Robb 95 percent CI

att= 4. Tiredness



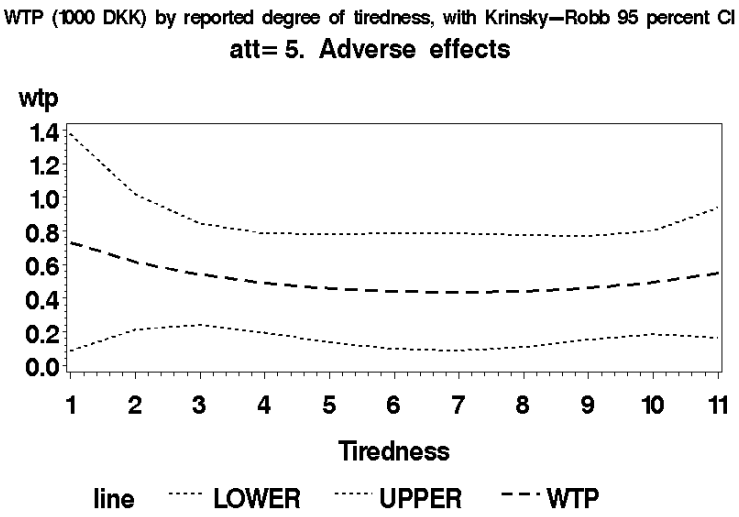
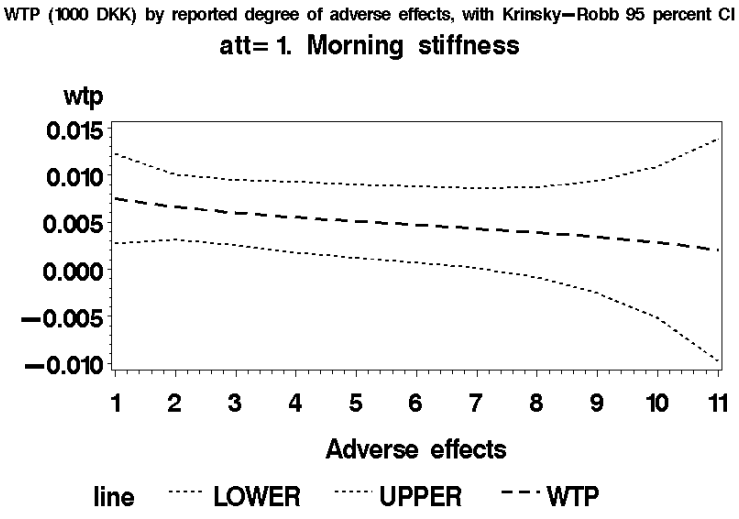
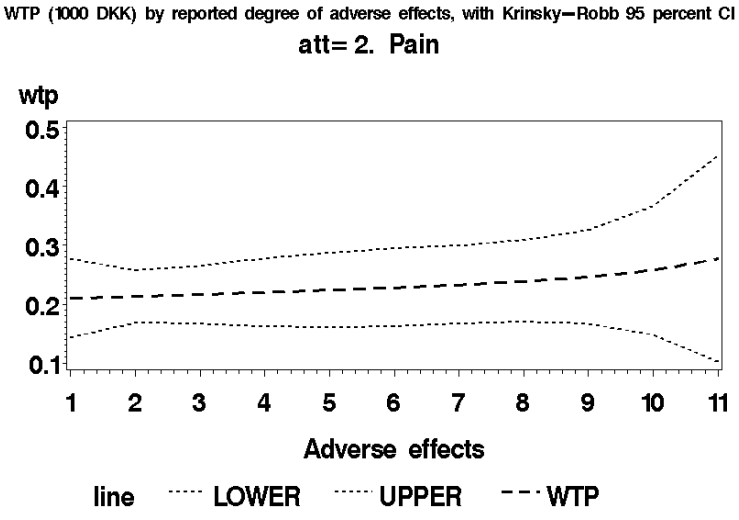
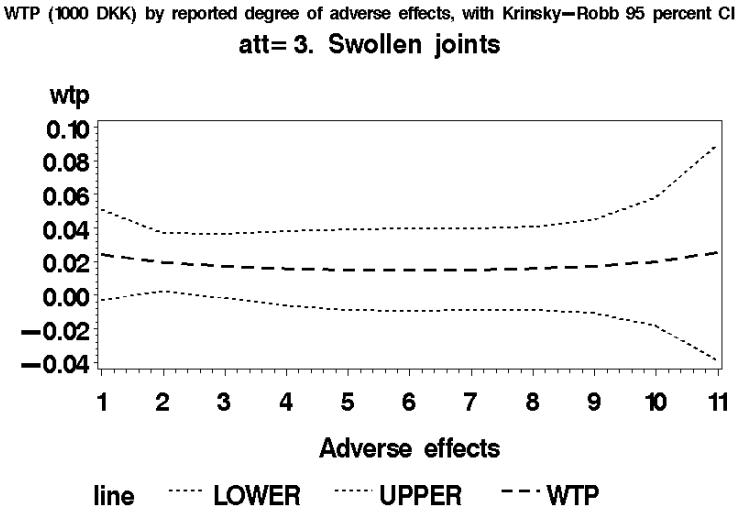


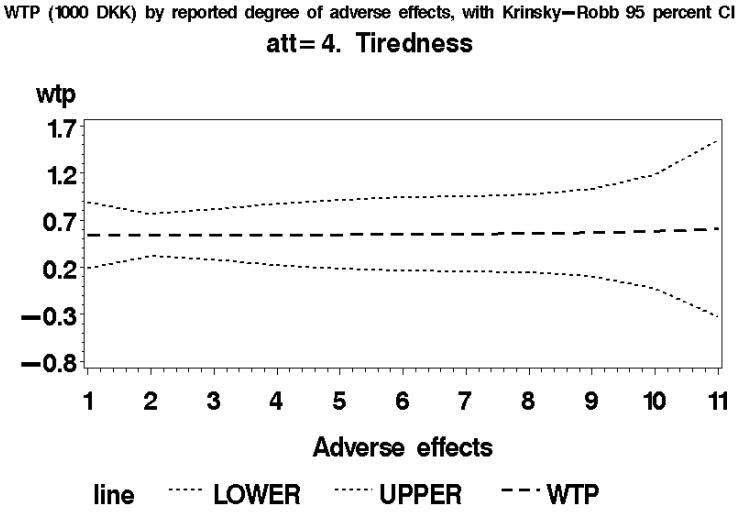
Figure 5 shows the WTP for reduction in attributes by degree of adverse effects. The degree of adverse effects did not seem to have a large impact on respondent's WTP. The largest influence appeared on attribute 1, morning stiffness, where a small decrease in WTP seemed to show as the extent of adverse effects increased. WTP for a reduction in pain level or a reduction in swollen joints increased a little with high levels of experienced adverse effects, whereas WTP for a reduction in adverse effects declined a little. WTP for a reduction in tiredness appeared almost unaffected by the degree of adverse effects.

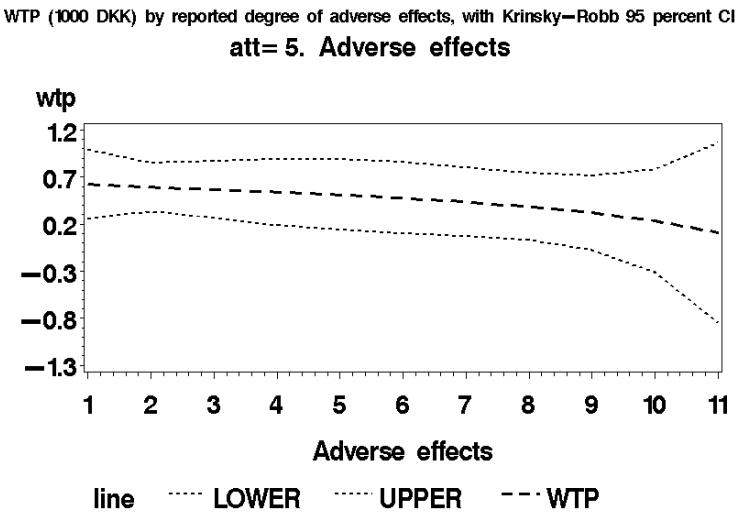
Figure 5. WTP (1000 DKK) for attributes 1-5, by reported degree of adverse effects (v19)





