Environment and Development Economics 22: 202–227 © Cambridge University Press 2016

doi:10.1017/S1355770X16000280

The effect of giving respondents time to think in a choice experiment: a conditional cash transfer programme in South Africa

ELIZABETH TILLEY
Eawag, Swiss Federal Institute of Aquatic Science and Technology, Überlandstrasse 133, 8600 Dübendorf, Switzerland; NADEL: Center for Development and Cooperation, ETH Zurich, Clausiusstrasse 37, 8092 Zurich, Switzerland.
Email: elizabeth.tilley@eawag.ch

IVANA LOGAR
Eawag, Swiss Federal Institute of Aquatic Science and Technology, Switzerland.
Email: ivana.logar@eawag.ch

ISABEL GÜNTHER
NADEL: Center for Development and Cooperation, ETH Zurich, Switzerland.
Email: isabel.guenther@nadel.ethz.ch

Submitted 25 June 2015; revised 21 December 2015, 2 July 2016; accepted 26 August 2016; first published online 23 November 2016

ABSTRACT. We conducted a choice experiment (CE) to estimate willingness to accept (WTA) values for a planned conditional cash transfer (CCT) programme designed to increase toilet use in South Africa. The payment is made conditional on using a toilet and bringing urine to a central collection point. In a split-sample approach, a segment of respondents were given time to think (TTT) (24 hours) about their responses, while the remaining respondents had to answer immediately. We found significant differences in the choice behaviour between the subsamples. To validate the stated preferences with actual behaviour, a CCT programme was implemented afterwards. The stated WTA estimates were far below those revealed by actual behaviour for both subsamples.

The authors would like to thank Marcella Veronesi and two anonymous reviewers for their valuable comments, the eThekwini Municipality for their logistical support, and all of the enumerators. They are grateful for the financial support of the Bill and Melinda Gates Foundation (Grant #OPP1011603). The funding source had no involvement in any aspect of the research or preparation of the article.

Contrary to our expectations, the TTT group had underestimated their actual WTA values by an even larger margin. The preferences for various attributes were nevertheless useful in informing the design of the real intervention.

1. Introduction

'Workfare' programmes for the poor, which promote the exchange of labour for monetary compensation, became prevalent throughout the world in the 20th century (Besley and Coate, 1992; Beaudry, 2002) but were almost exclusively operated by rich, industrialized countries until the early 1990s (Ravallion, 2015). More recently, poor countries have initiated their own social welfare systems. For example, India's Mahatma Gandhi National Rural Employment Guarantee Act guarantees 100 days of paid labour to every (eligible) household per year. As of 2011, more than 50 million Indian households have been paid for 2.5 billion days of labour. The programme has measurably improved rural income and contributed to rural agriculture and infrastructure (Zepeda *et al.*, 2013).

Although similar to workfare programmes, conditional cash transfers (CCTs) are distinguished by the fact that the payments are designed to encourage socially desirable behaviour, rather than simply to compensate the recipient for labour. A number of developing countries run CCT programmes that seek to improve education and health outcomes, while helping to smooth consumption (Rawlings and Rubio, 2005). Payments are usually conditioned on school attendance, clinic visits and/or nutrition workshop attendance, which ultimately contribute to the recipient's own wellbeing. Despite being criticized for requiring the recipient to bear the opportunity costs associated with participation and for making the cash transfer conditional on a behaviour change, CCTs have generally contributed to improved health and educational outcomes (Jehan *et al.*, 2012; Saavedra and Garcia, 2013). However, their long-term effects on poverty are unclear (Cueto, 2009).

CCTs have been one of the most studied development interventions of the last decades and have been rigorously evaluated (Fernald *et al.*, 2009; Vivalt, 2015). Despite this, we were unable to find any evidence of stated preference (SP) methods used for their design. Rather, the payment amounts are likely based on budget constraints and expert opinion instead of empirical results that link payments to the recipient preferences from a SP method, such as a choice experiment (CE).

In this study, a CE was conducted in South Africa in order to estimate the willingness to accept (WTA) payment for a CCT programme aimed at increasing toilet use among rural households in South Africa. The payment is directly linked to the quantity of urine generated (i.e., toilet use), and made conditional on collecting and bringing urine to a central collection point. We analyze the relative importance of the programme attributes and WTA values. Specifically, we examine the relative importance of three different payment forms (cash, a household item, or fertilizer) and payment frequency, as well as different features of the work required (walking time and delivery frequency).

Due to their hypothetical nature, SP methods are prone to various errors/biases that can affect their validity and reliability (Bateman *et al.*, 2002). Hypothetical bias represents the potential divergence between real and hypothetical payments (Cummings *et al.*, 1986). Because the proposed situation is hypothetical, the respondents may state a willingness to pay (WTP) or WTA amount in a survey that is biased, i.e., it exceeds (is less than) the amount that they would actually pay (accept) in a real situation. The WTA value is therefore an estimation of what, on a particular day, a respondent imagines her future disutility from a loss (e.g., opportunity cost) to be worth. Enumerator (yea-saying) bias is the result of the respondent trying to provide the 'expected' answer while strategic bias reflects tactical answers made by a respondent who tries to affect the outcome in their favour (Whittington, 2010; Loomis, 2011). All are potential criticisms of SP methods and, although there have been numerous studies investigating biases, there is no single method of eliminating them completely (Murphy *et al.*, 2005).

