University of St. Andrews

Discussion papers in Environmental Economics

http://www.st-andrews.ac.uk/gsd/research/envecon/eediscus/

Paper 2016-13

Giving respondents "time to think" reduces response randomness in repeated discrete choice tasks

Tobias Börger and Joseph Cook

Keywords: Discrete choice experiment, time to think, generalized multinomial logit, scale factor, Vietnam

JEL codes: D6, H1, Q51

Giving respondents "time to think" reduces response randomness in repeated discrete choice tasks

Tobias Börger[a,b*], Joseph Cook[c]

[a] Department of Geography and Sustainable Development, University of St Andrews, UK

[b] Plymouth Marine Laboratory, Plymouth, UK

[c] Evans School of Public Policy & Governance, University of Washington, Seattle, USA

* Corresponding author: tb79@st-andrews.ac.uk

Abstract: Earlier work has shown that giving respondents to stated preference surveys time to think about their choices shifts preferences by increasing price sensitivity and lowering WTP estimates. In this study we use the generalized multinomial logit model to assess the effect of time to think on preferences and response scale simultaneously. Response scale is an indicator of the weight of the deterministic component of indirect utility relative to the error component. Results show both effects. Time to think leads to different preference parameters and significantly increases scale as well. Respondents who are given overnight time to think about their choices make more consistent, i.e. less random, choices. Controlling for the scale effect lessens the impact of time to think on preference weights in the random utility models.

Keywords: Discrete choice experiment, time to think, generalized multinomial logit, scale factor, Vietnam

1. Introduction

The majority of stated preference surveys ask respondents to make choices with relatively little time for reflection and study. Surveys using in-person interviews and the increasingly-common web-based approaches typically move directly from presentation of the hypothetical scenarios to choice tasks. Mail surveys implicitly give respondents as much time as they need to complete the survey, but the role of reflection is poorly observed by the researcher, if at all, so it is difficult to disentangle mode effects from the role of reflection.

There has been growing interest recently in measuring the randomness of choice responses in repeated discrete choice tasks. More specifically, researchers have begun using models that attempt to

estimate the scale parameter, which is inversely correlated to the variance of the error term, separately from the vector of attribute preferences in the random utility model (Fiebig et al. 2010). The scale parameter is an indicator of the weight of the deterministic relative to the error component in the random utility model. These models have been used to measure the effect on scale of the number and order of choice tasks (Czajkowski et al. 2014a), experience with the public good in question (Czajkowski et al. 2015), choice certainty (Beck et al. 2013), level of education (Czajkowski et al. 2014b), and attribute non-attendance (Scarpa et al. 2012). A number of studies examined whether respondent engagement, as proxied by observed response times in web-based surveys, is associated with less noisy responses (Holmes 1998, Haiijer et al. 2000, Rose and Black 2006, Brown et al. 2008, Bech et al. 2011, Campbell et al. 2013, Hess and Stathapoulos 2013, Börger 2016). With the exception of Bech et al. (2011), these studies have generally found that respondents who take more time to complete choice tasks have less random responses (i.e. higher scale) and are less likely to be respondents who ignore attributes. One weakness of the studies in demonstrating the effect of response time on scale is that response times are endogenous; respondents who are more likely to take sufficient time to answer questions may differ from rushed respondents in unobserved ways.

Another group of studies, beginning with Whittington et al. (1992), exogenously changed the amount of time provided for the task by giving a random subset of respondents overnight to think about the valuation scenario. This "time-to-think" (TT) approach was primarily done with single discrete choice referendum questions (Whittington et al.1992, 1993, Lauria et al.1999, Svedsäter 2007, Lee et al. 2015). Cook et al. (2007) gave half of respondents time to think about a series of discrete choice tasks, and Day et al. (2012) similarly provided "advanced disclosure" of choice sets to a subset to test ordering effects. Most studies find that willingness-to-pay (WTP) is lower with additional time to think; Cook et al. (2012) found a roughly 40% drop in average WTP for different vaccine types with TT across parallel research designs in four countries.

What is the effect of additional, exogenously-provided time on choice randomness? Svedsäter (2007) and Cook et al. (2012) found that respondents with more time provided by the survey protocol were more certain of their responses. Cook et al. (2007) further found that respondents with more time were less likely to make a pattern of choices that violated preference transitivity, stability or monotonicity. Day et al. (2012) found that advanced disclosure of the choice set mitigates some but not all of the robust ordering effects found among the subset without disclosure. Cook et al. (2007) detected statistically-significant differences in the variance of the error term (scale) using the Swait and Louviere (1993) test, but did not explore the issue further.

No study, however, has examined how exogenous changes in response times might shift *both* preferences and response consistency. In this paper we separate the effect of additional time on preferences from the effect on scale following the approach of Czajkowski et al. (2016) and adapting the generalized multinomial logit (GMNL) model of Fiebig et al. (2010). We re-analyze the dataset used in Cook et al. (2007), which measured household demand for cholera and typhoid vaccines in a

medium-sized city in central Vietnam, discussed more below. Although the valuation topic is in the domain of public health, we feel the methodological results will be of greater interest to stated preference practitioners who are more often environmental and resource economists. We find consistent evidence that respondents with time to think have a larger scale parameter: their responses are less noisy and better explained by the attributes of the vaccines displayed in the choice tasks. Nevertheless, even after accounting for scale, we find that time to think continues to shift preferences. In particular, giving respondents more time (to potentially confer with a spouse or family member) makes them more sensitive to price and more likely to opt-out and purchase neither of the two hypothetical vaccines on offer.

The remainder of the paper is structured as follows. Sections 2 and 3 introduce the research design and the choice modeling approach, respectively. Section 4 presents the results, and Section 5 provides some discussion and conclusions.

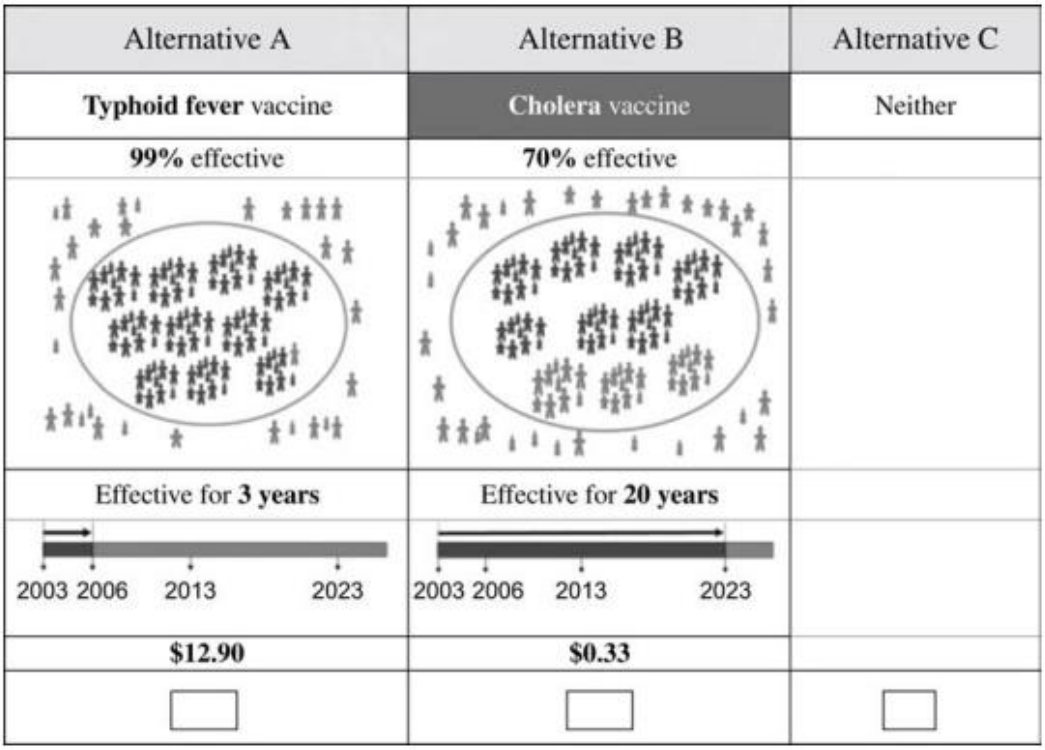
2. Research design

The data analyzed here come from a study, conducted in 2003, of household demand for "next generation" cholera and typhoid vaccines in Hue, a medium-sized city in central Vietnam (Cook et al. 2007). We refer readers to the original paper for more details on the study site, population and vaccine policy background, and briefly review here only the most relevant features of the study. The basic research design was a repeated, discrete choice task where respondents were asked to choose between a typhoid fever vaccine, a cholera vaccine or neither vaccine. In addition to vaccine type, the three attributes were the effectiveness of the vaccine (50%, 70%, or 99%), the duration of its effectiveness (3 or 20 years), and its financial price to the household (US$0.33, $3.22, and $12.90, in 2003 dollars). A stated-preference approach was appropriate because these "next generation" vaccines were generally unavailable at the time, and vaccine researchers were not yet certain of the actual effectiveness and duration of the vaccines in the field. The in-person household survey discussed vaccines, the risk of contracting cholera and typhoid, and the concept of vaccine protectiveness at length. A sample choice card is shown in Figure 1.

Subjects ($n = 400$) were asked to complete a total of six such choice cards. Four of these tasks were drawn from a main-effects, orthogonal task design that maximized statistical efficiency. Two additional, non-orthogonal, tasks were added specifically to test for "preference errors" (transitivity, stability, monotonicity). The order of six tasks was randomized at the subject level. Subjects were randomly assigned to complete one of three "blocks" of six tasks, or approximately $n = 133$ per block. Half of respondents (randomly-assigned, $n = 200$) completed all six choice tasks with the enumerator during the first interview after completing one practice task. This still represents standard practice for in-person stated preference surveys; we follow Cook et al. (2007) in referring to this as the "no time to think" treatment (NT). The other half were given the same introductory material and completed one choice task, but were then given the six cards overnight and encouraged to think about

their answers and discuss the choice with their spouse. The tasks were printed on laminated cards, and subjects were given markers that they could use to mark their choices. In all but one case, these "time to think" (TT) subjects were interviewed the next day. As Cook et al. (2007) noted, it is possible that respondents did not preserve the original, random ordering of the tasks but instead ordered them or looked at them all together, rather than serially.