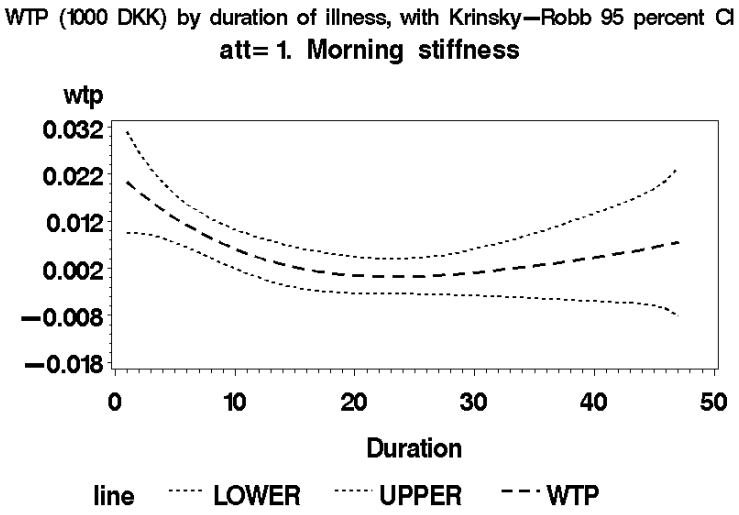

WTP (1000 DKK) by reported degree of adverse effects, with Krinsky−Robb 95 percent CI
att= 3. Swollen joints
wtp
0.10
0.08
0.06
0.04
0.02
0.00
−0.02
−0.04
1 2 3 4 5 6 7 8 9 10 11
Adverse effects
line LOWER UPPER WTP

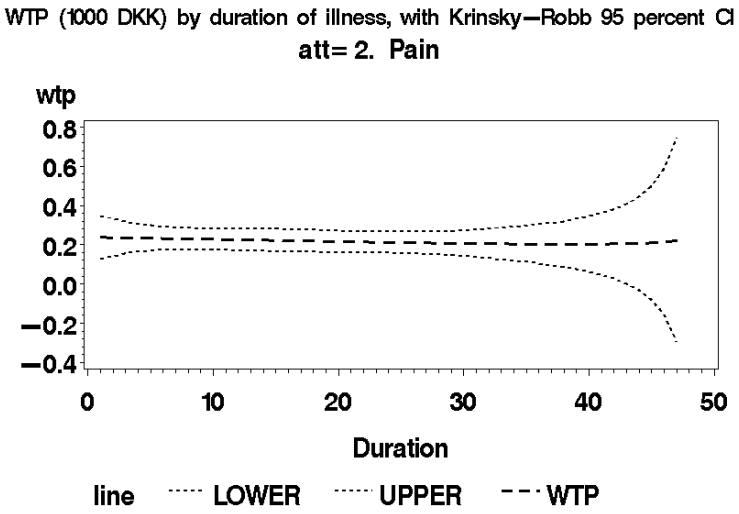

WTP (1000 DKK) by reported degree of adverse effects, with Krinsky−Robb 95 percent CI
att= 4. Tiredness
wtp
1.7
1.2
0.7
0.2
−0.3
−0.8
1 2 3 4 5 6 7 8 9 10 11
Adverse effects
line LOWER UPPER WTP



Looking at the effect of duration of illness (figure 6), WTP for a reduction in morning stiffness decreased until a duration of about 20 to 30 years where after it increased again. WTP for a reduction in pain appeared almost unchanged no matter how long the respondent had rheumatoid arthritis. A slightly fall in WTP for reduction in swollen joints or reduction in adverse effects was seen as duration of illness increased, whereas WTP for a reduction in tiredness increased with years suffering from rheumatoid arthritis.

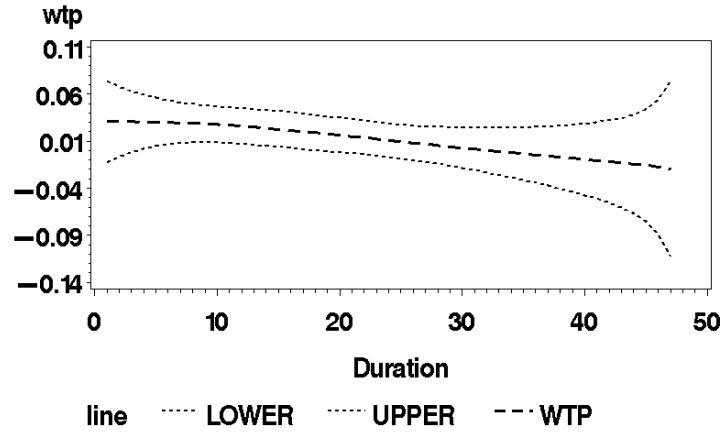
Figure 6.: WTP (1000 DKK) for attributes 1-5, by duration of illness (v5)





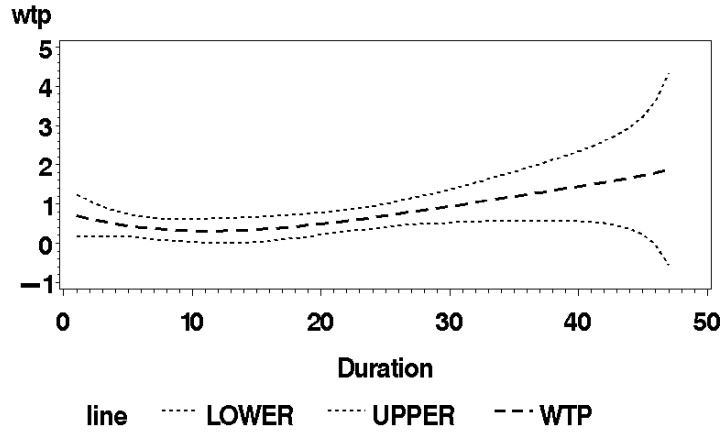
WTP (1000 DKK) by duration of illness, with Krinsky—Robb 95 percent CI

att= 3. Swollen joints



WTP (1000 DKK) by duration of illness, with Krinsky—Robb 95 percent CI

att= 4. Tiredness



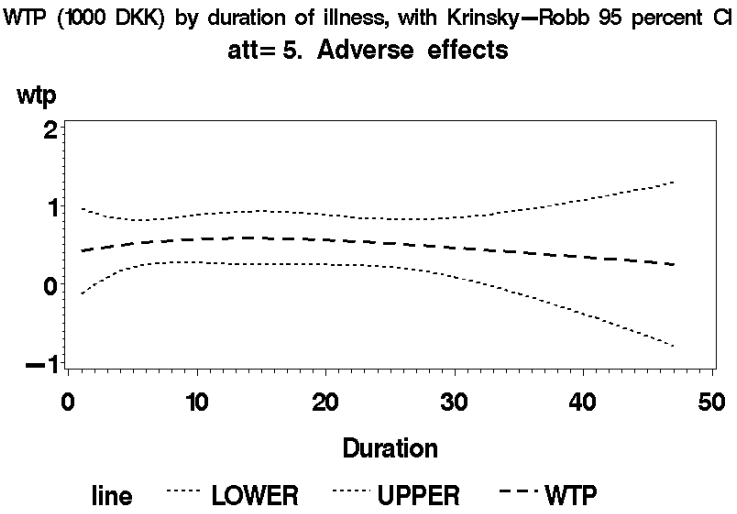
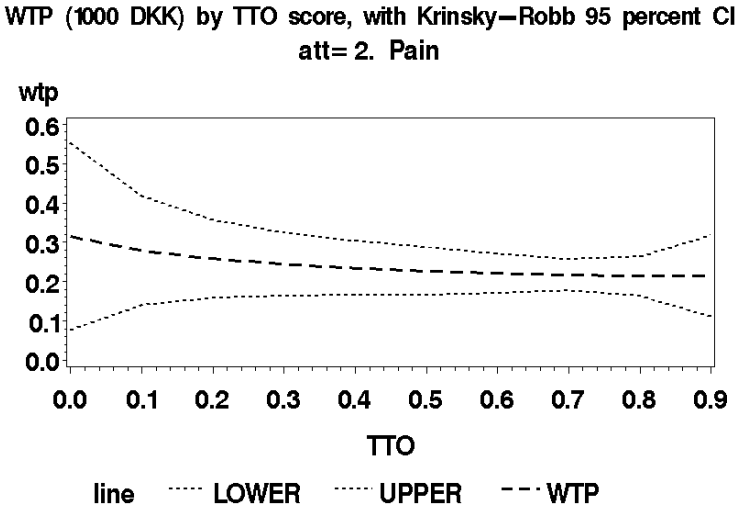
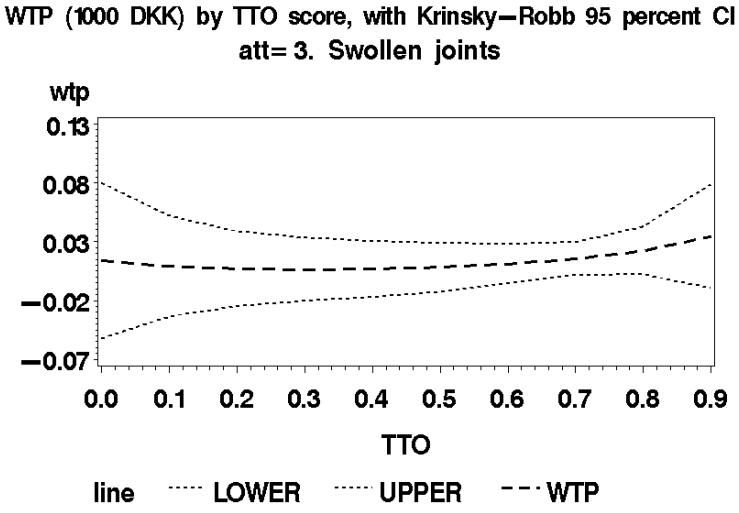
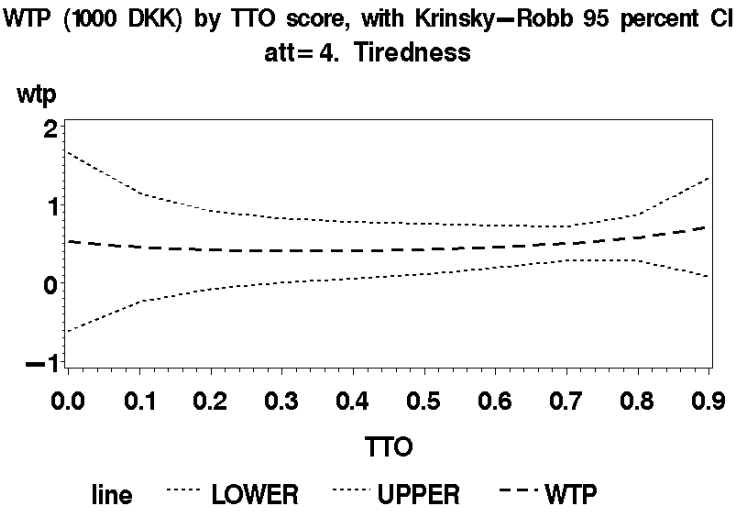