Giving respondents time to think (TTT) about their answers (Cook *et al.*, 2011) is one approach that has been proposed to minimize hypothetical bias, as the added time and ability to discuss with family members could allow respondents to better understand and contextualize the choices presented. If the same enumerator conducts a survey, any difference in response can be attributed to the effect of giving respondents TTT. Studies using the contingent valuation method (Whittington *et al.*, 1992; Lucas *et al.*, 2007; Islam *et al.*, 2008; Donfouet *et al.*, 2014) and a CE (Cook *et al.*, 2006) found that the WTP values among respondents who were given TTT were significantly lower (up to 40 per cent) than among those who were asked to answer immediately. Still, one study from Ghana found no significant difference between estimated WTP values using a contingent valuation method when TTT was given (Whittington *et al.*, 1993). However, in order to determine whether or not TTT generates more accurate responses, the SPs of respondents need to be compared with their actual behaviour. Although comparisons of welfare measures derived from SP valuation methods to actual payments have been well studied (e.g., Griffin *et al.*, 1995; Carlsson and Martinsson, 2001; Vossler *et al.*, 2012), we are not aware of any study in which SPs are compared to actual behaviour while taking into account the TTT component. In addition, this is to our knowledge the first study that analyzes the effect of TTT on WTA values. Furthermore, given the fact that 'task familiarity' will be low (Schläpfer and Fischhoff, 2012), we expect the TTT to have a significant effect on the SPs in this context. Stated differently, the proposed scenarios for a future CCT programme linked to toilet use are quite novel and likely difficult to assess immediately; an extra 24 hours of consideration should have a noticeable impact on respondents' ability to understand and appraise them.

This study addresses several gaps in the literature. By using a split-sample approach, we are able to test the effect of giving a segment of respondents TTT on their SPs and estimated WTA values. In addition, we compare the SPs to the actual behaviour in an implemented CCT programme with the same households in order to assess whether TTT generates welfare estimates that are more accurate than those among respondents with no TTT and, therefore, if this approach is capable of reducing hypothetical bias.

2. Study site

2.1. Background

Safe, accessible toilets are crucial for public health and yet severely lacking in many poor countries (WHO/UNICEF, 2014). Open defecation that results from a lack of toilets increases the risk of disease transmission but also of environmental pollution, especially if waste enters water bodies that can become eutrophic and dangerous to consume (Briscoe, 1984). Tackling the problem requires investment, but also information about the habits and preferences of the target population. Such information is difficult to obtain because sanitation is a taboo subject and potentially difficult to discuss with an unknown enumerator. As a result, there is a relative dearth of both actual behaviour and SP information related to sanitation despite the dire need to understand more about habits, investments and the use of toilets. The challenge in urban areas is often one of understanding how much people are willing to pay for sanitation, whereas in rural areas the challenge is how to incentivize people who may not have used, or be comfortable with, an indoor toilet, to use one. Numerous awareness campaigns have focused on trying to educate the rural poor about the importance of sanitation (Peal *et al.*, 2010); in this study we use a CE to determine the value of a CCT payment that would be required to increase and sustain toilet use, specifically, the use of a novel Urine-Diverting Dry Toilet (UDDT).

UDDTs are distinguished by the separation in the toilet bowl that directs urine through a pipe in the front and allows faeces to fall directly into dehydration chambers below. Keeping the urine away from the faeces ensures that they remain dry, relatively odourless, and contained within a small volume (Tilley *et al.*, 2014). UDDTs are sustainable and appropriate for dry conditions, and were for that reason installed by the eThekwini Municipality (in the province of KwaZulu-Natal in eastern South Africa). This was done in an effort to reduce their sanitation backlog that had accumulated during the years of apartheid (Gounden *et al.*, 2006). However, because of the unusual toilet pan and the fact that they are not the aspirational 'flush toilet', they are not generally well received by those who use them (Roma *et al.*, 2013).

Toilets that have been built and are not being used represent a large investment loss for the municipality. Determining the WTA of the UDDT-owning population to use their toilets more was the first step in assessing whether or not a CCT programme could be used to incentivize increased UDDT use. Because the UDDTs are designed to separate urine from faeces, the urine quantity could be easily measured in terms of volume and used as a proxy for toilet use. However, measuring urine quantities at each individual toilet would be logistically impossible and/or extremely expensive. Therefore, in order to measure UDDT use, we asked the toilet users to transport the urine tank to a centralized collection point, for which they would in turn receive a payment.

2.2. Sample

Our study population lives in the rural areas of the eThekwini Municipality, KwaZulu-Natal, South Africa, and includes households that have received (but did not choose) UDDTs from the Municipality. More than

75,000 UDDTs have been installed in the rural parts of eThekwini that are not connected to the centralized sewer system (Grau *et al*., 2015). Our sample was drawn from six communities (with about 500–2000 households each) that are located throughout the Municipality and are representative of the households that were targeted by the UDDT programme (i.e., they are similarly poor, rural and distant from the sewer network). Before this study began, the Municipality had randomly selected about 10 per cent of the houses with UDDTs in each of the six communities to retrofit with the urine collection tanks. Our sampling strategy was to interview all of the households that had been fitted with the urine collection equipment by the Municipality in order to ensure a sufficiently large sample size. The communities included in the study are located between 30 and 50 km from Durban, the largest city in the province. The CE was presented to the households as part of a questionnaire in which a variety of data related to water, sanitation and socio-economic characteristics were collected.

The analysis is based on 803 face-to-face interviews that were conducted by 12 enumerators.