Figure 1: Example choice card (from Cook et al.2007)



3. Choice modeling

Because the use of random utility models (McFadden 1974) is well-established, we focus our discussion on the generalized multinomial logit (GMNL) model framework (Fiebig et al. 2010, Gu et al. 2013). Respondent n derives utility from one of three alternatives i (vaccine A, vaccine B or neither) in six choice situations t according to

$$U_{nit} = \beta_n{}' x_{nit} + \varepsilon_{nit}. \quad (1)$$

x_{nit} is a vector containing alternative- and respondent-specific variables and β_n is the associated parameter vector. The elements of the coefficient vector are also referred to as utility weights since they indicate the influence of a particular attribute (as an element of the vector x_{nit}) on individual utility. Individual utility consists of a representative part, $\beta_n{}' x_{nit}$, observed by the analyst, and an idiosyncratic error term ε_{nit}. If ε_{nit} is independently and identically distributed according to the Type I Extreme

Value distribution, and further assuming that respondents choose the option that maximizes their utility, the probability that respondent n chooses alternative i in choice situation t is given by

$$P_{nit}|\beta_n = \frac{exp(\beta_n{}'x_{nit})}{\sum_{j=1}^{J} exp(\beta_n{}'x_{njt})} \quad \forall\, j = 1, \dots, i, \dots, J\,; t = 1, \dots, T\,; n = 1, \dots, N. \tag{2}$$

The random utility model allows for a number of specifications of the coefficient vector β_n relating to different assumptions about the structure of respondents' preferences. The GMNL framework nests several different well-known choice models as special cases when β_n, the vector of the individual coefficients in the sample, is specified as

$$\beta_n = \sigma_n\beta + \gamma\eta_n + (1-\gamma)\sigma_n\eta_n. \tag{3}$$

The elements of β_n may deviate from the sample mean β by η_n, which is a random variable with zero mean and standard deviation to be estimated. η_n serves to account for random heterogeneity in preferences, i.e. $Var(\eta_n)$. σ_n is a respondent-specific scale factor that shifts the whole vector of preference weights up or down in magnitude compared to the error term in the utility function (1). It is inversely correlated to the variance of that error term. γ is a weighting parameter constrained between 1 and 0, and discussed more below.

Different assumptions regarding the elements of (3) lead to different choice models, which represent differing conceptualizations of preference structures and heterogeneity across respondents. Assuming constant scale, i.e. $\sigma_n = 1$, leads to $\beta_n = \beta + \eta_n$, which is the random parameters logit (RPL) model (Revelt and Train 1998).[1] This model allows preference weights to vary randomly across respondents. η_n is assumed to be randomly distributed with zero mean and a standard deviation to be estimated. Different assumptions regarding the distributional form of η_n are possible. Many applications use the normal distribution, i.e. $\eta_n \sim N(0, \sigma)$.

Constraining neither the variance of η_n nor the scale factor σ_n defines the generalized multinomial logit model (GMNL). In this model, the parameter γ determines whether the scaling of σ_n affects only the means of the utility parameters (GMNL-I for $\gamma = 1$, and (3) simplifies to $\beta_n = \sigma_n\beta + \eta_n$) or both the mean and the standard deviations of the coefficients (GMNL-II for $\gamma = 0$, and (3) becomes $\beta_n = \sigma_n(\beta + \eta_n)$). In response to a criticism of the GMNL by Hess and Rose (2012) explained blow, we will use the GMNL-I model for the purposes of this analysis.[2]

The GMNL framework further allows for the parameterization of the scale factor as

$$\sigma_n = exp(\bar{\sigma} + \theta' z_n + \tau\varepsilon_{0n}). \tag{4}$$

[1] Note that in this case γ drops from equation (3) and hence is not defined.

[2] While γ can theoretically take any value between 0 and 1, constraining it to either of these values defines the two models, the interpretations of which are rather straightforward.

Since σ_n must not change sign but simply shifts the whole coefficient vector up or down, it is defined as the exponential function of a normalizing constant $\bar{\sigma}$, a systematic component $\theta' z_n$ and a random component $\tau\varepsilon_{0n}$. The latter follows a normal distribution, so the error term is normally distributed, i.e. $\varepsilon_{0n} \sim N(0,1)$, which in turn implies that the scale factor follows a lognormal distribution, i.e. $\sigma_n \sim LN(\bar{\sigma}, \tau^2)$ (Fiebig et al. 2010).

Importantly, the models allow for respondent-specific covariates of scale to be introduced in z_n. We use a dummy for the TT treatment as our explanatory variable. We employ a number of different versions of the GMNL model. In two models the random part of the scale parameter, τ, is constrained to zero so that scale is explained only by the respondent-specific TT variables, but in a third version τ is estimated freely. This is in response to the criticism by Hess and Rose (2012) that random preference ($Var(\eta_n)$) and random scale heterogeneity (τ) cannot be separated in the GMNL model. One way to address this is by constraining τ to zero, i.e. allowing only for random preference heterogeneity and explained (rather than random) scale heterogeneity by means of z_0. (We have also found that convergence is achieved more easily in this restricted model in Stata.) We do this in a model with $\gamma = 0$ and another specification with $\gamma = 1$. In this specification, scale is explained by the TT-dummy as a covariate of σ_n while preferences parameters are assumed to be correlated across respondents in the same treatment. These are essentially RPL models with scale determined by the TT dummy. Since the variance of the error term is inversely correlated to the scale factor σ_n, the former is effectively explained by TT. We will therefore refer to these models as heteroskedastic RPL. We also apply the GMNL-I model with a freely estimated τ to test whether a potential systematic effect of TT persists if σ_n is partly determined by a random component. This specification avoids the confounding of random preference and scale heterogeneity by setting $\gamma = 1$ so that the scaling only applies to the deterministic part of the preference weights, i.e. $\beta_n = \sigma_n \beta + \eta_n$.

In all four models the systematic effect of the time-to-think treatment on preferences is controlled for by interacting the treatment dummy with all but one of the vaccine attributes. To identify the model one attribute has to be excluded from these two-part interactions (Czajkowski et al. 2016). Based on the results of the mixed logit hierarchical Bayes models published earlier (Cook et al. 2007), as well as theory, we expect the effect of additional time to be strongest on price and the probability of opting out (the alternate-specific constant for no vaccine) because additional time may allow respondents to think more carefully about their budget constraints and the opportunity costs of buying one of the hypothetical vaccines. The effect of TT on preferences around the duration, effectiveness and type of vaccine are less clear *a priori*. Additional time may allow respondents to think about which of these attributes are most important to them, which may increase the utility weight of, say, duration, for some respondents while decreasing duration's utility weight for others. Similarly, some respondents may use additional time to reflect on the history of their family or friends with cholera or typhoid, leading some to weight vaccines against, say, cholera more heavily than those without TT, and some to

weight cholera vaccines less heavily. The models below explore different specifications for which attribute to leave out of the TT interaction, focusing on effectiveness, vaccine type and duration.

The introduction of an individual-specific scale factor creates potential correlation between attribute coefficients. In fact, scaled coefficients in the GMNL model are always correlated because they are all multiplied by the same scale factor σ_n.[3] The model therefore explicitly specifies the randomly-distributed preference coefficients to be correlated. However, because a two-treatment experimental design was used in the survey, this correlation pattern is constrained as follows. Attributes interacted with the treatment dummy are assumed to be independent when they are in different treatments, i.e. some of the off-diagonals of the Cholesky matrix are constrained to zero. See Appendix I for more details. We use 1,000 Halton draws to simulate the likelihood with the *gmnl* command in Stata 13 (Gu et al. 2013).

4. Results

4.1. Sample characteristics

Because of randomization, the samples are well-balanced on gender and age. By chance, however, the TT subsample was somewhat less educated than the NT subsample (Table 1). Because our hypothesis is that allowing more time to study the choices should increase scale and lower response noise, the effect of the time-to-think treatment on scale should be underestimated: we would also expect that respondents in the NT were already less likely to make choices at random because of their higher levels of education.

Table 1: Sample characteristics (from Cook et al. 2007)

	No time to think (NT)	**Time to think (TT)**
Observations	200	200
Age		
Mean age (years)	45	45
Age <35	9%	14%
Age 35-39	17%	16%
Age 40-44	24%	20%
Age 45-49	18%	20%
Age >50	33%	32%
Education		
Never attended school	4%	8%
Primary school	19%	23%
Secondary school	52%	59%
University and postgrad	26%	11%

[3] Note also the fact that RPL and GMNL models use simulated maximum likelihood, which leads to potentially correlated coefficients (Hensher et al. 2015).

4.2. Choice models

RPL model – Not accounting for differences in scale

We begin with the results of the basic RPL model in the left-hand column of Table 2. These parallel the original results in Cook et al. (2007), though are not exact replications since that paper used a mixed logit hierarchical Bayes approach estimated in GAUSS. The attribute variables are interacted with the TT treatment dummy (TT). This gives coefficients of, for instance, price in the time-to-think sample (TT*PRICE) and in the no-time-to-think sample (NT*PRICE). In order to identify the model we do not interact the effectiveness attribute (EFF70 and EFF99). All coefficients are assumed to be normally distributed and estimates of means and standard deviations are reported. The off-diagonal elements of the Cholesky matrix are omitted for brevity but available on request.

The RPL results suggest that giving additional time affects preference weights. For example, respondents are more sensitive to price with time to think: the price coefficient is almost twice as large in absolute magnitude in the TT treatment (TT*PRICE) than without time to think (NT*PRICE), a statistically-significant difference (Wald-test, $p = .002$). The ASC indicating preference for none of the vaccines changes from significantly negative without time to think (NT*ASC) to not statistically different from zero with time to think (TT*ASC): respondents with time to think are more likely to opt out and purchase neither of the two vaccines. This difference between treatments is statistically-significant (Wald-test: $p = .033$). The coefficient of preferences for the cholera vaccine over the typhoid vaccine (VACC_CH) is only significant in the NT treatment, whereas with time to think respondents do not prefer the cholera vaccine. This difference is not significant (Wald-test: $p = .369$). Preferences for a duration of the vaccines of 20 years differ across treatments with respondents with time to think preferring longer duration significantly, whereas this is insignificant in the NT treatment (Wald test: DUR20, $p = .003$). This model does not explicitly control for scale.