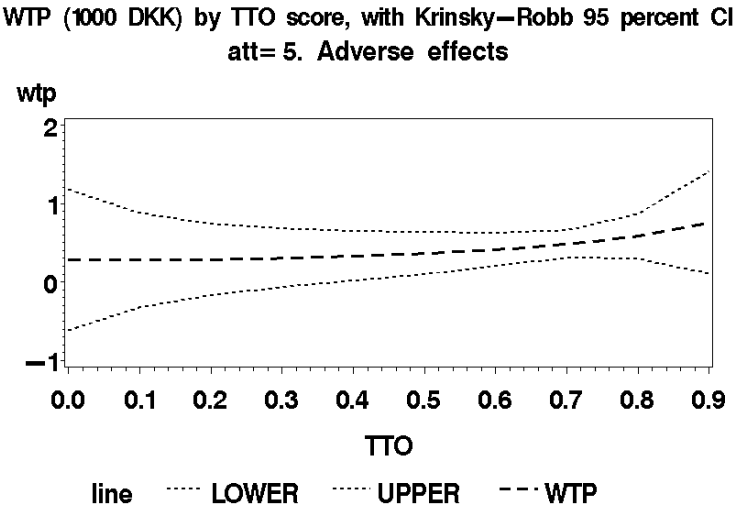
Figure 7 shows WTP for attributes by self/reported health (TTO score). As there were only 3 respondents with TTO scores below 0 and 5 respondents with TTO scores over 0.9, the confidence intervals were very large for these. To enhance interpretability, the figure therefore only shows WTP for the TTO interval from 0 to 0.9. For morning stiffness, there was a peak in WTP around a TTO value of 0.5, which indicated that persons with very poor or very good health had the highest WTP to reduce morning stiffness. For pain, the WTP was significantly positive and slightly falling with increased health. For swollen joints, tiredness and adverse effects, the WTP slightly increased with increasing health, but it was noticed that the WTP's for these three attributes were hardly significantly larger than zero throughout the range of TTO scores.

Figure 7.: WTP (1000 DKK) for attributes 1-5, by self-reported health (TTO scores)





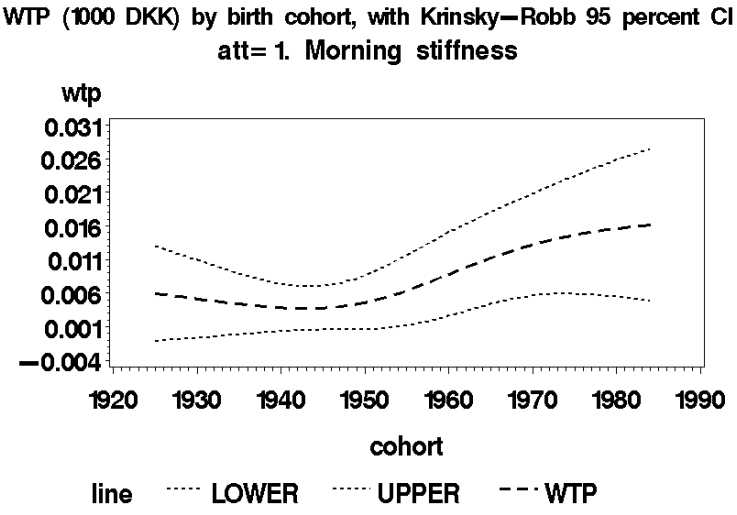


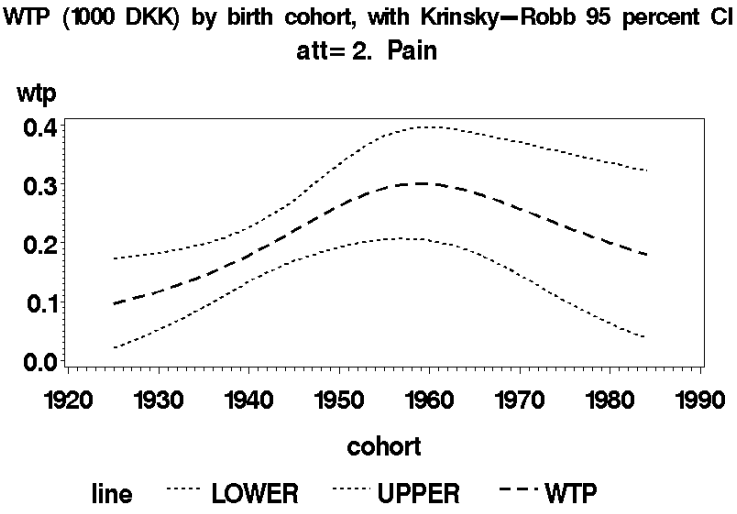


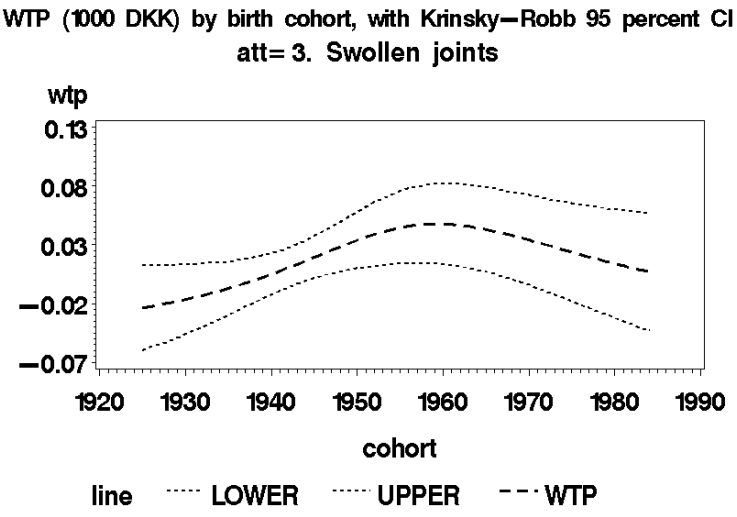
The older the respondent was the higher the WTP for a reduction in morning stiffness or tiredness (figure 8). Looking at the WTP for a reduction in pain, swollen joints or adverse effect, the WTP seemed to increase until a certain level where it peaked and afterwards

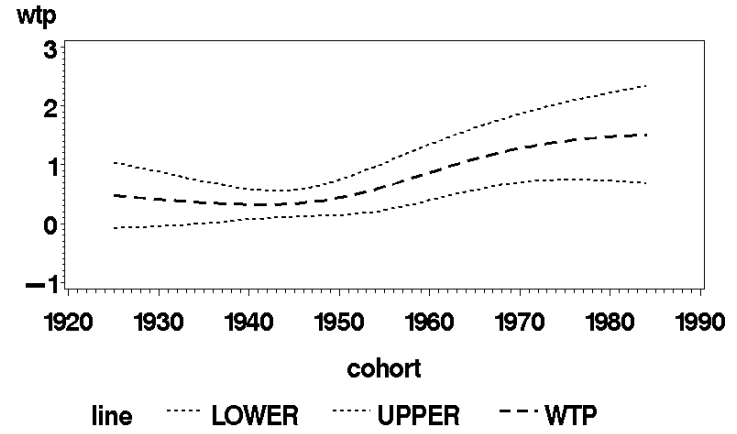
decreased again. The peak points were for respondents born in 1960, 1958 and 1955 respectively.