The interviews lasted approximately one hour each. The survey was carried out among adults over the age of 18 who lived in the household permanently. Furthermore, a subsample of randomly selected households of the 803 interviewed households was given 24 hours to think before providing their responses: the 'time to think' (TTT) subsample in the remainder of the article. Given that this subsample required at least two visits by each enumerator and was hence more time consuming, we aimed for the smallest sample size that would still ensure a sufficient number of observations for obtaining reasonable choice model estimates. That is, we aimed for about 25 per cent of the total sample within each area; this resulted in a total of 177 valid observations in the TTT sample.

For the TTT respondents, the enumerators administered the questionnaire and then explained the CE questions. Copies of the choice sets were left with the household so that the respondent could review them and consider the options overnight. The following day, the same enumerator returned to the household to record the choices.

The rest of respondents had to provide responses to the CE questions immediately, as is usually the case in most SP studies. Since they were not given TTT overnight, this sample is referred to as the 'No-TTT' subsample.

An overview of the main household characteristics for the two subsamples is presented in table 1. A t-test is applied to test for differences between the subsamples; the calculated p-values are presented in the final column. The results of a t-test indicate that the two subsamples are not statistically different, and therefore differences in their SPs can be attributed to the effect of TTT rather than differences in sample characteristics.

The variable 'number of UDDTs' is an important, although not straightforward, indicator of sanitation access. The UDDTs built by the municipality are robust, concrete designs that are wind- and waterproof, making them excellent storage facilities. The more UDDT toilets a household has, the more likely they may be to use at least one of them for sanitation; families with only one UDDT may have already converted it to storage and therefore may not be able to use it for sanitation purposes.

Table 1. *Household characteristics for the subsamples without and with time to think*

	No-TTT		*TTT*		*t-test*
	Mean	*SD*	*Mean*	*SD*	*p-value*
Proportion of female respondents	0.67	0.47	0.65	0.48	0.684
Respondent's age	39.72	15.33	39.09	14.86	0.627
Number of UDDTs	1.16	0.37	1.11	0.31	0.071*
Number of household members	5.19	3.09	4.88	2.86	0.232
Number of adults (over 18)	2.78	1.67	2.63	1.55	0.281
% of adults with a job	19.91	30.82	24.65	34.39	0.634
Asset index (0–1)	0.65	0.19	0.66	0.19	0.627
State benefits (R/100)	10.43	14.11	9.19	11.87	0.287
Proportion of households with electricity	0.86	0.35	0.91	0.29	0.088*
Proportion of adults with bank account	0.7	0.46	0.72	0.45	0.567
% of adults who finished university	4.83	15.85	7.12	20.36	0.113
% of adults who completed primary school	75.87	32.80	78.04	30.72	0.431
Proportion willing to work with urine	0.71	0.45	0.70	0.46	0.848
Number of observations	626		177		

Notes: $^{*}p < 0.1$. *SD* = standard deviation.

The 'asset index' is used as a proxy for income and was calculated using principle component analysis on 26 household items. 'State benefits' represents the total value of cash transfers that the household receives from the government (e.g., child support, unemployment). The values reported are in South African Rand (R) and are divided by 100. This variable could be considered as another proxy measurement for income. Assets and state benefits are highly, negatively correlated and therefore we only include state benefits in our final estimations. In other words, households who have more assets are less poor and hence entitled to fewer benefits than very poor households who have few assets and receive more state support.

The variable 'proportion of respondents willing to work with urine' shows that 70 per cent of respondents said that they are willing to work with urine themselves. The share is higher than expected given previous research (Roma *et al.*, 2013), although, given the absence of local employment, not altogether surprising. Following the CE, a question about which person in the household would be most likely to actually do the work was included. The results indicate a fairly even split between men and women (45 per cent and 55 per cent, respectively), indicating that the work would not be associated with either gender exclusively.

3. Method

3.1. *Choice experiment design*

The CE method elicits people's preferences based on choices that they make between two or more alternative descriptions of a good or service.

Table 2. *Attributes and attribute levels used in the CE*

Attribute	*Levels*		
Delivery frequency	0 visits/week	1 visit/week (20L/visit)	4 visits/week (5L/visit)
Walking time	0 minutes	5 minutes	30 minutes
Payment type	Fertilizer	Item	Cash
Payment frequency	1 time/month	4 times/month	
Payment value/20L	0 Rand	1 Rand	5 Rand

It forces respondents to make tradeoffs between different product or situational characteristics (e.g., a more expensive, but shorter trip vs. a cheaper but longer trip), which are referred to as attributes, and the levels that these attributes take (Hensher *et al.*, 2005). Based on these tradeoffs, we can derive WTA estimates for each attribute separately or for any combination of attribute levels.

The CE included five different attributes of the planned CCT intervention. The CE design with the full set of attributes and attribute levels is presented in table 2. In our study, the software JMP (www.jmp.com) was used to generate 40 choice sets each containing three alternatives (including a status quo alternative) with varying combinations of attribute levels. The 40 choice sets were divided into 10 choice packages, each consisting of four choice sets, and were randomly distributed among respondents. Therefore, each respondent in our sample received four different choice sets. Note that the choice sets were well balanced between the TTT and the No-TTT samples (i.e., the same proportion of both subsamples has received each choice set). Although there is no standard for the number of alternatives per choice set or choice sets per respondents, some have determined the optimal to be five and six, respectively (Chung *et al.*, 2010).

The attributes reflect both the work and payment features of the hypothetical CCT intervention. The work-related attributes describe the frequency of delivering the urine to the collection point where the urine volume is measured (delivery frequency) and the time required to get there (walking time). The payment-related attributes describe whether the payment would be received in the form of fertilizer (produced from the urine collected), a food item, or cash (payment type), how often the payment would be made (payment frequency) and the amount of payment (value).