Heteroskedastic RPL models – Taking into account deterministic scale differences

The second and third column of Table 2 report heteroskedastic RPL models with $\gamma = 0$ and $\gamma = 1$, respectively. While these models account for random preference heterogeneity by assuming normally distributed coefficients, they do not allow for random heterogeneity in scale, i.e. τ is constrained to zero. Consequently the only potential determinant of scale is the treatment dummy (TT). This specification is used in response to the criticism by Hess and Rose (2012) that the GMNL cannot distinguish between random preference and scale heterogeneity.

In general, patterns of significant and insignificant coefficients are similar to the one in the basic RPL model, with some exceptions. The ASC is significantly negative in both treatments. Nevertheless, the preference for buying any vaccine is significantly weaker with time to think than without. While the significance pattern of all other coefficients is the same as in the RPL model,

differences in the price and duration coefficients between NT and TT in the GMNL model are smaller in magnitude compared to the basic RPL model. In the heteroskedastic RPL(0) and RPL(1) models respondents with time to think are still more sensitive to the price, though the differences are only significant at the 5%- and 10%-level, respectively (Wald-tests: $p = .034$ in the RPL(0) and $p = .073$ in the RPL(1) models). They are again more likely to opt-out and choose neither vaccine (Wald-test, ASC: $p < .001$). The fit of these models to the data is slightly better than the basic RPL model. Both models show a significant effect of TT on scale, confirming our hypothesis that time to think leads to more consistent, i.e. less random stated choices.

Table 2: RPL, heteroskedastic RPL and GMNL models excluding level of effectiveness (EFF70 and EFF99) from interaction with the TT treatment dummy

	RPL		**het. RPL (0)**		**het. RPL (1)**		**GMNL**	
	Coef.	**s.e.**	**Coef.**	**s.e.**	**Coef.**	**s.e.**	**Coef.**	**s.e.**
Mean of coefficients								
NT*PRICE	-0.438***	(0.064)	-0.367***	(0.053)	-0.484***	(0.069)	-0.397***	(0.057)
TT*PRICE	-0.847***	(0.136)	-0.528***	(0.081)	-0.678***	(0.117)	-0.594***	(0.104)
NT*ASC	-2.635***	(0.524)	-2.662***	(0.520)	-2.643***	(0.476)	-3.274***	(0.612)
TT*ASC	-0.946	(0.598)	-0.769***	(0.222)	-0.980**	(0.436)	-0.873***	(0.306)
NT*VACC_CH	0.387*	(0.213)	0.400***	(0.154)	0.384**	(0.181)	0.400**	(0.174)
TT*VACC_CH	0.135	(0.185)	0.098	(0.094)	0.166	(0.156)	0.167	(0.116)
NT*DUR20	0.148	(0.173)	0.139	(0.124)	0.179	(0.154)	0.173	(0.153)
TT*DUR20	0.919***	(0.218)	0.617***	(0.130)	0.671***	(0.179)	0.639***	(0.142)
EFF70	0.386**	(0.173)	0.245**	(0.098)	0.274*	(0.142)	0.286**	(0.128)
EFF99	3.196***	(0.379)	2.116***	(0.282)	2.966***	(0.420)	2.653***	(0.369)
Standard deviation of coefficients								
NT*PRICE	0.499***	(0.078)	0.419***	(0.068)	0.539***	(0.083)	0.425***	(0.075)
TT*PRICE	0.684***	(0.095)	0.440***	(0.069)	0.796***	(0.124)	0.822***	(0.156)
NT*ASC	7.553***	(1.017)	6.039***	(0.845)	7.140***	(0.912)	7.161***	(0.980)
TT*ASC	7.159***	(0.913)	4.604***	(0.677)	8.411***	(1.145)	9.655***	(1.380)
NT*VACC_CH	1.522***	(0.300)	1.149***	(0.191)	1.540***	(0.234)	1.365***	(0.236)
TT*VACC_CH	0.554**	(0.253)	0.566***	(0.119)	0.584***	(0.251)	1.034***	(0.257)
NT*DUR20	1.201***	(0.246)	0.792***	(0.183)	1.070***	(0.204)	0.972***	(0.197)
TT*DUR20	1.259***	(0.260)	0.848***	(0.141)	1.323***	(0.235)	1.271***	(0.267)
EFF70	1.362***	(0.270)	0.674***	(0.136)	1.334***	(0.238)	1.259***	(0.266)
EFF99	3.282***	(0.414)	2.461***	(0.320)	3.330***	(0.380)	2.991***	(0.399)
(off-diagonal elements of the Cholesky matrix)								
Covariates of scale								
TT			1.158***	(0.267)	0.272**	(0.130)	0.622***	(0.159)
/tau			0	(constr.)	0	(constr.)	-0.550***	(0.067)
/gamma			0	(constr.)	1	(constr.)	1	(constr.)
Log-likelihood	-1,493		-1,490		-1,490		-1,489	
Observations	7,200		7,200		7,200		7,200	
Halton draws	1,000		1,000		1,000		1,000	
Parameters	65		66		66		67	
Adjusted R^2	0.405		0.406		0.406		0.406	
BIC	3,376		3,375		3,376		3,380	

***, ** and * indicating the 1%-, 5%- and 10%-level of confidence. [a] interacted with ASC_CHANGE. Adjusted R^2 is computed as $R^2 = 1 - (LL_m - k)/LL_0$, where LL_m and LL_0 are the log-likelihoods of the full model, and the intercept-only model respectively, and k the number of parameters. Bayesian Information Criterion (BIC) is calculated as $BIC = -2LL_m + k \cdot \ln(N)$ with N denoting the number of respondents. The use of BIC was preferred to Akaike Information Criterion because it imposes a stronger penalty on the inclusion of more parameters in the model.

GMNL model – Allowing for random scale

The right-hand column of Table 2 shows the GMNL model with random scale, i.e. the parameter τ is estimated. Note that γ is constrained to 1, which yields the GMNL-I model, so that the scale parameter only applies to the coefficient means but not to their random components (Fiebig et al. 2010). We also ran models where γ was estimated freely or γ set to zero (the latter yields a model in which random scale is multiplied with random preference heterogeneity, which Hess and Rose (2012) argue cannot be separated), but these specifications did not converge.

The GMNL model finds both an effect of TT on scale and additionally significant random scale heterogeneity. This supports the earlier finding in Cook et al. (2007) that time to think induced respondents to give more consistent choices. The pattern of significance of the coefficients is the same as in the heteroskedastic RPL models. The difference between coefficients across treatments is significant for price, ASC and duration (Wald-tests: $p = .044$, $p < .001$ and $p = .015$, respectively) and insignificant for the cholera vaccine (Wald-test: $p = .263$)

As a robustness check, models excluding vaccine type (cholera vs. typhoid) and duration from the interaction with the TT dummy were run. Tables A.1 and A.3 in Appendix III show the corresponding models. While the exact pattern of significant and insignificant coefficients in these models differs to some extent, the model not interacting duration (Table A.1) shows the same systematic effect of time to think on mean scale. We do not find this effect, however, when the cholera vaccine is excluded from interaction with TT in Table A.3. Nevertheless, the respective GMNL models for both specifications also find a significant random component of scale (τ).

Our main finding from this analysis is that time to think affects responses via a second channel (scale): the effect of the covariate TT on the scale parameter σ is positive and statistically significant in two out of three model specifications. Time to think shifts the vector of attribute coefficients and leads to more consistent choices and a smaller weight of the error term in the estimated linear-additive underlying utility function. Respondents who are given time to think are able to state choices that are based more firmly on the content of the choice cards, namely the attributes and their respective levels. Working with such data, the analyst is able to explain a larger portion of utility provided by different types of vaccines with the observed data.

4.3. Effects of TT on WTP for vaccine types

The estimated parameters above can be used to compute WTP of different vaccine types. However, since all coefficients are assumed to be normally distributed in both the RPL and GMNL models the implicit price of attribute k cannot simply be computed as the negative ratio of its coefficient and the price coefficient, i.e. $WTP_k = -\beta_k {\beta_{Price}}^{-1}$. The resulting distribution of individual implicit prices would have infinite moments because the coefficient distribution in the denominator (β_{Price})

encompasses zero (Meijer and Rouwendahl 2006, Carson and Czajkowski 2013). We therefore employ a bootstrapping approach, which is inspired by the procedure in Czajkowski et al. (2016) but amended to some extent, to simulate median WTP for different vaccine types. 10^4 draws are taken from the multivariate normal distribution described by the mean of the estimated parameters and the asymptotic variance-covariance matrix of the maximum likelihood model. For each of these draws the individual-specific parameters (conditional on individual choice responses) are computed (Revelt and Train 2000, Campbell 2007). Using these coefficients, the respondent-specific ($n = 400$) WTP for each vaccine type is computed according to Appendix II and the medians stored. The means and 95% confidence intervals of these 10^4 simulated sample medians are displayed in Table 4 for the case of excluding the effectiveness attribute from the TT-dummy interaction. The same set of bootstrapped WTP estimates based on models excluding vaccine or effectiveness from the interaction are displayed in Tables A.2 and A.4 in Appendix III.

In Table 3, WTP for vaccine type increases with effectiveness and duration as in the original study in Cook et al. (2007). In general, WTP for any vaccine type is lower with time to think than without, with the notable exception of the cholera vaccine, 50% effectiveness and 20 years' duration in the RPL model. Vaccine types for which mean WTP in the TT treatment falls outside of the 95%-confidence interval of WTP in the NT treatment are marked with an asterisk. Table 3 shows the heteroskedastic RPL (0) separates best between the NT and TT treatments in terms of simulated WTP in this way. This picture is supported by WTP estimates simulated from models excluding duration and cholera vaccine from interaction with TT in Tables A.2 and A.4.