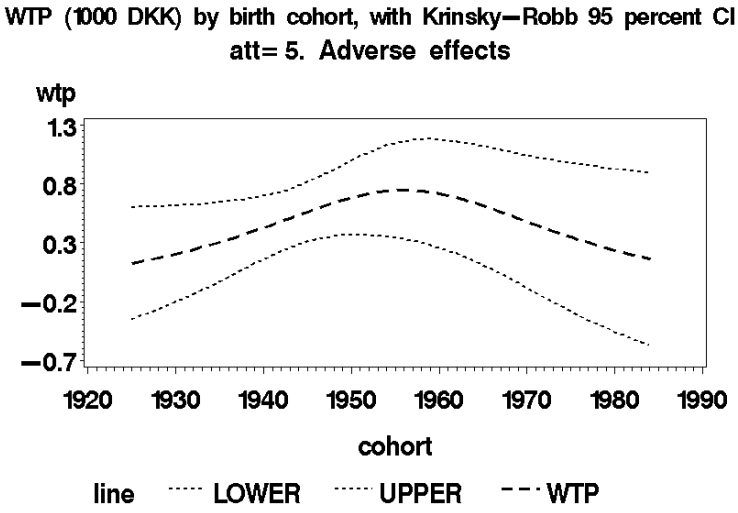
Figure 8.: WTP (1000 DKK) for attributes 1-5, by birth cohort (v3)





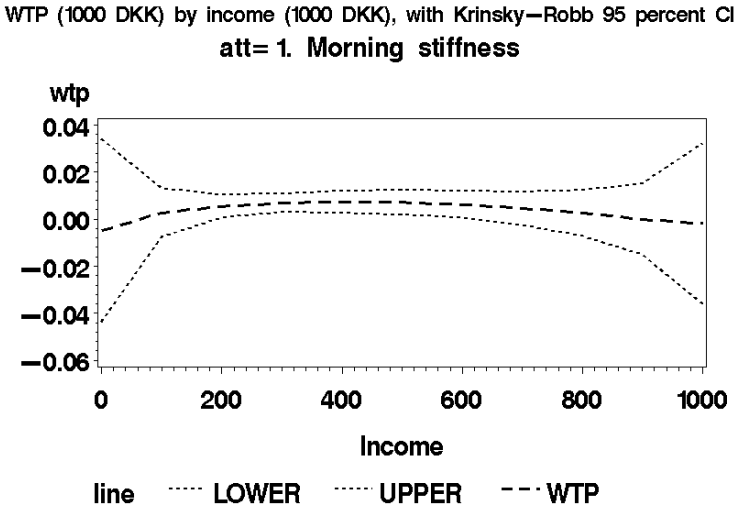

WTP (1000 DKK) by birth cohort, with Krinsky−Robb 95 percent CI
att= 3. Swollen joints
wtp
0.13
0.08
0.03
−0.02
−0.07
1920
1930
1940
1950
1960
1970
1980
1990
cohort
line
LOWER
UPPER
WTP

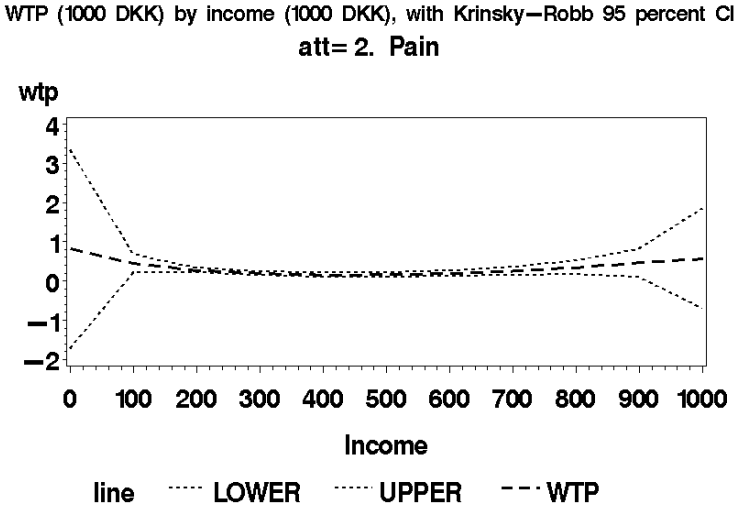

WTP (1000 DKK) by birth cohort, with Krinsky−Robb 95 percent CI
att= 4. Tiredness
wtp
3
2
1
0
−1
1920
1930
1940
1950
1960
1970
1980
1990
cohort
line
LOWER
UPPER
WTP



As illustrated in figure 9, WTP for a reduction in morning stiffness, swollen joints or tiredness rised with increasing income but seemed to decrease again for morning stiffness or swollen joints after an income level of about 700.000 DKK had been reached. After an income level higher than 550.000 DKK their WTP for a reduction in tiredness appeared to stabilize. WTP for a reduction in adverse effects seemed almost unaffected by the income level, whereas WTP for a reduction in pain level decreased to begin with, at small income levels, then slightly increased as income increased.

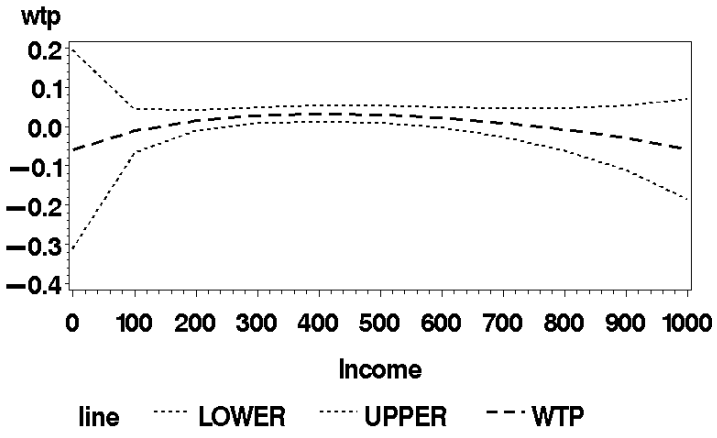
Figure 9. WTP (1000 DKK) for attributes 1-5, by income (1000 DKK) (v92)





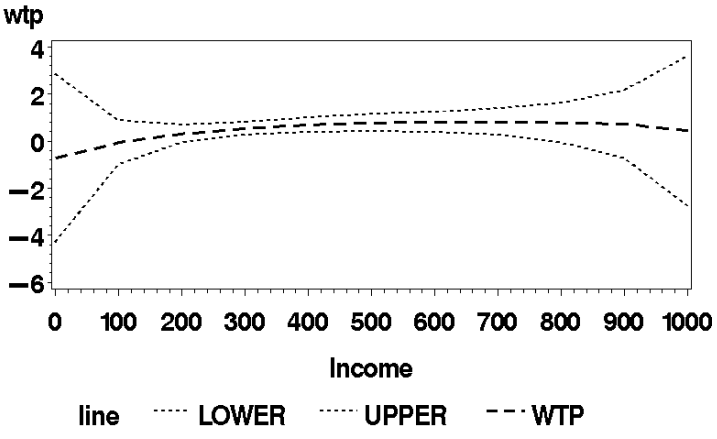
WTP (1000 DKK) by income (1000 DKK), with Krinsky−Robb 95 percent CI