In order to make sure that literacy was not a barrier to survey participation, we developed visual choice cards that included no words and only simple numbers. A local artist depicted the 80 different choice alternatives, plus the status quo option. An example of a choice card (with two choice alternatives and the status quo) is shown in figure 1. In this specific choice card, option A shows a family member walking for 30 minutes to the collection point once a week and receiving a payment of 5R once a week (i.e., four times/month). Option B shows a family member coming to the collection point once a week, walking for 5 minutes and receiving an item worth 20R, paid out once a month (i.e., 5R/20 L container). Note

https://doi.org/10.1017/S1355770X16000280 Published online by Cambridge University Press

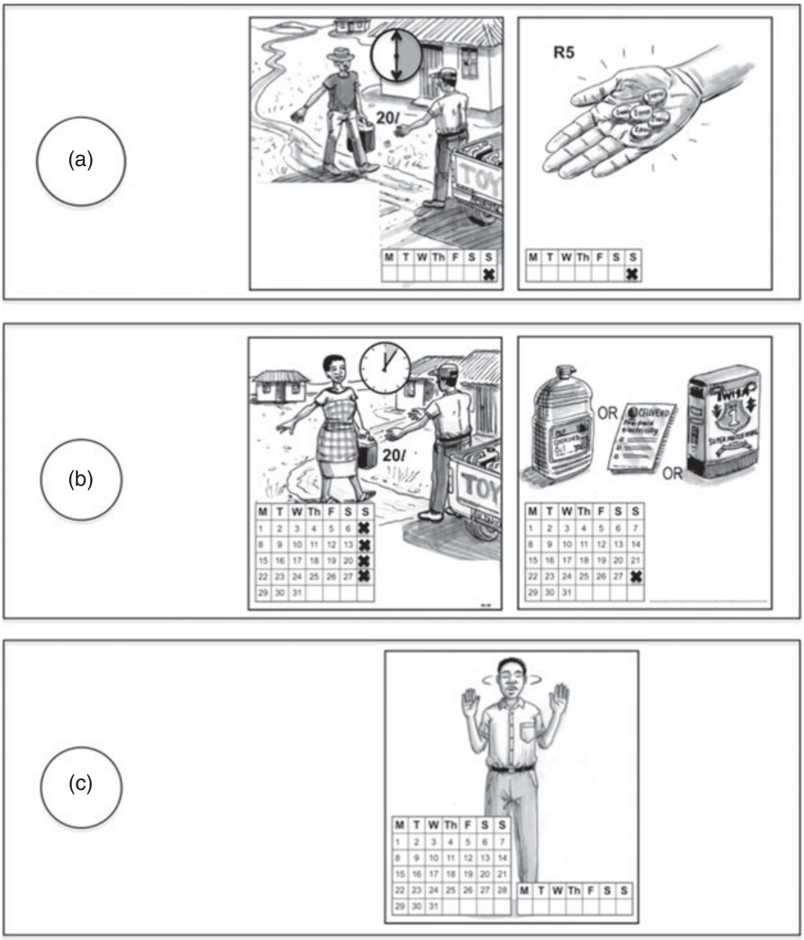

Figure 1. *Example of choice card (out of 40)*

that although the price of the item was not written on the choice card, the enumerator reminded the respondent of the value, and the prices were already well known within the communities. Option C is the status quo option, which was the same across all 40 choice sets. The status quo option shows the respondent refusing to do any work and receiving no payment. The empty calendars indicate that there is no delivery frequency and no payment frequency.

The work-related attributes were included with the intention of understanding how the CCT programme should best be implemented to maximize participation. The effect of the walking time was tested in order to help determine how many collection points would be required within a community, so that that the walking time would not exceed the average acceptable limit.

https://doi.org/10.1017/S1355770X16000280 Published online by Cambridge University Press

Cash is the most common payment vehicle in CCT programmes although it is not the only one. Food has commonly been used as an incentive in experimental health programmes (Martins *et al.*, 2009; Banerjee *et al.*, 2010). Comparison studies have addressed the impacts and cost effectiveness of CCTs in the form of cash, food or vouchers (Sutton *et al.*, 2008; Hidrobo *et al.*, 2014; Hoddinott *et al.*, 2014), although recipients' SPs remain understudied. Despite the logistical challenges, our local partner, the eThekwini Water and Sanitation Unit (EWS) was most in favour of providing fertilizer to the households with the idea that families would use it to grow food. Household items (eggs, oil, cornmeal, sugar, bread, toilet paper) were included with the intention of improving access to basic staples. We included cash to test the relative preferences between the three payment types.

The payment frequency was varied between weekly (four times a month) and monthly to understand how often respondents preferred to receive payment. The payment value for a 20 L tank delivered to a collection point was varied between 0, 1 and 5R. At the time of the study the minimum wage at the municipality was 152R per day, meaning that a 5R payment for a 5-minute walk would be equivalent to three times the minimum wage, while 5R for a 30-minute walk would be equivalent to half the minimum wage (calculated on an hourly basis). Note that only about 20 per cent of the adults in our sample had a paid job at the time of the interview and payment would not be received per visit, but per full tank (i.e., 20 L). Hence, if the delivery frequency were four times per week, in which only 5 L was delivered each time, the payment would only be received for the full amount delivered that week (20 L). The attribute levels presented on the choice cards represent the extremes of what we would expect respondents to do in real life: walking with 20 kg of household goods is not the norm, although, for some without water in their homes, fetching and carrying 20 L of water is not uncommon either (GWANET, 2015).