Table 3: WTP for vaccine types (model specification: no interaction between effectiveness and TT)

Vaccine type	RPL				het. RPL (0)			
	NT		TT		NT		TT	
	Mean	95% CI	Mean	95% CI	Mean	95% CI	Mean	95% CI
Chol., 50%, 3yr	0.59	[-0.03 - 2.94]	0.14	[-0.12 - 0.90]	5.55	[3.21 - 8.01]	0.44	[-0.06 - 1.20]*
Chol., 50%, 20yr	1.23	[-0.01 - 4.31]	1.30	[0.09 - 2.64]	6.33	[3.96 - 8.94]	1.99	[0.88 - 3.02]*
Chol., 70%, 3yr	9.21	[6.42 - 12.02]	4.35	[2.84 - 5.83]	11.50	[9.06 - 14.35]	4.96	[4.05 - 5.84]*
Chol., 70%, 20yr	10.25	[7.67 - 13.00]	6.38	[5.16 - 7.67]*	12.10	[9.53 - 15.16]	6.87	[6.03 - 7.78]*
Chol., 99%, 3yr	11.74	[8.83 - 14.81]	6.93	[5.54 - 8.43]*	13.85	[11.18 - 17.05]	7.51	[6.74 - 8.33]*
Chol., 99%, 20yr	12.33	[9.60 - 15.29]	8.30	[7.04 - 9.72]*	14.28	[11.63 - 17.73]	8.20	[7.44 - 9.15]*
Typh., 50%, 3yr	0.32	[-0.02 - 1.77]	0.02	[-0.15 - 0.46]	3.61	[1.14 - 6.17]	0.30	[-0.10 - 0.96]*
Typh., 50%, 20yr	0.89	[0.00 - 3.43]	0.95	[0.00 - 2.18]	4.40	[1.81 - 7.06]	1.74	[0.63 - 2.76]*
Typh., 70%, 3yr	8.13	[5.56 - 10.69]	3.97	[2.54 - 5.38]*	9.54	[7.09 - 12.36]	4.64	[3.83 - 5.42]*
Typh., 70%, 20yr	9.00	[6.29 - 11.87]	5.93	[4.70 - 7.25]*	10.18	[7.51 - 13.24]	6.58	[5.77 - 7.45]*
Typh., 99%, 3yr	10.51	[7.85 - 13.41]	6.49	[5.13 - 7.97]*	11.83	[9.34 - 14.92]	7.35	[6.50 - 8.26]*
Typh., 99%, 20yr	10.92	[8.40 - 13.77]	7.78	[6.56 - 9.25]*	12.29	[9.69 - 15.68]	8.07	[7.19 - 9.15]*

Vaccine type	het. RPL (1)				GMNL			
	NT		TT		NT		TT	
	Mean	95% CI	Mean	95% CI	Mean	95% CI	Mean	95% CI
Chol., 50%, 3yr	0.84	[-0.04 - 2.97]	0.01	[-0.09 - 0.27]	2.66	[0.00 - 5.54]	0.00	[-0.03 - 003]
Chol., 50%, 20yr	1.77	[-0.01 - 4.45]	0.71	[0.00 - 1.85]	3.56	[0.64 - 6.41]	0.18	[0.00 - 0.82]*
Chol., 70%, 3yr	8.72	[6.66 - 11.01]	3.95	[2.47 - 5.31]*	11.12	[8.52 - 13.89]	2.91	[1.50 - 4.26]*
Chol., 70%, 20yr	9.62	[7.59 - 11.94]	5.90	[4.73 - 7.10]*	11.82	[9.20 - 14.67]	4.71	[3.13 - 6.04]*
Chol., 99%, 3yr	12.34	[10.14 - 14.89]	6.75	[5.43 - 8.14]*	15.36	[12.52 - 18.69]	5.73	[4.01 - 7.21]*
Chol., 99%, 20yr	12.73	[10.56 - 15.37]	8.28	[6.99 - 9.70]*	15.90	[13.02 - 19.30]	7.32	[5.78 - 8.73]*
Typh., 50%, 3yr	0.36	[-0.04 - 1.73]	-0.02	[-0.10 - 0.01]	1.33	[-0.02 - 3.78]	-0.01	[-0.04 - 0.00]
Typh., 50%, 20yr	1.09	[0.00 - 3.26]	0.45	[-0.01 - 1.40]	2.09	[0.00 - 4.77]	0.07	[-0.01 - 0.46]
Typh., 70%, 3yr	7.40	[5.53 - 9.47]	3.60	[2.32 - 4.76]*	9.49	[6.97 - 12.20]	2.43	[1.10 - 3.71]*
Typh., 70%, 20yr	8.17	[6.12 - 10.48]	5.45	[4.36 - 6.56]*	10.22	[7.57 - 13.00]	4.16	[2.64 - 5.51]*
Typh., 99%, 3yr	10.87	[8.87 - 13.27]	6.25	[4.99 - 7.61]*	13.77	[11.09 - 16.90]	5.14	[3.45 - 6.65]*
Typh., 99%, 20yr	11.26	[9.23 - 13.80]	7.78	[6.50 - 9.17]*	14.28	[11.54 - 17.55]	6.77	[5.18 - 8.23]*

Unit: 2003 USD. CI - Confidence interval. * TT mean falls outside of NT confidence interval.

5. Discussion and conclusions

This paper set out to examine the effects on preference patterns and scale of giving respondents in a discrete choice experiment time to think. We made use of an existing dataset assessing values of potential cholera and typhoid vaccines in the Vietnamese city of Hue. There are essentially two main results of this analysis. First, there is substantial evidence that time to think has a positive effect on scale. In econometric terms, giving respondents time to think increases choice consistency in the sense that the relative weight of the deterministic part of the underlying indirect utility function increases relative to the error component. Hence the analyst is able to make better predictions of choice probabilities. To identify the model, we needed to exclude at least one set of vaccine attributes from the interaction with TT. We chose to test models excluding attributes for which there was no clear expectation that they might be influenced by time to think. We find the main result of the positive scale effect of TT when we exclude duration and effectiveness, though not when we exclude vaccine type.

On the level of the survey interview, these results indicate that respondents who are given one night's time to think about the choice tasks take into account the attributes and their levels more when stating their choices. So in addition to the effects of time to think already established in Cook et al. (2007), such as fewer violations of internal validity of utility theory, higher stated certainty, lower demand for vaccines and higher price elasticity, we also find a positive effect on choice consistency. Respondents who are given time to think state, in effect, less random choices.

Whether this main result implies that respondents really absorb and process this information differently, however, is unclear based on the above analysis. A potential way to shed light onto this issue is to study the effect of time to think on attribute attendance, as it has been shown that response time increases the likelihood of a respondent to have considered more attributes when completing the choice tasks (Börger 2016). While objective attribute attendance patterns can be established by means of equality-constrained latent class models (Scarpa et al. 2009, 2012), this dataset does not include any questions pertaining to whether respondents attended to a particular attribute or not, which would be necessary to investigate stated attribute non-attendance (e.g. Hensher et al. 2005, Colombo et al. 2013). Another potential way to investigate this question is the innovative use of eye-tracking technology for stated preference surveys (Uggeldahl et al. 2016).

As a second main result we found that preference parameters (and their differences between treatments) change when scale is controlled for in the heterogeneous RPL and GMNL models compared to the RPL where the scale effect is not modeled. For most attributes, the difference in preference weights between the NT and TT treatments decreases when scale is controlled for. This implies that the differences in utility weights caused by providing time to think are only partially a result of the shifting of preferences but are also due to a significant difference in scale between the treatments. Earlier studies could therefore have attributed too much of the effect of TT on preferences alone, and the headline changes in total WTP could be overstated (i.e. approximately 40% in Cook et al. 2012). Recall, though, that most existing TT studies, including three of the four in Cook et al. 2012, used single discrete choices (i.e. contingent valuation) where scale and choice randomness may be somewhat less of a concern. Furthermore, we find that providing TT causes substantial drops in mean WTP in this dataset, even when the effect of TT on scale is controlled for. Appendix Table A.5 calculates the percentage decrease in bootstrapped mean WTP for a given vaccine type moving from NT to TT, for all models. We do not report the percentage change when the WTP distributions are not statistically different (95% confidence intervals overlap). Among cases where the distributions are different, WTP is always lower in the TT subsample, and the percentage decreases range from 29% to 100% (the mean WTP for the vaccine with TT is zero). In our preferred specification (excluding effectiveness from the TT interaction), and counting overlapping distributions as a zero percentage change, the average decrease across all vaccine types is 39%.

In addition to these two main outcomes, this study also sheds some light on the debate around the GMNL model. While it has been argued that the model cannot distinguish between random scale and preference heterogeneity (Hess and Rose 2012), our study shows that regardless of this issue, the model is capable of identifying systematic differences in scale. This is achieved by either excluding random scale heterogeneity from the model or – if scale is assumed to be random – by scaling only the deterministic component of the utility weights.

Given our results, as well as those from earlier studies, why are so few researchers implementing the TT protocol? After all, a decrease in average WTP of even 20% could have important policy implications where valuation estimates are being used, for example, to assess natural resource damages or evaluate large infrastructure investments. As discussed in Cook et al (2012), there are logistical concerns in implementing the protocol for in-person interviews since NT interviews must be done first to prevent the NT respondents from learning about the survey from others and implicitly getting time to think. This would not be a concern, though, if all respondents were given time to think. There is also the obvious concern that two shorter in-person interviews will be costlier than one longer interview because of the time and hassle in arranging a second interview and traveling to the home. On the other hand, if TT improves the consistency of responses as shown above, it may lower standard errors and enable to practitioners to reduce sample sizes of surveys and lower the variable costs of a survey effort.

A deeper argument against the protocol, however, would be that researchers wish to observe preferences that mimic how citizens might vote in the types of hypothetical referenda that public goods valuation techniques began with. Suppose a jurisdiction were voting on a proposal to spend public money to restore a major waterway to improve habitat for fish and other wildlife. Would voters examine the information in advance, thinking carefully about the impacts the program might have on wildlife, how credible the program is, and the expected increase in their tax burden and subsequent loss of income? Or would voters walk into the booth knowing nothing about the proposal and make a snap decision based only on the ballot language? If the answer is the latter, then one could argue that a TT protocol artificially induces reflection that would not be present in reality, and thus artificially lowers WTP. In reality, of course, there will be a distribution of voter behavior that would vary by person and by issue. Some voters will be poorly-prepared for nearly all decisions at the ballot box, others think carefully in advance about even the most mundane elected office, and others think very carefully in advance about issues they care about but barely think about others. We argue that giving TT most accurately reflects the potential for citizens to deliberate about decisions, or for that matter for consumers to deliberate about the decision to purchase private goods. Some TT respondents will ignore the opportunity to reflect, depending on their time constraints or motivation about the survey topic, just like they would at the ballot box. Those who would reflect on an issue in reality are likely to take the same opportunity in the TT survey protocol.