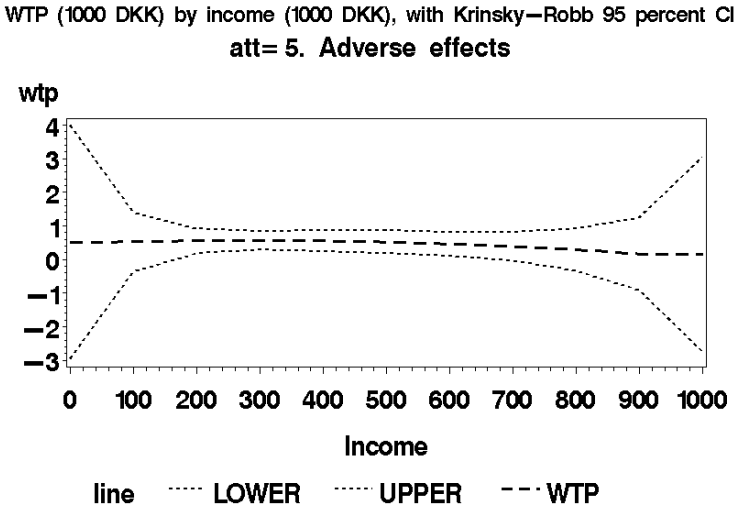
att= 3. Swollen joints



WTP (1000 DKK) by income (1000 DKK), with Krinsky−Robb 95 percent CI

att= 4. Tiredness




WTP (1000 DKK) by income (1000 DKK), with Krinsky−Robb 95 percent CI
att= 5. Adverse effects
wtp
4
3
2
1
0
−1
−2
−3
0
100
200
300
400
500
600
700
800
900
1000
Income
line
LOWER
UPPER
WTP

5. Discussion

Reliability at the input level was investigated by considering how consistent the respondent's answers to the DCE-questions were over time. Results (table 2-5) showed that a good correspondence between the choices in the surveys was found, also confirmed by the highly significant chi-square statistics. Thus, on the input level DCE was reliable.

Focussing on DCE´s reliability at the output level, first Wald test results were presented (Table 6). The test was analogue to the F-tests presented by Bryan et al [6]. However, since we, in our model, controlled for individual effects, we preferred the asymptotic Wald test to ensure results were consistent (and not necessarily unbiased or minimum variant in a finite-sample context). It appeared that survey 2 did not differ from survey 1. The same was the case when comparing survey 1 and survey 3. Taking a closer look at the confidence intervals (Table 7) showed, that the confident intervals for the calculated WTP at survey time 1 and 2 did overlap, and again looking at survey 3, results showed that there were no differences between survey 1 and 3. Hence, overall, the results were constant over surveys 1, 2 and 3, i.e DCE was reliable at the output level.

How the WTP value calculated was influenced by different individual characteristics were shown in tables 8-11. We focused on the variables selected in the DCE model, i.e. if the respondent had monthly expenditures for prescriptive drugs, the respondent's gender, civil status and occupation. Interestingly, the results showed that WTP appeared to differ among the groups of respondents depending on their individual characteristics. Figure 1-9 illustrates continuously variables selected in the DCE model and their influence on respondent's WTP. The tendencies illustrated in the figures were as expected. For example, we would expect that the longer the respondent's duration of illness the more likely his WTP would decrease or only slightly increase, since the longer he had suffered from rheumatoid arthritis the less he would probably expect to gain from a reduction in the attributes (for example the longer his duration of illness the less his chances for a connection to the job market and thus the less the expected gain from a reduction in attribute levels).

The length of time between the surveys was an important issue when investigating reliability. Too short an interval might result in a memory effect, where the respondents were able to recall there earlier answers, and hence the reliability of the method would become misleading. On the other hand, if the time interval was too long changes could appear that would affect underlying preferences. In the present study the time interval between the three face-to-face interviews in which the respondents participated was 4 months. This interval was chosen since it was assessed that the interval would be long enough to ensure that no memory effect was present. In these periods of four month the respondents' health status could in principle have changed. Since the DCE scenarios were concerned with the effect of RA treatment and thus the respondents' health status may have influenced their choices, a change in health status could have an important effect. A possible health status change was taken into account by asking in detail about the

respondents' clinical condition in each interview, being able to use this information to ensure that the reliability measure could be adjusted to only reflect changes in 'reproducibility'. It was, however, tested whether the average TTO score remained constant through the three surveys (results not reported). This hypothesis could not be rejected, and hence we assumed that health status was unchanged.

Respondent 'drop outs' in survey 2 and 3 may have influenced the results. It might for example have been the case that these respondents felt, because of devaluation of their health, that they would not be able to participate in another interview. However, we don't know the reasons for why they didn't want to participate in the second (and/or third) survey.

The attributes and their relative weights for the three surveys were presented in Appendix B Table B1. It appeared that most variables had a significant effect on utility. However, a few variables were not significant, but still we chose to include them in the model, since from a theoretical point they appeared to be important. For example the respondents income. The income variable was of importance since the respondents WTP would be expected to be influenced by the variable, if the respondent, that is, keept his budget restriction in mind when looking at the cost involved in the choice scenarios. This variable appeared to be significant in the non-adjusted logit model, but came out as insignificant when the model was controlled for random effects.

To the authors' knowledge measuring reliability in the health care field has only been done once before by Bryan et al [6]. Bryan et al also investigated reliability over time. Their results were promising, indicating a high reliability level at both the input data and result levels. The results presented in this paper, equally demonstrated a high level of reliability at both input and output levels. This study however, used a data material with a much longer time interval between measurements and a set of hypothetical alternatives that seemed more complex than in the before mentioned study, making the results presented in this paper even stronger. Further, when comparing the two studies, the reader's attention should be drawn to the fact, that different statistic measures have been applied when investigating reliability at the input level. In the present study χ^2 were applied, where Bryan et al used the kappa (K) statistic. We believe that χ^2 is a more objective measure, since it could be used to explain a probability for classification-by-chance, where the K statistic is more of an ad-hoc measure. Also, at the output/result level, Wald test results were presented which however were analogue to the F-tests presented by Bryan et al. Further, at the result level, in this study we made what could be considered an encompassing of Bryan et al's models P1 and P2. This encompassing was important to avoid bias, since if the WTP values over surveys were related to individual characteristics (which they were), a model as P1 would be biased and F-test not reliable. Even though different test statistics have been applied in the two studies the authors believe a comparison is justified.

This paper demonstrates a high level of conjoint reliability in DCE at the input as well as the output level. Future work exploring other forms of reliability in conjoint measurement is encouraged to resolve the remaining methodological issues related to CA.

6. References

1. Beggs S, Cardell S, Hausman J. Assessing the potential demand for electric cars. Journal of Econometrics 1981;16:1-19.

2. Train K. Qualitative choice analysis. Theory, econometrics and an application to automobile demand. Cambridge. Massachussets,1986.

3. Cramer JS. The logit model for economists. Edward Arnold. London, 1991.

4. Frankfort-Nachmias C, Nachmias D. Research Methods in the Social Sciences. 5th edition. Arnold, 1996.

5. Bateson JE, Reibstein D, Boulding W. Conjoint analysis reliability and validity: a framework for future research. In Review of Marketing, Houston MJ (ed). American Marketing Association: Chicago, 1987; 451-481.

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

7. Furst DE, Breedveld FC, Kalden JR, Smolen JS, Burmester G R, Bijlsma JWJ, Dougados M, Emery P, Keystone EC, Klareskog L, Mease PJ. Updated consensus statement on biological agents, specifically tumour necrosis factor • (TNF•) blocking agents and interleukin -1 receptor antagonist (IL-1ra) for the treatment of rheumatic diseases. Ann Rheum Dis 2004;63(Suppl II):ii2-ii12

8. Slothuus U, Larsen ML, Junker P. Willingness to pay for arthritis symptom alleviation. Comparison of Closed-ended Questions With and Without Follow-up. International Journal of Technology Assessment in Health Care 2000;16(1):60-72.

9. Ryan M, Hughes J. Using conjoint analysis to assess women's preferences for miscarriage management. Health Econ 1997;6(3):261-274.

10. Pedersen KM, Wittrup-Jensen K, Brooks R, Gudex C Vaerdisaetning af sundhed [Valuation of health]. SDU 2003. (in Danish)

Appendix A

Test for equal distribution across eight groups with respect to cohort, gender and duration.

A Wald test for equality of average of v3 (birth year), v4 (female) and v5 (duration of illness) across eight groups, is performed for each survey:

For each survey, a composite zero hypothesis regarding v3 specifies that the average of v3 is the same for the eight groups, i.e. $H_0 : \overline{x}_2 = \overline{x}_1, \overline{x}_3 = \overline{x}_1, ..., \overline{x}_8 = \overline{x}_1$, or $H_0 : L\overline{x} = 0$, where L is a (K-*1*) by K (K is the number of groups, i.e. K=8) matrix with -1 in the first column, 1 in element (k,k+1), and 0 otherwise, and $\overline{x}$ is the vector of $\overline{x}_1, \overline{x}_2, \overline{x}_3, .., \overline{x}_8$.