3.2. Estimation approach

In this unlabelled CE we measure the utility of choosing one of the two work-payment alternatives relative to the utility of the status quo option. The fully specified utility for an individual q who chooses alternative j in choice situation t is:

$$U_{qjt} = \boldsymbol{\beta}'_q \mathbf{x}_{qjt} + \boldsymbol{\gamma}'_q \mathbf{z}_{qjt} + \varepsilon_{qjt} \quad (1)$$

where $\mathbf{x}_{qjt}$ is the vector of choice attributes, $\boldsymbol{\beta}'_q$ is a vector of coefficients, $\mathbf{z}_{qjt}$ is the vector of socio-economic variables (observed), $\boldsymbol{\gamma}'_q$ is a vector of coefficients associated with the socio-economic characteristics of person q, which are invariant for that person, and ε_{qjt} is a random term.

To analyze the choices made we employed a mixed logit choice model that allows the estimated attribute coefficients to vary over respondents, reflecting the heterogeneity of individual preferences (Hensher *et al.*, 2005). The probability that person q chooses alternative i from a set of alternatives $j = (1 \ldots J)$ in a series of choice tasks $t = (1, \ldots, T)$ is conditional on the

https://doi.org/10.1017/S1355770X16000280 Published online by Cambridge University Press

parameter vector β, such that:

$$P_{qi}(\beta) = \prod_{t=1}^{T} \left[\frac{exp\left(\boldsymbol{\beta}'_q \mathbf{x}_{qit} + \boldsymbol{\gamma}'_q \mathbf{z}_{qit}\right)}{\sum_j exp\left(\boldsymbol{\beta}'_q \mathbf{x}_{qjt} + \boldsymbol{\gamma}'_q \mathbf{z}_{qjt}\right)} \right]. \quad (2)$$

Since we do not observe β, the conditional probability is integrated over all possible values of β, such that:

$$P_{qi} = \int P_{qi}(\beta)\, f(\beta)\, d\beta \quad (3)$$

where $f(\beta)$ is a density function. Equation (3) usually requires a simulation and is solved by generating draws of β from its distribution (Train, 2009). Halton sequences are used in our estimation because they produce more precise results than independent random draws in the estimation of mixed logit models (Bhat, 2001).

To evaluate the models, a distribution for each attribute coefficient was specified so that a mean and standard deviation could be estimated for each coefficient. We specified the attribute time, delivery frequency, and payment frequency as being normally distributed. The indicator variables representing cash, item and fertilizer payments were assumed to follow a uniform distribution, which is most commonly used for dummy variables (Hensher *et al.*, 2005). Note that the value attribute was specified as fixed, as is common practice in the literature.

4. Results

4.1. Choice model

In our model specification, all choice attributes were included. The results of choice models for the TTT and No-TTT samples are shown in table 3.

For each sample, models that include only choice attributes (columns 2, 4, 6 and 8), and extended models with both choice attributes and a variety of relevant socio-economic characteristics (columns 1, 3, 5 and 7), are presented. In addition, table 3 reports the results of models with an alternative-specific constant (columns 1, 2, 5 and 6) and without the constant (columns 3, 4, 7 and 8). The dependent variable is the choice of an alternative in a choice set (1 if an alternative is chosen, 0 otherwise). The attribute coefficients represent the marginal utilities of choosing a work-payment alternative relative to the status quo option. Each payment type (cash, item, fertilizer) was coded as a dummy variable; the item-based payment was set as the base category. For all choice model specifications and both subsamples delivery frequency is highly significant and negative, indicating that, as the delivery frequency increases (i.e., more trips per week), the probability of choosing a work-payment alternative over the status quo decreases. Furthermore, the standard deviation of this coefficient (*SD* Delivery Frequency) is significant, which indicates that the variable is random and preferences for it vary significantly across individuals in both subsamples.

Table 3. *Mixed logit estimation results*

	No-TTT				*TTT*			
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*
Random parameters								
Delivery frequency	−0.207***	−0.235***	−0.203***	−0.195***	−0.349***	−0.351***	−0.346***	−0.281***
	(0.040)	(0.044)	(0.040)	(0.040)	(0.093)	(0.104)	(0.092)	(0.097)
Payment frequency	−0.077**	−0.091***	−0.074**	−0.04	−0.134*	−0.202**	−0.139*	−0.063
	(0.030)	(0.032)	(0.029)	(0.030)	(0.073)	(0.084)	(0.072)	(0.068)
Cash payment	−0.016	−0.037	−0.026	0.157	0.731**	0.978**	0.733**	1.252***
– relative to item	(0.135)	(0.143)	(0.132)	(0.127)	(0.353)	(0.414)	(0.357)	(0.381)
Fertilizer payment	−0.416***	−0.545***	−0.405***	−0.376***	−0.449	−0.571*	−0.457	−0.233
– relative to item	(0.139)	(0.146)	(0.134)	(0.139)	(0.288)	(0.322)	(0.289)	(0.287)
Walking time	−0.006	−0.005	−0.005	−0.005	−0.024*	−0.032*	−0.024*	−0.027*
	(0.005)	(0.005)	(0.005)	(0.005)	(0.013)	(0.017)	(0.013)	(0.016)
Non-random parameters								
Constant	−0.513	1.388***			−2.08	2.326***		
	(0.787)	(0.243)			(1.776)	(0.690)		
Value of payment	0.162***	0.175***	0.158***	0.205***	0.337***	0.372***	0.333***	0.416***
	(0.023)	(0.023)	(0.023)	(0.023)	(0.061)	(0.062)	(0.061)	(0.062)
Age	0.004		−0.001		0.060**		0.040*	
	(0.011)		(0.009)		(0.027)		(0.022)	
No. of UDDT	0.669		0.540		−0.346		−1.305	
	(0.464)		(0.361)		(1.198)		(0.977)	
Family size	−0.318***		−0.344***		−0.227		−0.218	
	(0.077)		(0.072)		(0.171)		(0.165)	