Finally, the increasing use of online formats is implicitly giving respondents additional time, just like mail surveys. As noted in the introduction, a number of recent studies have found that longer internet response times are associated with less noisy choices, though these response times are endogenous and could have serious measurement error if longer times are due not to reflection but inattention and multi-tasking on the device the survey is being taken on. We believe a split-sample, online DCE study imposing some minimum reflection period, whether 10 minutes or 1 day, would be a useful contribution to the literature.

Acknowledgments

We thank Dale Whittington for helpful comments, Arne Risa Hole for assistance with implementing his gmnl command in Stata, and Mikolaj Czajkowski for comments and his generosity with both his Matlab code (freely available on the internet) and his time in helping us adapt it for our application. All errors are of course our own.

References

Bech, M., T. Kjaer, and J. Lauridsen (2011) Does the number of choice sets matter? Results from a web survey applying a discrete choice experiment. *Health Economics* 20, 273-286.

Beck, M. J., J.M. Rose and D.A. Hensher (2013) Consistently inconsistent: The role of certainty, acceptability and scale in choice. *Transportation Research Part E: Logistics and Transportation Review* 56, 81-93.

Börger, T. (2016) Are fast responses more random? Testing the effect of response time on scale in an online choice experiment. *Environmental and Resource Economics* 65(2), 389-413.

Brown, T. C., D. Kingsley, G. L. Peterson, N. E. Flores, A. Clarke and A. Birjulin (2008) Reliability of individual valuations of public and private goods: Choice consistency, response time, and preference refinement. *Journal of Public Economics* 92, 1595-1606.

Campbell, D. (2007) Willingness to Pay for Rural Landscape Improvements: Combining Mixed Logit and Random-Effects Models. *Journal of Agricultural Economics* 58 (3), 467-483.

Campbell, D., M.R. Morkbak and S.D. Olsen (2013) How quick can you click? The role of response time in online stated choice experiments. Bioecon Conference Cambridge, UK.

Canh, D.G., D. Whittington, L.T.K. Thoa, N. Utomo, N.T. Hoa, C. Poulos, D.T.D. Thuy, D. Kim and A. Nyamete (2006) Household demand for typhoid fever vaccines in Hue, Vietnam. *Health Policy and Planning* 21(3), 241–255.

Carson, R.T. and M. Czajkowski (2013) A new baseline model for estimating willingness to pay from discrete choice models. Paper presented at the International Choice Modelling Conference 2013, Sydney.

Colombo, S., M. Christie and N. Hanley (2013) What are the consequences of ignoring attributes in choice experiments? Implications for ecosystem service valuation. *Ecological Economics*, 96, 25-35.

Cook, J., D. Whittington, D.G. Canh, F.R. Johnson and A. Nyamete (2007) Reliability of stated preferences for cholera and typhoid vaccines with time to think in Hue, Vietnam. *Economic Inquiry* 45(1), 100-114.

Cook, J., M. Jeuland, B. Maskery and D. Whittington (2012) Giving stated preference respondents "Time to Think": Results from four countries. *Environmental and Resource Economics* 51, 473-496.

Czajkowski, M., M. Giergiczny and W.H. Greene (2014a) Learning and fatigue effects revisited: Investigating the effects of accounting for unobservable preference and scale heterogeneity. *Land Economics* 90(2), 324–351.

Czajkowski, M., T. Kądziela and N. Hanley (2014b) We want to sort! Assessing households' preferences for sorting waste. *Resource and Energy Economics* 36, 290-306.

Czajkowski, M., N. Hanley and J. LaRiviere (2015) The effects of experience on preference uncertainty: Theory and empirics for public and quasi-public environmental goods. *American Journal of Agricultural Economics* 97, 333-351.

Czajkowski, M., N. Hanley and J. LaRiviere (2016) Controlling for the effects of information in a public goods discrete choice model. *Environmental and Resource Economics* 63, 523-544.

Day, B., I.J. Bateman, R.T. Carson, D. Dupont, J.J. Louviere, S. Morimoto, R. Scarpa, P. Wang (2012) Ordering effects and choice set awareness in repeat-response stated preference studies. *Journal of Environmental Economics and Management* 63(1), 73–91.

Gu, Y., A.R. Hole, S. Knox (2013) Estimating the generalized multinomial logit model in Stata. *The Stata Journal* 13(2): 382-397.

Fiebig, D.G., M.P. Keane, J.J. Louviere and N. Wasi (2010) The generalized multinomial logit model: Accounting for scale and coefficient heterogeneity. *Marketing Science* 29, 393-421.Gu, Y., A.R. Hole and S. Knox (2013) Estimating the generalized multinomial logit model in Stata. *The Stata Journal* 13(2), 382-397.

Haaijer, R., W. Kamakura and M. Wedel (2000) Response latencies in the analysis of conjoint choice experiments. *Journal of Marketing Research* 37, 376-382.

Hensher, D.A., J.M. Rose and W.H. Greene (2005) The implications on willingness to pay of respondents ignoring specific attributes. *Transportation*, 32, 203-222.

Hensher, D.A., J.M. Rose and W.H. Greene (2015) Applied choice analysis, Cambridge University Press, Cambridge.

Hess, S. and J.M. Rose (2012) Can scale and coefficient heterogeneity be separated in random coefficients models? *Transportation* 39(6), 1225-1239.

Hess, S. and A. Stathopoulos (2013) Linking response quality to survey engagement: A combined random scale and latent variable approach. *Journal of Choice Modelling* 7, 1-12.

Holmes, T., K. Alger, C. Zinkhan and E. Mercer (1998) The effect of response time on conjoint analysis estimates of rainforest protection values. *Journal of Forest Economics* 4, 7-28.

Lauria, D.T., D. Whittington, K. Choe, C. Turingan and V. Abiad. (1999) Household demand for improved sanitation services: A case study of Calamba, Phillipines. In: K. Willis and I. Bateman (eds.) *Valuing environmental preferences: theory and practice of the contingent valuation method*. Oxford, Oxford University Press, 540–584.

Lee, J.-S., V. Mogasale, J.K. Lim, M. Carabali, C. Sirivichayakul, D.D. Anh, K.S. Lee, V.D. Thiem, K. Limkittikul, L.H. Tho, I.D. Velez, J.E. Osorio, P. Chanthavanich, L.J. da Silva and B. Maskery (2015) A multi-country study of the household willingness-to-pay for dengue vaccines: Household surveys in Vietnam, Thailand, and Colombia. *PLOS Neglected Tropical Diseases* 9(6), e0003810. http://doi.org/10.1371/journal.pntd.0003810.

Kim, D., D.G. Canh, C. Poulos, L.T.K. Thoa, J. Cook, N.T. Hoa, A. Nyamete, D.T.D. Thuy, J. Deen, D.D. Trach, J. Clemens, V.D. Thiem, D.D. Anh and D. Whittington (2008) Private demand for cholera vaccines in Hue, Vietnam. *Value in Health* 11(1), 119–128.

McFadden, D. (1974) Frontiers in econometrics, economic theory and mathematical economics. In: P. Zarembka, P. (eds), Academic Press, New York, 105-142.

Meijer, E. and J. Rouwendal (2006) Measuring welfare effects in models with random coefficients. *Journal of Applied Econometrics* 21, 227-244.

Revelt, D. and K. Train (1998) Mixed logit with repeated choices: Households' choices of appliance efficiency level. *Review of Economics and Statistics* 80(4), 647-657.

Revelt, D. and K. Train (2000) Customer-specific taste parameters and mixed logit: Households' choice of electricity supplier. University of California, Berkeley.

Rose, J. and I. Black (2006) Means matter, but variance matter too: Decomposing response latency influences on variance heterogeneity in stated preference experiments. *Marketing Letters* 17, 295-310.

Scarpa, R., T. J. Gilbride, D. Campbell and D. A. Hensher (2009) Modelling attribute non-attendance in choice experiments for rural landscape valuation. *European Review of Agricultural Economics* 36, 151-174.

Scarpa, R., R. Zanoli, V. Bruschi and S. Naspetti (2012) Inferred and stated attribute non-attendance in food choice experiments. *American Journal of Agricultural Economics* 95(1), 165-180.

Svedsäter, H. (2007) Ambivalent statements in contingent valuation studies: inclusive response formats and giving respondents time to think. *Australian Journal of Agricultural and Resource Economics* 51, 91-107.

Swait, J. and J.J. Louviere (1993) The role of the scale parameter in the estimation and comparison of multinomial logit models. *Journal of Marketing Research* 30, 305-314.

Uggeldahl, K, C. Jacobsen, T.H. Lundhede, S.B. Olsen (2016) Choice certainty in discrete choice experiments: Will eye tracking provide useful measures? *Journal of Choice Modelling* 20, 35-48.

Whittington, D., V.K. Smith, A. Okorafor, A. Okore, J.L. Liu and A. McPhail (1992) Giving respondents time to think in contingent valuation studies: A developing country application. *Journal of Environmental Economics and Management* 22, 205-225.

Whittington, D., D.T. Lauria, K. Choe, J.A. Hughes and V. Swarna (1993) Household sanitation in Kumasi, Ghana: A description of current practices, attitudes and perceptions. *World Development* 21(5), 733–748.

Appendix I: Constraining correlation patterns between attribute coefficients

A treatment dummy indicating the TT and NT samples is created and interacted with all but one choice attributes. Since this dummy is also used as covariate of scale, interacting with all attributes would lead to an unidentified model (Czajkowski et al. 2016). This differentiation between an attribute in the two treatments means any interacted coefficient represents the effect of only those choices made by respondents in that respective treatment. Consequently, in any RPL model with correlated coefficients only those choice attributes that were displayed in the same treatment can possibly be correlated. At the same time, any treatment-specific attribute coefficient is potentially correlated with the non-interacted attribute. Hence certain pairwise correlations between variables in the model need to be constrained to zero. In practical estimation, these constraints are imposed on elements of the lower triangular Cholesky matrix C, yet what is effectively to be constrained are the corresponding elements of the variance-covariance matrix VC. Therefore a link between the constraints imposed on the lower triangular of C and the corresponding elements of VC needs to be established.