The zero hypothesis is tested for each survey using a Wald test: $W = (L\overline{x})'(LVL')^{-1}(L\overline{x})$, where V is a diagonal matrix containing the variances of $\overline{x}_1, \overline{x}_2, ..., \overline{x}_8$ as diagonal elements. The Wald test follows a chi-square distribution with K-1=7 degrees of freedom.

Likewise, Wald tests are calculated for proportion of female and for average duration of illness.

The means, standard deviations and number of observations for each group within each survey are reported at the bottom of the paper.

The Wald tests are as follows:

survey	variable	wald	df	prob
1	cohort	6.380303	7	0.4961097
2	cohort	6.380303	7	0.4961097
3	cohort	6.380303	7	0.4961097
1	gender	1.434134	7	0.9845028
2	gender	1.434134	7	0.9845028
3	gender	1.434134	7	0.9845028
1	duration	11.506124	7	0.1180152
2	duration	11.506124	7	0.1180152
3	duration	11.506124	7	0.1180152

It is seen that the zero hypothesis cannot be rejected for any of the variables (the probability is large than 10 percent for all variables), i.e. it can be concluded that the respondents are equally distributed across the eight groups with respect to cohort, gender and durations

Descriptive statistics:

survey=1 Gruppe nr.=1

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
V3	Cohort	22	46.5909091	10.2057831	29.0000000	71.0000000
V4	Gender	23	0.6521739	0.4869848	0	1.0000000
v5	Duration	22	13.5909091	10.0411491	3.0000000	40.0000000

survey=1 Gruppe nr.=2

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
V3	Cohort	20	50.2000000	12.2242210	37.0000000	84.0000000
V4	Gender	21	0.6190476	0.4976134	0	1.0000000
v5	Duration	20	9.4500000	6.6052212	1.0000000	24.0000000

survey=1 Gruppe nr.=3

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
V3	Cohort	22	50.5909091	8.5113878	34.0000000	77.0000000
V4	Gender	23	0.7391304	0.4489778	0	1.0000000
v5	Duration	22	10.3636364	7.5501211	1.0000000	27.0000000

survey=1 Gruppe nr.=4

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
V3	Cohort	22	53.4545455	9.6792723	39.0000000	77.0000000
V4	Gender	23	0.6521739	0.4869848	0	1.0000000
v5	Duration	22	11.7272727	9.5527698	1.0000000	31.0000000

survey=1 Gruppe nr.=5

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
V3	Cohort	21	49.1428571	11.3149711	35.0000000	75.0000000
V4	Gender	21	0.7142857	0.4629100	0	1.0000000
v5	Duration	21	15.1904762	10.4767316	1.0000000	35.0000000

survey=1 Gruppe nr.=6

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
V3	Cohort	20	50.7500000	13.5253174	34.0000000	80.0000000
V4	Gender	20	0.6500000	0.4893605	0	1.0000000
v5	Duration	19	11.3684211	8.3613564	2.0000000	30.0000000

survey=1 Gruppe nr.=7

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
V3	Cohort	24	48.2916667	10.0107732	25.0000000	72.0000000
V4	Gender	24	0.6250000	0.4945354	0	1.0000000
v5	Duration	24	15.7500000	10.7430463	2.0000000	47.0000000

survey=1 Gruppe nr.=8

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
V3	Cohort	21	48.8571429	11.3678745	34.0000000	69.0000000
V4	Gender	21	0.7142857	0.4629100	0	1.0000000
v5	Duration	21	14.6666667	8.5634884	4.0000000	34.0000000

survey=2 Gruppe nr.=1

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
V3	Cohort	22	46.5909091	10.2057831	29.0000000	71.0000000
V4	Gender	23	0.6521739	0.4869848	0	1.0000000
v5	Duration	22	13.5909091	10.0411491	3.0000000	40.0000000

```
------------------------------------ survey=2 Gruppe nr.=2 -----------------------------
Variable    Label          N          Mean          Std Dev        Minimum      Maximum
-------------------------------------------------------------------------------------
V3          Cohort         20         50.2000000    12.2242210    37.0000000  84.0000000
V4          Gender         21          0.6190476     0.4976134           0     1.0000000
v5          Duration       20          9.4500000     6.6052212     1.0000000  24.0000000
-------------------------------------------------------------------------------------

------------------------------------ survey=2 Gruppe nr.=3 -----------------------------
Variable    Label          N          Mean          Std Dev       Minimum        Maximum
-------------------------------------------------------------------------------------
V3          Cohort         22         50.5909091    8.5113878     34.0000000  77.0000000
V4          Gender         23          0.7391304    0.4489778           0     1.0000000
v5          Duration       22         10.3636364    7.5501211     1.0000000  27.0000000
-------------------------------------------------------------------------------------

------------------------------------ survey=2 Gruppe nr.=4 -----------------------------
Variable    Label          N          Mean          Std Dev       Minimum        Maximum
-------------------------------------------------------------------------------------
V3          Cohort         22         53.4545455    9.6792723     39.0000000  77.0000000
V4          Gender         23          0.6521739    0.4869848           0     1.0000000
v5          Duration       22         11.7272727    9.5527698     1.0000000  31.0000000
-------------------------------------------------------------------------------------

------------------------------------ survey=2 Gruppe nr.=5 -----------------------------
Variable    Label          N          Mean          Std Dev         Minimum      Maximum
-------------------------------------------------------------------------------------
V3          Cohort         21         49.1428571    11.3149711    35.0000000  75.0000000
V4          Gender         21          0.7142857     0.4629100           0     1.0000000
v5          Duration       21         15.1904762    10.4767316     1.0000000  35.0000000
-------------------------------------------------------------------------------------

------------------------------------ survey=2 Gruppe nr.=6 -----------------------------
Variable    Label          N          Mean          Std Dev         Minimum      Maximum
-------------------------------------------------------------------------------------
V3          Cohort         20         50.7500000    13.5253174    34.0000000  80.0000000
V4          Gender         20          0.6500000     0.4893605           0     1.0000000
v5          Duration       19         11.3684211     8.3613564     2.0000000  30.0000000
-------------------------------------------------------------------------------------

------------------------------------ survey=2 Gruppe nr.=7 -----------------------------
Variable    Label           N         Mean          Std Dev         Minimum      Maximum
-------------------------------------------------------------------------------------
V3          Cohort          24        48.2916667    10.0107732    25.0000000  72.0000000
V4          Gender          24         0.6250000     0.4945354           0     1.0000000
v5          Duration        24        15.7500000    10.7430463     2.0000000  47.0000000
-------------------------------------------------------------------------------------

------------------------------------ survey=2 Gruppe nr.=8 -----------------------------
Variable    Label           N         Mean          Std Dev         Minimum      Maximum
-------------------------------------------------------------------------------------
V3          Cohort          21        48.8571429    11.3678745    34.0000000  69.0000000
V4          Gender          21         0.7142857     0.4629100           0     1.0000000
v5          Duration        21        14.6666667     8.5634884     4.0000000  34.0000000
-------------------------------------------------------------------------------------

------------------------------------ survey=3 Gruppe nr.=1 -----------------------------
Variable    Label           N         Mean          Std Dev         Minimum      Maximum
-------------------------------------------------------------------------------------
V3          Cohort          22        46.5909091    10.2057831    29.0000000  71.0000000
V4          Gender          23         0.6521739     0.4869848           0     1.0000000
v5          Duration        22        13.5909091    10.0411491     3.0000000  40.0000000
-------------------------------------------------------------------------------------

------------------------------------ survey=3 Gruppe nr.=2 -----------------------------
Variable    Label           N         Mean          Std Dev         Minimum      Maximum
-------------------------------------------------------------------------------------
V3          Cohort          20        50.2000000    12.2242210    37.0000000  84.0000000
V4          Gender          21         0.6190476     0.4976134           0     1.0000000
v5          Duration        20         9.4500000     6.6052212     1.0000000 24.0000000
-------------------------------------------------------------------------------------
```

---------- survey=3 Gruppe nr.=3 ----------

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
V3	Cohort	22	50.5909091	8.5113878	34.0000000	77.0000000
v4	Gender	23	0.7391304	0.4489778	0	1.0000000
v5	Duration	22	10.3636364	7.5501211	1.0000000	27.0000000

---------- survey=3 Gruppe nr.=4 ----------

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
V3	Cohort	22	53.4545455	9.6792723	39.0000000	77.0000000
V4	Gender	23	0.6521739	0.4869848	0	1.0000000
v5	Duration	22	11.7272727	9.5527698	1.0000000	31.0000000

---------- survey=3 Gruppe nr.=5 ----------

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
V3	Cohort	21	49.1428571	11.3149711	35.0000000	75.0000000
V4	Gender	21	0.7142857	0.4629100	0	1.0000000
v5	Duration	21	15.1904762	10.4767316	1.0000000	35.0000000