(*continued*)

https://doi.org/10.1017/S1355770X16000280 Published online by Cambridge University Press

Table 3. *Continued*

	No-TTT				TTT			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Willing to work w/urine	3.218***		3.252***		5.790***		5.432***	
	(0.403)		(0.390)		(1.037)		(0.999)	
Benefits (R/100)	0.03141*		0.0340**		−0.004		0.020	
	(0.017)		(0.017)		(0.045)		(0.047)	
Standard deviation of random parameters								
SD Delivery frequency	0.426***	0.513***	0.407***	0.467***	0.521***	0.633***	0.516***	0.566***
	(0.052)	(0.057)	(0.054)	(0.056)	(0.131)	(0.138)	(0.129)	(0.137)
SD Payment frequency	0.009	0.101	0.023	0.145*	0.324**	0.386***	0.293**	0.250*
	(0.116)	(0.104)	(0.135)	(0.082)	(0.134)	(0.147)	(0.132)	(0.152)
SD Cash	0.151	1.093*	0.200	0.483	2.717***	3.623***	2.777***	3.503***
	(1.240)	(0.628)	(1.232)	(0.945)	(0.911)	(1.108)	(0.886)	(1.122)
SD Fertilizer	2.214***	2.471***	2.117***	2.413***	2.078***	2.51***	2.057**	2.236***
	(0.280)	(0.296)	(0.274)	(0.281)	(0.781)	(0.767)	(0.800)	(0.728)
SD Walking time	0.053***	0.056***	0.0451***	0.057***	0.058***	0.099***	0.063***	0.094***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.018)	(0.022)	(0.019)	(0.021)
N	2504	2504	2504	2504	708	708	708	708
Log likelihood	−2086	−2239	−2087	−2252	−555	−581	−555	−593
Restricted LL	−2619	−2751	−2619	−2751	−778	−778	−778	−778
McFadden pseudo R^2	0.204	0.186	0.203	0.181	0.287	0.254	0.286	0.237

Notes: *, ** and *** denote $p < 0.1$, $p < 0.05$ and $p < 0.01$, respectively. The values in brackets represent standard errors.

Model estimates for the full set of covariates are presented for the full sample and each of the subsamples in the appendix. While some of the covariates tested, particularly the indicator variables for the areas, are of varying significance, we focus the discussion on the results in table 3 in an effort to highlight the more interesting variations between the TTT and No-TTT subsamples, as well as the effect of the alternative-specific constant (ASC), which has important implications for the calculation of the WTA values later on. In general, the socio-economic covariates are not used in the determination of the WTA values; only those for the attributes are used, so the inclusion of additional covariates is to highlight potential socio-economic drivers related to choice behaviour; since the choice packages were randomly distributed across households, the inclusion of (additional) socio-economic variables does not change attribute coefficients (and WTP).

The payment frequency coefficients are significant and negative in most model specifications, meaning that both subsamples have a preference to receive payment less frequently rather than more frequently. Thus, the respondents' utility decreases as the frequency of payment increases. This result is somewhat counterintuitive, but points towards either a preference for the microsavings mechanism of delayed payments, or indicates that respondents gave more importance to the value of payments depicted on the choice cards rather than the payment frequency, perhaps without realizing that the equivalent weekly payment would be the same, if not lower, than alternatives based on weekly payments. While the value of the payment is consistently significant regardless of the specification, the significance levels of the payment frequency vary. This implies that the value attribute was more important to the respondents and likely drove their choices, especially when the tradeoffs were complex.

The results concerning the type of payment indicate important differences between the two treatment groups. The estimated coefficients for payments in cash relative to the item-based payments were positive and highly significant for the TTT subsample, but not for the No-TTT subsample. Since cash can be spent according to individuals' own preferences whereas items cannot, one would expect the utility of the cash payment to be higher than the utility of an item-based payment with an equivalent value. However, the insignificant coefficients for the cash payments for the No-TTT sample indicate that payments in cash as compared to an item-based payment do not significantly increase respondents' utility and hence the probability of choosing a work-payment alternative. Stated differently, the respondents viewed payments in the form of cash and items as equally desirable. This finding might imply that respondents without TTT did not have a chance to fully consider the utility of cash as opposed to the respondents with TTT.

However, receiving payments in the form of fertilizer rather than in the form of an item significantly reduces the utility for the No-TTT subsample. All respondents could be classified as rural, but agricultural is not common in this area (Okem *et al.*, 2013). The result is therefore not altogether surprising. Conversely, respondents who were given TTT did not perceive the fertilizer as being significantly different from items (but still of considerably lower utility than cash).

Furthermore, it is interesting to note that that the coefficients for walking time are significant for the TTT subsample but not for the No-TTT sample. TTT seemed to allow respondents to realize the opportunity cost of their time and, as a result, increased walking time to a collection point led to a lower probability of choosing a work-payment alternative, which was not the case for the No-TTT group. However, it is important to note that the standard deviations associated with the coefficients for walking time are significant in both subsamples, indicating that the preferences for walking time vary significantly across respondents within each subsample. As expected, the amount of payment is consistently positive and highly significant in all eight choice models. This result indicates that the higher the value offered, the higher the probability that respondents will choose a work-payment alternative to the status quo.