Due to the particular set of constraints required for the present analysis the order in which the attributes enter the choice model matters as illustrated by the following examples. In example 1, the model has five variables: $v1$, $v2$, $v3$, $v4$ and $v5$. $v1$ and $v3$ are attribute variables in treatment 1 while $v2$ and $v4$ are correlated attributes in treatment 2. $v5$ is the non-interacted variable and hence correlated to attributes in both treatments. With the element a_{ij} denoting the correlation between variables vi and vj the constrained Cholesky matrix is

$$C_1 = \begin{bmatrix} a_{11} & 0 & 0 & 0 & 0 \\ 0 & a_{22} & 0 & 0 & 0 \\ a_{31} & 0 & a_{33} & 0 & 0 \\ 0 & a_{42} & 0 & a_{44} & 0 \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{bmatrix}. \quad \text{(A.1)}$$

Given that the relationship between C and VC is $C \times C' = VC$, the pattern of zero and non-zero elements of C_1 only translates into VC_1 consistently if the non-interacted (i.e. unconstrained) variable is listed last. In this case, the variance-covariance matrix is

$$VC_1 = \begin{bmatrix} {a_{11}}^2 & 0 & a_{31}a_{11} & 0 & a_{51}a_{11} \\ 0 & {a_{22}}^2 & 0 & a_{22}a_{42} & a_{22}a_{52} \\ a_{31}a_{11} & 0 & {a_{31}}^2 + {a_{33}}^2 & 0 & a_{51}a_{31} + a_{53}a_{33} \\ 0 & a_{42}a_{22} & 0 & {a_{42}}^2 + {a_{44}}^2 & a_{52}a_{42} + a_{54}a_{44} \\ a_{51}a_{11} & a_{52}a_{22} & a_{51}a_{31} + a_{53}a_{33} & a_{52}a_{42} + a_{54}a_{44} & {a_{51}}^2 + {a_{52}}^2 + {a_{53}}^2 + {a_{54}}^2 + {a_{55}}^2 \end{bmatrix}. \quad \text{(A.2)}$$

The pattern of zero and non-zero (i.e. to be estimated in the model) elements of the lower triangular of VC_1 is the same as in C_1. The importance of the order of variables in the model is demonstrated in Example 2. If the variable order is different, say $v5 - v1 - v3 - v2 - v4$, and the correlation pattern is the same as in Example 1, the Cholesky matrix with the appropriate constraints is

$$C_2 = \begin{bmatrix} a_{55} & 0 & 0 & 0 & 0 \\ a_{15} & a_{11} & 0 & 0 & 0 \\ a_{25} & 0 & a_{22} & 0 & 0 \\ a_{35} & a_{31} & 0 & a_{33} & 0 \\ a_{45} & 0 & a_{42} & 0 & a_{44} \end{bmatrix} \quad \text{and} \quad C_2{}' = \begin{bmatrix} a_{55} & a_{15} & a_{25} & a_{35} & a_{45} \\ 0 & a_{11} & 0 & a_{31} & 0 \\ 0 & 0 & a_{22} & 0 & a_{42} \\ 0 & 0 & 0 & a_{33} & 0 \\ 0 & 0 & 0 & 0 & a_{44} \end{bmatrix}. \tag{A.3}$$

This matrix includes constraints on the same pairs of variables as in Example 1. It results in a variance-covariance matrix

$$VC_2 = \begin{bmatrix} {a_{55}}^2 & a_{15}a_{55} & a_{25}a_{55} & a_{35}a_{55} & a_{45}a_{55} \\ a_{15}a_{55} & {a_{15}}^2 + {a_{11}}^2 & a_{25}a_{15} & a_{35}a_{15} + a_{31}a_{11} & a_{45}a_{15} \\ a_{25}a_{55} & a_{25}a_{15} & {a_{25}}^2 + {a_{22}}^2 & a_{35}a_{25} & a_{45}a_{25} + a_{54}a_{22} \\ a_{35}a_{55} & a_{35}a_{15} + a_{31}a_{11} & a_{35}a_{25} & {a_{35}}^2 + {a_{31}}^2 + {a_{33}}^2 & a_{45}a_{35} \\ a_{35}a_{55} & a_{45}a_{15} & a_{45}a_{25} + a_{42}a_{22} & a_{45}a_{35} & {a_{45}}^2 + {a_{42}}^2 + {a_{44}}^2 \end{bmatrix}. \tag{A.4}$$

It is evident that the pattern of zero and non-zero elements of the lower triangular of VC_2 is not the same as in the Cholesky matrix C_2. Finally note that the order of the interacted variables ($v1$, $v2$, $v3$ and $v4$) is irrelevant to this result. That is, a Cholesky matrix with, say, the variable order $v1 - v3 - v2 - v4$ carries its pattern of zero and non-zero elements over into the respective variance-covariance matrix.

Appendix II: Computation of median WTP for vaccine types

WTP for cholera and typhoid vaccines with different characteristics are simulated because both the attribute and price coefficients are assumed to be normally distributed. In a first step, attribute coefficients have to be transformed because effects coding is used for all but the price attribute. The adjusted ASC coefficient is

$$\beta_{ASC\_adj} = \beta_{ASC} - (\beta_{Price} \cdot \mu_{Price}) \tag{A.5}$$

where μ_{Price} is the sample mean of the price variable, 3.22 in our data set. In addition, coefficients for the implicit levels of the non-monetary attributes, i.e. the typhoid vaccine (β_{ty}), an effectiveness level of 50% (β_{eff50}) and the duration of 3 years (β_{d3}), have to be computed from the estimated coefficients as follows.

$$\beta_{ty} = -\beta_{ch} \tag{A.6}$$

$$\beta_{eff50} = -(\beta_{eff70} + \beta_{eff99}) \tag{A.7}$$

$$\beta_{d3} = -\beta_{d20} \tag{A.8}$$

With this complete set of coefficients, WTP of a forced choice between the no-buy option a single vaccine of type v, with effectiveness level eff and duration d can be computed as

$$WTP^{fc}_{v,eff,d} = -\frac{\beta_v + \beta_{eff} + \beta_d - ASC_{adj}}{\beta_{Price}}. \tag{A.9}$$

To arrive at the expected WTP for any type of vaccine we further need the probability of not opting out for that generic vaccine as

$$P_{v,eff,d}(not\ opt\ out) = 1 - \frac{exp(ASC_{adj})}{exp(ASC_{adj}) + exp(\beta_v + \beta_{eff} + \beta_d)} \tag{A.10}$$

Based on (A.9) and (A.10), expected WTP of a vaccine of type v, with effectiveness level eff and duration d is

$$E(WTP_{v,eff,d}) = P_{v,eff,d}(not\ opt\ out) \cdot WTP^{fc}_{v,eff,d}$$

$$E(WTP_{v,eff,d}) = \left[1 - \frac{exp(ASC_{adj})}{exp(ASC_{adj}) + exp(\beta_v + \beta_{eff} + \beta_d)}\right] \cdot \left[-\frac{\beta_v + \beta_{eff} + \beta_d - ASC_{adj}}{\beta_{Price}}\right]. \tag{A.11}$$

Appendix III: Alternative models excluding vaccine and effectiveness from interaction with the TT treatment dummy

Table A.1: RPL, heteroskedastic RPL, and GMNL models excluding duration (DUR) from interaction with the TT treatment dummy

	RPL		het. RPL (0)		het. RPL (1)		GMNL	
	Coef.	**s.e.**	**Coef.**	**s.e.**	**Coef.**	**s.e.**	**Coef.**	**s.e.**
Mean of coefficients								
NT*PRICE	-0.437***	(0.063)	-0.364***	(0.055)	-0.418***	(0.062)	-0.478***	(0.074)
TT*PRICE	-0.886***	(0.146)	-0.335***	(0.098)	-0.278	(0.191)	-0.144	(0.156)
NT*ASC	-3.341***	(0.552)	-2.829***	(0.564)	-2.722***	(0.532)	-3.523***	(0.649)
TT*ASC	-1.012	(0.639)	-0.337	(0.208)	-0.551	(0.400)	-0.283	(0.320)
NT*EFF70	0.022	(0.210)	0.029	(0.172)	0.020	(0.217)	0.010	(0.220)
TT*EFF70	0.881***	(0.339)	0.359**	(0.150)	0.229	(0.171)	0.104	(0.123)
NT*EFF99	3.036***	(0.433)	2.361***	(0.340)	2.786***	(0.400)	2.834***	(0.432)
TT*EFF99	3.782***	(0.645)	1.556***	(0.452)	1.088	(0.735)	0.624	(0.680)
NT*VACC_CH	0.332*	(0.176)	0.216	(0.140)	0.440**	(0.175)	0.475**	(0.191)
TT*VACC_CH	0.364*	(0.211)	0.064	(0.075)	0.087	(0.081)	0.062	(0.073)
DUR20	0.489***	(0.130)	0.274***	(0.089)	0.262	(0.166)	0.154	(0.164)
Standard deviation of coefficients								
NT*PRICE	0.568***	(0.100)	0.406***	(0.068)	0.473***	(0.076)	0.493***	(0.085)
TT*PRICE	0.444***	(0.101)	0.334***	(0.098)	0.778***	(0.115)	0.761***	(0.123)
NT*ASC	7.218***	(1.092)	6.218***	(0.907)	6.748***	(0.958)	7.265***	(1.163)
TT*ASC	10.448***	(1.760)	3.790***	(1.149)	9.308***	(1.488)	9.586***	(1.431)
NT*EFF70	1.248***	(0.416)	0.619**	(0.254)	0.911***	(0.303)	0.741***	(0.260)
TT*EFF70	0.610*	(0.357)	0.566***	(0.202)	1.383***	(0.354)	1.075***	(0.340)
NT*EFF99	3.006***	(0.567)	2.307***	(0.349)	3.300***	(0.515)	2.865***	(0.430)
TT*EFF99	3.776***	(0.756)	1.703***	(0.508)	4.196***	(0.686)	3.941***	(0.646)
NT*VACC_CH	1.258***	(0.281)	1.121***	(0.181)	1.415***	(0.237)	1.450***	(0.257)
TT*VACC_CH	0.470	(0.345)	0.496***	(0.166)	0.885***	(0.279)	1.275***	(0.264)
DUR20	1.513***	(0.229)	0.732***	(0.201)	1.207***	(0.177)	1.212***	(0.176)
(off-diagonal elements of the Cholesky matrix)								
Covariates of scale								
TT			1.305***	(0.351)	1.276*	(0.680)	1.926*	(1.089)
/tau			0	(constr.)	0	(constr.)	-0.594***	(0.089)
/gamma			0	(constr.)	1	(constr.)	1	(constr.)
Log-likelihood	-1,480		-1,487		-1,491		-1,485	
Observations	7,200		7,200		7,200		7,200	
Halton draws	1,000		1,000		1,000		1,000	
Parameters	77		78		78		79	
Adjusted R^2	0.406		0.403		0.402		0.403	
BIC	3,422		3,441		3,449		3,444	

***, ** and * indicating the 1%-, 5%- and 10%-level of confidence. [a] interacted with ASC_CHANGE. Adjusted R^2 is computed as $R^2 = 1 - (LL_m - k)/LL_0$, where LL_m and LL_0 are the log-likelihoods of the full model, and the intercept-only model respectively, and k the number of parameters. Bayesian Information Criterion (BIC) is calculated as $BIC = -2LL_m + k \cdot \ln(N)$ with N denoting the number of respondents. The use of BIC was preferred to Akaike Information Criterion because it imposes a stronger penalty on the inclusion of more parameters in the model.