---------- survey=3 Gruppe nr.=6 ----------

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
V3	Cohort	20	50.7500000	13.5253174	34.0000000	80.0000000
V4	Gender	20	0.6500000	0.4893605	0	1.0000000
v5	Duration	19	11.3684211	8.3613564	2.0000000	30.0000000

---------- survey=3 Gruppe nr.=7 ----------

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
V3	Cohort	24	48.2916667	10.0107732	25.0000000	72.0000000
V4	Gender	24	0.6250000	0.4945354	0	1.0000000
v5	Duration	24	15.7500000	10.7430463	2.0000000	47.0000000

---------- survey=3 Gruppe nr.=8 ----------

variable	Label	N	Mean	Std Dev	Minimum	Maximum
V3	Cohort	21	48.8571429	11.3678745	34.0000000	69.0000000
V4	Gender	21	0.7142857	0.4629100	0	1.0000000
v5	Duration	21	14.6666667	8.5634884	4.0000000	34.0000000

Appendix B

Tabel B1 Logistic regression model with adjustment for random individual variation

Variable	Parameter	Std. Err	95% conf. limits		Prob.
			lower	upper	
Intercept	0.2248	0.0673	0.0929	0.3568	0.0008
Att1	-0.0362	0.0219	-0.0791	0.0067	0.0977
Att1survey2	0.0005	0.0018	-0.0031	0.0041	0.7742
Att1survey3	0.0021	0.0024	-0.0026	0.0069	0.3826
Att1duration of illness	0.0011	0.0006	0.0000	0.0022	0.0445
Att1duration of illness^2	-0.0000	0.0000	-0.0001	0.0000	0.1756
Att1reported degree of morning stiffness	-0.0002	0.0014	-0.0029	0.0026	0.9010
Att1reported degree of morning stiffness^2	-0.0001	0.0001	-0.0003	0.0002	0.4504
Att1reported degree of pain	0.0016	0.0021	-0.0024	0.0057	0.4280
Att1reported degree of pain^2	-0.0001	0.0001	-0.0004	0.0002	0.4732
Att1reported degree of swollen joints	-0.0012	0.0015	-0.0042	0.0018	0.4397
Att1reported degree of swollen joints^2	0.0001	0.0001	-0.0002	0.0003	0.5057
Att1reported degree of tiredness	-0.0003	0.0016	-0.0035	0.0029	0.8565
Att1reported degree of tiredness^2	0.0000	0.0001	-0.0002	0.0003	0.7529
Att1reported degree of adverse effects	-0.0008	0.0016	-0.0039	0.0023	0.6196
Att1reported degree of adverse effects^2	0.0001	0.0001	-0.0002	0.0003	0.6114

Att1prescriptive drug	-0.0033	0.0049	-0.0129	0.0062	0.4940
Att1TTO	0.0387	0.0145	0.0103	0.0672	0.0076
Att1TTO^2	-0.0361	0.0164	-0.0682	-0.0039	0.0280
Att1birth cohort	0.0013	0.0008	-0.0002	0.0028	0.0799
Att1birth cohort^2	-0.0000	0.0000	-0.0000	-0.0000	0.0449
Att1gender	0.0030	0.0024	-0.0018	0.0077	0.2194
Att1civil status	-0.0011	0.0031	-0.0071	0.0050	0.7271
Att1occupation_2	-0.0050	0.0047	-0.0143	0.0042	0.2852
Att1occupation_3	-0.0074	0.0036	-0.0144	-0.0004	0.0371
Att1occupation_4	-0.0021	0.0046	-0.0112	0.0070	0.6549
Att1income	-0.0327	0.0310	-0.0935	0.0282	0.2926
Att1income^2	0.0337	0.0359	-0.0366	0.1040	0.3471
Att2	-0.2378	0.2671	-0.7612	0.2857	0.3734
Att2survey2	0.0097	0.0266	-0.0424	0.0618	0.7155
Att2survey3	-0.0217	0.0374	-0.0951	0.0517	0.5624
Att2duration of illness	-0.0020	0.0058	-0.0134	0.0094	0.7342
Att2duration of illness^2	-0.0000	0.0002	-0.0004	0.0003	0.7941
Att2reported degree of morning stiffness	-0.0478	0.0188	-0.0846	-0.0109	0.0111
Att2reported degree of morning stiffness^2	0.0051	0.0016	0.0020	0.0083	0.0014
Att2reported degree of pain	-0.0228	0.0240	-0.0698	0.0243	0.3433
Att2reported degree of pain^2	0.0016	0.0018	-0.0018	0.0050	0.3590
Att2reported degree of swollen joints	-0.0259	0.0194	-0.0641	0.0122	0.1820
Att2reported degree of swollen joints^2	0.0025	0.0015	-0.0004	0.0054	0.0935
Att2reported degree of tiredness	-0.0330	0.0216	-0.0754	0.0093	0.1263
Att2reported degree of	0.0022	0.0019	-0.0016	0.0060	0.2573

tiredness^2					
Att2reported degree of adverse effects	-0.0191	0.0205	-0.0593	0.0211	0.3522
Att2reported degree of adverse effects^2	0.0025	0.0016	-0.0007	0.0057	0.1300
Att2prescriptive drug	0.1614	0.0451	0.0729	0.2499	0.0004
Att2TTO	-0.1445	0.2250	-0.5855	0.2965	0.5209
Att2TTO^2	0.2329	0.2323	-0.2223	0.6882	0.3159
Att2birth cohort	0.0012	0.0086	-0.0157	0.0180	0.8914
Att2birth cohort^2	-0.0000	0.0001	-0.0002	0.0001	0.7689
Att2gender	0.0301	0.0319	-0.0325	0.0926	0.3461
Att2civil status	-0.0063	0.0382	-0.0811	0.0685	0.8693
Att2occupation_2	-0.1962	0.0696	-0.3326	-0.0598	0.0048
Att2occupation_3	0.1435	0.0569	0.0319	0.2551	0.0117
Att2occupation_4	0.0912	0.0569	-0.0203	0.2027	0.1088
Att2income	1.0811	0.4739	0.1523	2.0098	0.0225
Att2income^2	-1.3265	0.5616	-2.4273	-0.2257	0.0182
Att3	0.2580	0.1183	0.0261	0.4900	0.0292
Att3survey2	0.0139	0.0102	-0.0060	0.0338	0.1722
Att3survey3	-0.0100	0.0109	-0.0114	0.0314	0.3588
Att3duration of illness	-0.0009	0.0017	-0.0042	0.0025	0.6094
Att3duration of illness^2	0.0000	0.0000	-0.0001	0.0001	0.4025
Att3reported degree of morning stiffness	0.0036	0.0065	-0.0092	0.0165	0.5769
Att3reported degree of morning stiffness^2	-0.0002	0.0006	-0.0013	0.0009	0.6729
Att3reported degree of pain	-0.0199	0.0106	-0.0408	0.0010	0.0618
Att3reported degree of pain^2	0.0017	0.0007	0.0002	0.0031	0.0231
Att3reported degree of swollen joints	-0.0020	0.0059	-0.0136	0.0097	0.7421
Att3reported degree of swollen	0.0001	0.0005	-0.0008	0.0011	0.7956

joints^2					
Att3reported degree of tiredness	0.0001	0.0069	-0.0135	0.0136	0.9893
Att3reported degree of tiredness^2	-0.0000	0.0006	-0.0012	0.0011	0.9415
Att3reported degree of adverse effects	-0.0094	0.0070	-0.0231	0.0042	0.1761
Att3reported degree of adverse effects^2	0.0008	0.0006	-0.0003	0.0018	0.1709
Att3prescriptive drug	-0.0040	0.0248	-0.0527	0.0446	0.8708
Att3TTO	0.0161	0.0586	-0.0988	0.1310	0.7836
Att3TTO^2	-0.0305	0.0621	-0.1522	0.0912	0.6228
Att3birth cohort	-0.0061	0.0037	-0.0133	0.0011	0.0985
Att3birth cohort^2	0.0000	0.0000	-0.0000	0.0001	0.1409
Att3gender	-0.0001	0.0102	-0.0200	0.0199	0.9961
Att3civil status	0.0178	0.0130	-0.0077	0.0433	0.1707
Att3occupation_2	0.0088	0.0222	-0.0347	0.0523	0.6922
Att3occupation_3	-0.0315	0.0180	-0.0667	0.0037	0.0795
Att3occupation_4	-0.0033	0.0224	-0.0472	0.0406	0.8826
Att3income	-0.2452	0.1223	-0.4849	-0.0055	0.0449
Att3income^2	0.2784	0.1280	0.0276	0.5293	0.0296
Att4	-1.6155	1.6625	-4.8740	1.6429	0.3312
Att4survey2	0.2372	0.1661	-0.0884	0.5629	0.1534
Att4survey3	0.3164	0.2148	-0.1047	0.7374	0.1408
Att4duration of illness	0.0494	0.0345	-0.0182	0.1170	0.1518
Att4duration of illness^2	-0.0024	0.0010	-0.0044	-0.0004	0.0213
Att4reported degree of morning stiffness	0.0478	0.1133	-0.1742	0.2699	0.6730
Att4reported degree of morning stiffness^2	-0.0037	0.0098	-0.0230	0.0155	0.7032
Att4reported degree of pain	0.0637	0.1614	-0.2527	0.3801	0.6932
Att4reported degree of pain^2	-0.0058	0.0120	-0.0294	0.0178	0.6300