In the No-TTT subsample, as the amount of state benefits increase, so too does the probability of choosing a work-payment alternative. This result is expected since poor people who depend on state benefits are assumed to be more interested in receiving payment than those who have other sources of income, and are less poor. This phenomenon is, however, not observed for the TTT subsample; it seems that richer respondents accepted low-value work-payment options, which the No-TTT respondents rejected. The time likely allowed them to realize that low payments were better than no payments. As expected, a willingness to 'work with urine' is an important and highly significant predictor of choosing a work-payment alternative for both subsamples.

After selecting each alternative, respondents were asked what the most important attribute for making their choice was. The frequency of the work was stated as the most important attribute in 33 per cent of the choice situations, the type of payment in 30 per cent of them, and the amount of payment in only 8 per cent of the cases. This points to the fact that the respondent may have fixated on a feature of the alternative and disregarded the other attributes (both positive and negative) which led to some of the less intuitive model results. This phenomenon is known as attribute non-attendance in the CE literature (e.g., Nguyen *et al.*, 2015).

It has been suggested in the literature that the estimated ASC represents 'yea-saying' or enumerator bias. That is, the respondent is agreeing to an alternative, but the utility cannot be associated with the attributes included (Morrison *et al.*, 2002). There is no consensus about the need to include the ASC in either the model or the subsequent WTA estimates. Some authors state that it should be excluded so that only attributes are valued (Kataria, 2009), while others argue that it is necessary to include the constant as it contributes to the WTA (Klinglmair *et al.*, 2015). In an effort to examine the unobserved sources of utility, we show the models both with and without the ASC in table 3. In a labelled CE the utility for each alternative would be alternative-specific, and therefore the constant would capture the utility of an alternative that is not described by the attributes included in the model. In an unlabelled CE, such as in our case, the constant represents the utility associated with choosing *any* alternative that is not the status quo. We can deduce two key findings from the models presented here. First, when moving from the fully specified model to the attribute-only model, the ASC absorbs the utility that is not explained by the socio-economic variables. Secondly, when comparing the models with and without the ASCs, we see that, generally, the significance and size of the coefficients do not change considerably. This indicates that, at least in this study, the unobserved utility can be reasonably explained by socio-economic characteristics of respondents, and that when the constant is not included, the pseudo R^2 statistic is not substantially affected.

4.2. Effect of time to think on choice behaviour

In order to determine whether there are significant differences between the attribute coefficients estimated for the TTT and No-TTT samples we employ the Swait–Louviere test procedure (Swait and Louviere, 1993). The test is a modified two-stage Chow test which accounts for differences in both the preference parameters and the scale of the utility functions. The first step allows us to test the hypothesis that the estimated coefficients for choice attributes are equal between the two samples. If this first hypothesis cannot be rejected, the equality of scale parameters between the utility functions of two samples can be tested in the second step. Otherwise, if the first hypothesis is rejected, we cannot determine whether the differences in utilities stem from differences in preference or scale parameters due to their confounding effect.

The results of the Swait–Louviere test show that the first null hypothesis is rejected at the 1 per cent level (Likelihood-Ratio test statistic equals −28.00 with 13 degrees of freedom; $p = 0.009$). This finding indicates that there are significant differences between the preference parameters for the TTT and No-TTT choice attributes. Based on these results, we conclude that the simple process of allowing respondents to ponder their decision for 24 hours resulted in choices that were significantly different from those of respondents who were given no extra time.

The differences in choice behaviour between the two groups of respondents can be attributed to a variety of factors. We attempted to portray the scenarios in as realistic a manner as possible in the choice cards, but no image can fully capture the reality of the experience. In the 24 hours that the respondents had to consider their choices, they had the opportunity to frame the choice alternatives within the context of their daily lives and, in the process, to potentially reduce the hypothetical bias. Even the simple act of walking to and from the local shop with a heavy bag of food could force the respondents to consider whether they would be willing to do the same or more work for the payment proposed. A 30-minute walk with 20 kg may not look difficult on the choice card, but the respondents may think otherwise as they strain to get a heavy parcel home in the midday sun. Conversely, if they are walking for several hours each day, a 30-minute detour might seem insignificant.

The time provided to think about responses would also allow for discussion with family members and neighbours who could have influenced the opinion of the respondent on the work proposed and hence the choices made in the CE. Fearing that they may be made responsible for urine transport if, based on the responses, the programme were to be implemented, family members may have urged the respondent to reject all but the most

https://doi.org/10.1017/S1355770X16000280 Published online by Cambridge University Press

lucrative scenarios. Regardless of the individual's reasons, respondents who had TTT made, on average, choices that were significantly different from those who provided answers immediately.

4.3. Effect of time to think on welfare estimates

The estimated coefficients from CEs can be used to estimate both marginal and mean welfare measures. Marginal welfare measures indicate the change in utility associated with a one-unit change in a given attribute, *ceteris paribus*. In this study we are more interested in the consumer surplus (CS) associated with a given alternative, i.e., the mean WTA of a work-payment option so that we can price it appropriately in the actual CCT intervention. The expected change in CS (ΔE) that would result from the work-payment alternative compared to status quo is calculated as (Train, 2009):

$$\Delta E\left(CS_q\right) = \frac{1}{-\alpha_q}\left[ln\left(\sum_{j=1}^{J^1} e^{V_{qj}^1}\right) - ln\left(\sum_{j=1}^{J^0} e^{V_{qj}^0}\right)\right] \quad (4)$$

where the superscripts 0 and 1 denote the status quo and work-payment alternatives, respectively, and V_{qj} represents the observed utility for person q for an alternative j. In the mean WTP estimates α_q represents the marginal utility of income. In our WTA estimation, we use the coefficient estimated for the amount of payment attribute (value) and reverse the sign to account for the fact that payments are received, rather than given (from the perspective of the respondent). The observed utility of the status quo option is assumed to be zero (i.e., utilities of work-payment alternatives are measured relative to the status quo).