WALD-tests for the RPL model in Table A.3 are as follows: PRICE: $p = .004$; ASC: $p = .016$; EFF70: $p = .025$, and EFF99: $p = .101$ and VACC_CH: $p = .791$.For the heteroskedastic RPL(0) model they are PRICE: $p = .760$; ASC: $p < .001$; EFF70: $p = .137$, and EFF99: $p = .055$ and VACC_CH: $p = .335$; and for the heteroskedastic RPL(1) model they are PRICE: $p = .469$; ASC: $p < .001$; EFF70: $p = .477$, and EFF99: $p = .023$ and VACC_CH: $p = .038$. For the coefficients in the GMNL model WALD-tests are: PRICE: $p = .052$, ASC: $p < .001$, EFF70: $p = .726$, and EFF99: $p = .002$ and VACC_CH: $p = .064$.

Table A.2: WTP for vaccine types (model specification: no interaction between duration and TT)

Vaccine type	RPL				het. RPL (0)			
	NT		TT		NT		TT	
	Mean	95% CI	Mean	95% CI	Mean	95% CI	Mean	95% CI
Chol., 50%, 3yr	3.03	[0.79 - 5.24]	0.57	[0.00 - 2.07]*	3.81	[1.34 - 6.37]	0.08	[-0.20 - 0.83]*
Chol., 50%, 20yr	2.37	[-0.02 - 5.13]	0.32	[-0.01 - 1.63]	3.77	[1.28 - 6.37]	0.01	[-0.21 - 0.49]*
Chol., 70%, 3yr	8.33	[5.99 - 10.77]	6.09	[4.08 - 8.17]	9.17	[6.75 - 12.07]	4.69	[2.72 - 6.57]*
Chol., 70%, 20yr	6.96	[4.37 - 9.99]	5.23	[2.72 - 7.58]	9.10	[6.65 - 12.07]	3.91	[2.00 - 5.95]*
Chol., 99%, 3yr	11.23	[8.11 - 14.66]	8.54	[5.56 - 11.50]	13.35	[10.16 - 17.14]	7.19	[5.01 - 9.40]*
Chol., 99%, 20yr	9.64	[6.91 - 13.15]	7.52	[4.38 - 10.69]	13.27	[10.08 - 17.16]	6.50	[4.64 - 8.70]*
Typh., 50%, 3yr	1.35	[-0.05 - 3.36]	0.39	[-0.01 - 1.58]	2.36	[0.09 - 4.85]	-0.02	[-0.23 - 0.39]*
Typh., 50%, 20yr	0.59	[-0.09 - 2.60]	0.24	[-0.01 - 1.29]	2.28	[0.03 - 4.84]	-0.05	[-0.23 - 0.17]*
Typh., 70%, 3yr	5.94	[3.54 - 8.60]	5.83	[3.96 - 7.75]	7.67	[5.37 - 10.35]	4.15	[2.51 - 5.80]*
Typh., 70%, 20yr	4.20	[1.95 - 6.95]	5.14	[2.82 - 7.27]	7.56	[5.25 - 10.36]	3.48	[1.90 - 5.27]*
Typh., 99%, 3yr	8.76	[6.02 - 12.16]	8.39	[5.34 - 11.29]	11.86	[8.88 - 15.49]	6.65	[4.62 - 8.81]*
Typh., 99%, 20yr	7.33	[5.00 - 10.25]	7.42	[4.16 - 10.62]	11.80	[8.79 - 15.45]	5.99	[4.25 - 8.20]*

Vaccine type	het. RPL (1)				GMNL			
	NT		TT		NT		TT	
	Mean	95% CI	Mean	95% CI	Mean	95% CI	Mean	95% CI
Chol., 50%, 3yr	4.34	[1.73 - 6.99]	0.73	[-0.17 - 2.35]*	5.04	[2.80 - 7.26]	0.30	[-0.30 - 1.64]*
Chol., 50%, 20yr	4.16	[1.39 - 6.94]	0.74	[-0.10 - 2.32]*	4.65	[1.85 - 7.01]	0.27	[-0.07 - 1.55]*
Chol., 70%, 3yr	10.19	[7.54 - 13.10]	3.71	[-0.11 - 6.87]*	10.32	[7.04 - 13.35]	2.29	[-0.09 - 5.89]*
Chol., 70%, 20yr	9.82	[6.98 - 12.93]	3.43	[-0.05 - 6.82]*	9.89	[6.13 - 13.09]	1.89	[-0.01 - 5.65]*
Chol., 99%, 3yr	15.04	[11.13 - 19.03]	4.49	[-0.02 - 8.97]*	14.82	[9.71 - 19.18]	3.04	[-0.02 - 8.09]*
Chol., 99%, 20yr	14.56	[10.39 - 18.69]	3.49	[-0.01 - 8.75]*	14.40	[8.97 - 18.90]	2.10	[0.00 - 7.44]*
Typh., 50%, 3yr	1.88	[-0.03 - 4.56]	0.42	[-0.20 - 1.66]	2.69	[0.30 - 4.98]	0.03	[-0.51 - 0.74]*
Typh., 50%, 20yr	1.60	[-0.03 - 4.37]	0.44	[-0.15 - 1.66]	2.31	[-0.01 - 4.69]	0.06	[-0.08 - 0.63]
Typh., 70%, 3yr	7.57	[4.78 - 10.55]	3.32	[-0.15 - 6.25]*	7.92	[4.59 - 10.74]	1.65	[-0.08 - 4.84]*
Typh., 70%, 20yr	7.06	[4.10 - 10.16]	3.13	[-0.08 - 6.18]*	7.48	[3.64 - 10.44]	1.23	[-0.02 - 4.54]*
Typh., 99%, 3yr	12.34	[8.46 - 16.28]	4.15	[-0.03 - 8.54]*	12.45	[7.49 - 16.64]	2.31	[-0.02 - 6.85]*
Typh., 99%, 20yr	11.82	[7.76 - 15.83]	3.24	[-0.01 - 8.37]*	12.08	[6.86 - 16.31]	1.45	[0.00 - 6.11]*

Unit: 2003 USD. CI - Confidence interval. * TT mean falls outside of NT confidence interval.

Table A.3: RPL, heteroskedastic RPL and GMNL models excluding cholera vaccine (VACC_CH) from interaction with the TT treatment dummy

	RPL		het. RPL (0)		het. RPL (1)		GMNL	
	Coef.	**s.e.**	**Coef.**	**s.e.**	**Coef.**	**s.e.**	**Coef.**	**s.e.**
Mean of coefficients								
NT*PRICE	-0.358***	(0.056)	-0.373***	(0.057)	-0.387***	(0.060)	-0.404***	(0.069)
TT*PRICE	-0.970***	(0.163)	-0.903***	(0.232)	-2.545	(3.582)	-0.706	(0.575)
NT*ASC	-2.809***	(0.568)	-2.488***	(0.497)	-2.778***	(0.552)	-2.895***	(0.574)
TT*ASC	-1.192*	(0.667)	-1.117**	(0.528)	-3.557	(5.215)	-0.520	(0.645)
NT*EFF70	0.117	(0.188)	0.118	(0.193)	-0.004	(0.200)	0.052	(0.197)
TT*EFF70	1.004***	(0.390)	1.290***	(0.430)	2.585	(3.822)	0.652	(0.532)
NT*EFF99	2.470***	(0.358)	2.295***	(0.350)	2.570***	(0.381)	2.567***	(0.417)
TT*EFF99	4.291***	(0.667)	4.276***	(1.093)	10.295	(14.390)	2.834	(2.229)
NT*DUR20	0.178	(0.133)	0.185	(0.134)	0.186	(0.141)	0.222	(0.152)
TT*DUR20	1.072***	(0.300)	0.998***	(0.343)	2.335	(3.386)	0.723	(0.623)
VACC_CH	0.237*	(0.122)	0.283**	(0.128)	0.418**	(0.164)	0.322**	(0.164)
Standard deviation of coefficients								
NT*PRICE	0.416***	(0.065)	0.422***	(0.070)	0.454***	(0.070)	0.420***	(0.080)
TT*PRICE	0.962***	(0.185)	0.771***	(0.194)	1.054***	(0.170)	0.937***	(0.178)
NT*ASC	6.256***	(0.840)	5.944***	(0.840)	6.281***	(0.892)	6.061***	(0.833)
TT*ASC	10.781***	(1.784)	10.437***	(2.790)	10.645***	(1.736)	11.850***	(1.834)
NT*EFF70	0.843***	(0.296)	0.726***	(0.286)	0.753**	(0.295)	0.831***	(0.286)
TT*EFF70	1.480***	(0.422)	1.320***	(0.520)	1.622***	(0.490)	1.490***	(0.461)
NT*EFF99	2.759***	(0.426)	2.444***	(0.402)	2.646***	(0.395)	2.389***	(0.412)
TT*EFF99	5.420***	(0.962)	4.576***	(1.338)	5.325***	(0.887)	5.249***	(0.969)
NT*DUR20	0.865***	(0.222)	0.792***	(0.213)	0.887***	(0.214)	0.837***	(0.252)
TT*DUR20	1.908***	(0.409)	1.609***	(0.468)	2.011***	(0.361)	2.162***	(0.406)
VACC_CH	1.367***	(0.215)	1.232***	(0.227)	1.386***	(0.208)	1.388***	(0.228)
(off-diagonal elements of the Cholesky matrix)								
Covariates of scale								
TT			0.286	(0.325)	-0.708	(1.380)	0.486	(0.754)
/tau			0	(constr.)	0	(constr.)	-0.694***	(0.079)
/gamma			0	(constr.)	1	(constr.)	1	(constr.)
Log-likelihood	-1,488		-1,485		-1,484		-1,490	
Observations	7,200		7,200		7,200		7,200	
Halton draws	1,000		1,000		1,000		1,000	
Parameters	77		78		78		79	
Adjusted R^2	0.403		0.404		0.404		0.401	
BIC	3,438		3,438		3,435		3,453	

***, ** and * indicating the 1%-, 5%- and 10%-level of confidence. [a] interacted with ASC_CHANGE. Adjusted R^2 is computed as $R^2 = 1 - (LL_m - k)/LL_0$, where LL_m and LL_0 are the log-likelihoods of the full model, and the intercept-only model respectively, and k the number of parameters. Bayesian Information Criterion (BIC) is calculated as $BIC = -2LL_m + k \cdot \ln(N)$ with N denoting the number of respondents. The use of BIC was preferred to Akaike Information Criterion because it imposes a stronger penalty on the inclusion of more parameters in the model.