Att4reported degree of swollen joints	-0.0092	0.1186	-0.2417	0.2233	0.9384
Att4reported degree of swollen joints^2	0.0009	0.0099	-0.0185	0.0202	0.9298
Att4reported degree of tiredness	-0.0678	0.1424	-0.3469	0.2114	0.6342
Att4reported degree of tiredness^2	0.0048	0.0137	-0.0221	0.0318	0.7242
Att4reported degree of adverse effects	-0.0722	0.1207	-0.3088	0.1644	0.5496
Att4reported degree of adverse effects^2	-0.0009	0.0096	-0.0198	0.0180	0.9237
Att4prescriptive drug	-0.1805	0.3593	-0.8847	0.5238	0.6155
Att4TTO	-0.2344	0.9347	-2.0664	1.5976	0.8020
Att4TTO^2	0.0707	1.0534	-1.9940	2.1354	0.9465
Att4birth cohort	0.1142	0.0482	0.0199	0.2086	0.0177
Att4birth cohort^2	-0.0012	0.0004	-0.0021	-0.0004	0.0058
Att4gender	-0.0681	0.1936	-0.4475	0.3113	0.7251
Att4civil status	0.0954	0.2611	-0.4162	0.6071	0.7147
Att4occupation_2	0.0150	0.3999	-0.7688	0.7987	0.9702
Att4occupation_3	-0.1642	0.2519	-0.6579	0.3295	0.5145
Att4occupation_4	-0.6506	0.3863	-1.4076	0.1065	0.0921
Att4income	-2.9125	2.6560	-8.1181	2.2931	0.2728
Att4income^2	1.9911	2.8086	-3.5137	7.4958	0.4784
Att5	1.2884	1.3176	-1.2941	3.8709	0.3282
Att5survey2	0.2391	0.1449	-0.0448	0.5231	0.0988
Att5survey3	0.1289	0.1618	-0.1883	0.4460	0.4257
Att5duration of illness	-0.0153	0.0252	-0.0646	0.0340	0.5431
Att5duration of illness^2	0.0001	0.0007	-0.0012	0.0015	0.8395
Att5reported degree of morning stiffness	-0.1013	0.0822	-0.2623	0.0598	0.2178
Att5reported degree of morning	0.0078	0.0073	-0.0066	0.0221	0.2880

stiffness^2					
Att5reported degree of pain	0.0295	0.1489	-0.2624	0.3215	0.8427
Att5reported degree of pain^2	-0.0014	0.0104	-0.0217	0.0189	0.8948
Att5reported degree of swollen joints	-0.0823	0.0853	-0.2495	0.0849	0.3348
Att5reported degree of swollen joints^2	0.0079	0.0072	-0.0061	0.0219	0.2691
Att5reported degree of tiredness	-0.0821	0.1007	-0.2794	0.1153	0.4150
Att5reported degree of tiredness^2	0.0091	0.0089	-0.0083	0.0265	0.3046
Att5reported degree of adverse effects	0.0311	0.1216	-0.2073	0.2695	0.7983
Att5reported degree of adverse effects^2	-0.0023	0.0086	-0.0193	0.0146	0.7861
Att5prescriptive drug	-0.0317	.03526	-0.7227	0.6593	0.9283
Att5TTO	-0.2257	1.2432	-2.6622	2.2109	0.8560
Att5TTO^2	-0.1229	1.2326	-2.5389	2.2930	0.9206
Att5birth cohort	-0.0272	0.0462	-0.1177	0.0634	0.5564
Att5birth cohort^2	0.0002	0.0004	-0.0006	0.0011	0.5613
Att5gender	-0.1492	0.1412	-0.4260	0.1276	0.2907
Att5civil status	-0.2825	0.2010	-0.6765	0.1115	0.1599
Att5occupation_2	-0.0944	0.2986	-0.6798	0.4909	0.7518
Att5occupation_3	0.0970	0.2790	-0.4499	0.6439	0.7282
Att5occupation_4	-0.1631	0.2854	-0.7224	0.3962	0.5676
Att5income	-1.1711	1.7305	-4.5628	2.2207	0.4986
Att5income^2	1.2343	1.9286	-2.5458	5.0143	0.5222
Att6	-3.1373	1.4160	-5.9126	-0.3619	0.0267
Att6survey2	0.0358	0.1014	-0.1630	0.2345	0.7244
Att6survey3	-0.0793	0.1193	-0.3132	0.1545	0.5061
Att6duration of illness	-0.0055	0.0227	-0.0500	0.0389	0.8077
Att6duration of illness^2	-0.0005	0.0007	-0.0019	0.0009	0.4987

Att6reported degree of morning stiffness	-0.0645	0.0693	-0.2003	0.0714	0.3524
Att6reported degree of morning stiffness^2	-0.0051	0.0061	-0.0068	0.0170	0.4036
Att6reported degree of pain	-0.0492	0.1148	-0.2743	0.1758	0.6681
Att6reported degree of pain^2	0.0055	0.0087	-0.0115	0.0225	0.5280
Att6reported degree of swollen joints	-0.0144	0.0805	-0.1722	0.1433	0.8578
Att6reported degree of swollen joints^2	0.0029	0.0068	-0.0105	0.0163	0.6738
Att6reported degree of tiredness	-0.1599	0.0843	-0.3251	0.0053	0.0578
Att6reported degree of tiredness^2	0.0125	0.0079	-0.0030	0.0279	0.1131
Att6reported degree of adverse effects	-0.0605	0.0810	-0.2193	0.0982	0.4549
Att6reported degree of adverse effects^2	0.0044	0.0065	-0.0083	0.0171	0.4951
Att6prescriptive drug	-0.2878	0.2654	-0.8081	0.2325	0.2783
Att6TTO	-1.2552	0.4871	-2.2098	-0.3005	0.0100
Att6TTO^2	1.3948	0.6325	0.1550	2.6345	0.0274
Att6birth cohort	0.1173	0.0453	0.0285	0.2062	0.0096
Att6birth cohort^2	-0.0010	0.0004	-0.0019	-0.0002	0.0110
Att6gender	0.0614	0.1328	-0.1989	0.3217	0.6438
Att6civil status	0.2708	0.1577	-0.0382	0.5798	0.0859
Att6occupation_2	0.0096	0.2016	-0.3855	0.4046	0.9622
Att6occupation_3	0.1206	0.1694	-0.2114	0.4526	0.4764
Att6occupation_4	-0.3348	0.2292	-0.7841	0.1146	0.1442
Att6income	-0.7337	1.5923	-3.8546	2.3872	0.6450
Att6income^2	0.1823	1.7300	-3.2085	3.5731	0.9161

Note to table:

att1 refers to attribute no.1 in Table 1, att2 refers to attribute no. 2, etc.
duration of illness = the length of time the respondent has been diagnosed with arthritis
reported degree of morning stiffness = the respondents own valuation of experiencing morning stiffness on a scale from 0 to 10 (10 being the worst)
reported degree of pain= the respondents own valuation of experiencing pain on a scale from 0 to 10 (10 being the worst)
reported degree of swollen joints= the respondents own valuation of experiencing swollen joints on a scale from 0 to 10 (10 being the worst)
reported degree of tiredness= the respondents own valuation of experiencing tiredness on a scale from 0 to 10 (10 being the worst)
reported degree of adverse effects= the respondents own valuation of experiencing adverse effects on a scale from 0 to 10 (10 being the worst)
prescriptive drug= does the respondents have a monthly expenditure for prescriptive drugs (yes,no)
TTO= the EURO Qol estimate (Danish weights [8] have been used for calculation purposes)
birth cohort=the respondents year of birth
gender= the respondents gender (male=0; female=1)
civil status=the respondents civil status, (singel=0, married/cohab =1)
occupation=self-employed=1:public/private employed=2:retired=3:other non-employed=4
income= the respondents yearly income before tax (in 1000 DKK)