WALD-tests for the RPL model in Table A.1 are as follows: PRICE: $p < .001$; ASC: $p = .061$; EFF70: $p = .037$; EFF99: $p = .008$; and DUR20: $p = .006$. For the coefficients in the heteroskedastic RPL(0) model the WALD-tests are: PRICE: $p = .014$; ASC: $p = .051$; EFF70: $p = .010$; EFF99: $p = .049$; and DUR: $p = .025$, and in the heteroskedastic RPL(1) model they are PRICE: $p = .546$; ASC: $p < .881$; EFF70: $p = .502$; EFF99: $p = .591$; and DUR: $p = .527$. In the GMNL model the WALD tests yield: PRICE: $p = .591$; ASC: $p = .001$; EFF70: $p = .296$; EFF99: $p = .900$; and DUR: $p = .423$.

Table A.4: WTP for vaccine types (model specification: no interaction between cholera vaccine and TT)

Vaccine type	RPL				het. RPL (0)			
	NT		TT		NT		TT	
	Mean	**95% CI**	**Mean**	**95% CI**	**Mean**	**95% CI**	**Mean**	**95% CI**
Chol., 50%, 3yr	2.34	[0.14 - 5.01]	0.01	[-0.06 - 0.11]*	2.48	[0.19 - 5.29]	-0.01	[-0.06 - 0.00]*
Chol., 50%, 20yr	3.22	[0.66 - 6.14]	0.40	[-0.01 - 1.61]*	3.36	[0.75 - 6.37]	0.11	[-0.01 - 0.71]*
Chol., 70%, 3yr	10.02	[7.41 - 12.96]	4.56	[3.21 - 5.90]*	9.44	[6.86 - 12.46]	5.08	[4.03 - 6.21]*
Chol., 70%, 20yr	10.87	[8.08 - 14.08]	6.41	[5.22 - 7.70]*	10.23	[7.47 - 13.45]	6.91	[5.91 - 7.93]*
Chol., 99%, 3yr	16.23	[12.91 - 20.12]	7.90	[6.53 - 9.42]*	14.84	[11.54 - 18.83]	8.36	[7.14 - 9.73]*
Chol., 99%, 20yr	16.88	[13.30 - 21.08]	9.58	[8.28 - 11.14]*	15.49	[11.97 - 19.80]	10.07	[8.85 - 11.41]*
Typh., 50%, 3yr	1.19	[-0.07 - 3.36]	-0.03	[-0.09 - 0.00]	1.29	[-0.06 - 3.53]	-0.01	[-0.06 - 0.00]
Typh., 50%, 20yr	1.89	[-0.01 - 4.53]	0.09	[-0.03 - 0.67]	2.07	[0.03 - 4.71]	0.01	[-0.03 - 0.08]*
Typh., 70%, 3yr	8.49	[5.76 - 11.55]	3.52	[2.06 - 4.87]*	7.99	[5.52 - 10.83]	4.12	[2.94 - 5.17]*
Typh., 70%, 20yr	9.30	[6.37 - 12.58]	5.42	[4.20 - 6.64]*	8.79	[6.06 - 11.89]	5.98	[4.85 - 7.04]*
Typh., 99%, 3yr	14.65	[11.40 - 18.33]	6.87	[5.47 - 8.33]*	13.45	[10.31 - 17.17]	7.42	[6.22 - 8.65]*
Typh., 99%, 20yr	15.14	[11.70 - 19.16]	8.56	[7.31 - 9.96]*	14.09	[10.70 - 18.12]	9.18	[7.93 - 10.43]*

Vaccine type	het. RPL (1)				GMNL			
	NT		TT		NT		TT	
	Mean	**95% CI**	**Mean**	**95% CI**	**Mean**	**95% CI**	**Mean**	**95% CI**
Chol., 50%, 3yr	1.48	[-0.23 - 4.61]	-0.21	[-2.09 - 0.49]	3.14	[0.89 - 5.41]	0.06	[-0.06 - 0.51]*
Chol., 50%, 20yr	1.88	[-0.17 - 5.53]	0.37	[-0.30 - 1.80]	3.95	[1.37 - 6.43]	0.35	[0.00 - 1.38]*
Chol., 70%, 3yr	6.34	[0.85 - 11.27]	3.49	[0.00 - 5.62]	9.73	[7.31 - 12.31]	3.09	[0.02 - 5.74]*
Chol., 70%, 20yr	6.68	[0.40 - 12.08]	4.67	[0.00 - 7.15]	10.47	[7.87 - 13.27]	4.46	[0.02 - 7.45]*
Chol., 99%, 3yr	11.00	[2.08 - 17.67]	5.61	[0.00 - 8.70]	15.35	[12.20 - 18.96]	5.46	[0.00 - 8.87]*
Chol., 99%, 20yr	11.24	[1.51 - 18.40]	6.75	[0.00 - 10.20]	15.90	[12.36 - 19.91]	6.78	[0.00 - 10.52]*
Typh., 50%, 3yr	0.67	[-0.07 - 2.65]	-0.17	[-1.50 - 0.03]*	1.59	[-0.05 - 3.88]	-0.01	[-0.06 - 0.07]
Typh., 50%, 20yr	0.92	[-0.02 - 3.47]	0.20	[-0.04 - 1.12]	2.27	[-0.01 - 5.00]	0.07	[-0.01 - 1.44]
Typh., 70%, 3yr	4.96	[0.45 - 9.14]	2.99	[0.00 - 4.91]	7.91	[4.91 - 10.67]	2.06	[0.00 - 4.61]*
Typh., 70%, 20yr	5.25	[0.45 - 9.95]	4.22	[0.00 - 6.65]	8.64	[5.29 - 11.76]	3.42	[0.00 - 6.43]*
Typh., 99%, 3yr	9.88	[2.26 - 15.62]	5.10	[0.00 - 7.97]	13.59	[10.00 - 17.30]	4.42	[0.00 - 7.68]*
Typh., 99%, 20yr	10.07	[2.16 - 16.25]	6.29	[0.00 - 9.68]	14.20	[10.16 - 18.34]	5.79	[0.00 - 9.55]*

Unit: 2003 USD. CI - Confidence interval. * TT mean falls outside of NT confidence interval.

Table A.5: Percentage change in mean bootstrapped WTP from giving time to think

	Excluding effectiveness			
	RPL	**het. RPL (0)**	**het. RPL (1)**	**GMNL**
Chol., 50%, 3yr	ns	92%	ns	ns
Chol., 50%, 20yr	ns	68%	ns	95%
Chol., 70%, 3yr	ns	57%	55%	74%
Chol., 70%, 20yr	38%	43%	39%	60%
Chol., 99%, 3yr	41%	46%	45%	63%
Chol., 99%, 20yr	33%	43%	35%	54%
Typh., 50%, 3yr	ns	92%	ns	ns
Typh., 50%, 20yr	ns	61%	ns	ns
Typh., 70%, 3yr	51%	51%	51%	74%
Typh., 70%, 20yr	34%	35%	33%	59%
Typh., 99%, 3yr	38%	38%	43%	63%
Typh., 99%, 20yr	29%	34%	31%	53%

	Excluding duration			
	RPL	**het. RPL (0)**	**het. RPL (1)**	**GMNL**
Chol., 50%, 3yr	81%	98%	83%	94%
Chol., 50%, 20yr	ns	100%	82%	94%
Chol., 70%, 3yr	ns	49%	64%	78%
Chol., 70%, 20yr	ns	57%	65%	81%
Chol., 99%, 3yr	ns	46%	70%	79%
Chol., 99%, 20yr	ns	51%	76%	85%
Typh., 50%, 3yr	ns	100%	ns	99%
Typh., 50%, 20yr	ns	100%	ns	ns
Typh., 70%, 3yr	ns	46%	56%	79%
Typh., 70%, 20yr	ns	54%	56%	84%
Typh., 99%, 3yr	ns	44%	66%	81%
Typh., 99%, 20yr	ns	49%	73%	88%

	Excluding vaccine type			
	RPL	**het. RPL (0)**	**het. RPL (1)**	**GMNL**
Chol., 50%, 3yr	100%	100%	ns	98%
Chol., 50%, 20yr	87%	97%	ns	91%
Chol., 70%, 3yr	54%	46%	ns	68%
Chol., 70%, 20yr	41%	32%	ns	57%
Chol., 99%, 3yr	51%	44%	ns	64%
Chol., 99%, 20yr	43%	35%	ns	57%
Typh., 50%, 3yr	ns	ns	100%	ns
Typh., 50%, 20yr	ns	99%	ns	ns
Typh., 70%, 3yr	59%	48%	ns	74%
Typh., 70%, 20yr	42%	32%	ns	60%
Typh., 99%, 3yr	53%	45%	ns	67%
Typh., 99%, 20yr	43%	35%	ns	59%

Note: Calculated as change in mean WTP (NT-TT/NT) from Tables 3, A.2 and A.4. ns – not significant; bootstrapped 95% confidence intervals overall. In some cases where the mean WTP was negative the implied change was greater than 100%; these were rounded down to 100%.