To guide the design of the future CCT programme, we estimated the CS, or mean WTA values for possible CCT packages. The confidence intervals around the mean values are calculated using the Krinsky and Robb (1986) procedure with 2,000 draws. The mean WTA values presented in table 4 were estimated for the alternatives in which delivery frequency is once a week and payments are made in the form of either fertilizer or cash, while other attribute levels are varied. These alternatives were selected as they would be the most realistic ones that would be used in a CCT implementation (e.g., the logistics associated with delivering a variety of household items would be difficult and frequent deliveries (four times weekly) would be too labour intensive to implement). A Poe test (Poe *et al.*, 2005) was used to statistically analyze the differences in WTA values between the two subsamples.

For reasons discussed earlier, the conditions under which an ASC should be included in a calculation of mean WTA are not well defined. Given the large and significant values that the constants take on (columns 2 and 6 in table 3), it would appear sensible to include it in a WTA estimation. However, since this value represents unobserved utility derived from making a work-payment choice, it does not have a practical meaning in the calculation of a CCT payment, i.e., the ultimate aim of this exercise. For this reason, we therefore estimate the CS associated with the work-payment

Table 4. *Mean WTA values for different alternatives, by sample*

	TTT		*No-TTT*	
	Payment freq. 1/month	*Payment freq. 1/week*	*Payment freq. 1/month*	*Payment freq. 1/week*
Fertilizer payment				
$t = 5$ mins	**1.716****	**2.168****	**3.107*****	**3.709*****
	[0.365–3.068]	[0.446–3.889]	[1.536–4.679]	[1.754–5.664]
$t = 30$ mins	**3.361*****	**3.813*****	**3.714*****	**4.316*****
	[1.370–5.352]	[1.648–5.977]	[1.921–5.508]	[2.217–6.415]
Cash payment				
$t = 5$ mins	−**1.858**	−**1.406**	**0.508**	**1.109**
	[−3.779–0.064]	[−3.443–0.631]	[−0.693–1.708]	[−0.400–2.619]
$t = 30$ mins	−**0.213**	**0.239**	**1.115**	**1.716***
	[−2.670–2.245]	[−2.227–2.704]	[−0.545–2.775]	[−0.136–3.568]

Notes: *, ** and *** denote $p < 0.1$, $p < 0.05$ and $p < 0.01$, respectively. The values in brackets represent 95% confidence intervals. t indicates walking time to collection point in minutes.

alternatives using the coefficients derived from the attribute-only models that exclude the constants (i.e., columns 4 and 8 in table 3).

The estimated mean WTA values for payments in the form of fertilizer are presented in the upper half of table 4. Comparing the first two rows, i.e., a 5-minute walking time to a 30-minute walking time, it is clear that for all attribute combinations the payment required by respondents is always higher when the walking time is increased, as would be expected. However, comparing the estimated values for different payment frequencies for a given walking time (i.e., adjacent columns) reveals that the estimated WTA values for weekly payments are higher than for monthly payments, reflecting the trends observed earlier in section 4.1. For a 5-minute walking time (TTT subsample), respondents would require 1.7R per 20 L if the total was paid out once a month (i.e., 6.8R at the end of the month) as compared to 2.2R per 20 L for weekly payments (i.e., 8.8R at the end of the month), for a weekly urine delivery. The difference of 2R is the penalty that the respondent inflicts on the CCT programme for getting small, weekly payments rather than a larger, monthly payment. Comparing the welfare measures between the two subsamples, all the estimated WTA values are higher for the No-TTT subsample, albeit not significantly (the Poe test results are available from the authors upon request).

The results for cash-based payments are presented in the bottom half of table 4. Essentially, none of the estimated WTA values for cash payments for either subsample and for any combination of the choice attribute levels is significant (with an exception of WTA of No-TTT sample for a 30-minute walking time and payment frequently once a week, which is significant at the 10 per cent level), indicating that we cannot statistically distinguish these values from zero. Stated differently, the respondents had such low stated WTA values when paid in cash that they would be willing to do the work practically for free. This finding is partly due to the high utility expressed by the respondents for cash; i.e., the preference for cash is so high that this attribute feature drives down the mean WTA estimates. The high utility of cash (i.e., large, positive values in the denominator) is responsible for the insignificant WTA estimates. The Poe test results show that WTA values for cash payments and a 5-minute walking time are significantly higher for the No-TTT sample, but not for alternatives with a 30-minute walking time.

These results are somewhat unexpected and point to the fact that giving respondents TTT resulted in lower WTA values, for both payment types. A review by Cook *et al.* (2011) showed that mean WTP values decreased by an average of 40 per cent when TTT was given. Although there has been no previous study that examined the effect of TTT on WTA, one would expect that the time would allow respondents to realize the value of their loss (e.g., time, resources) and demand a higher payment for participation in the CCT programme. Instead, it appears that those who had extra time to respond became more inclined to participate in the programme for lower payments.

4.4. Actual behaviour and hypothetical bias

Following the CE, we implemented *actual* CCT interventions in three of the six communities that took part in the CE. Interventions were different in each of the areas; details about each can be found in Tilley (2015). We focus on the results from one area (referred to in the appendix as Area 6). The interventions tested in Area 6 were based directly on the results of the CE, whereas interventions tested in the other areas were designed to test attributes not included in the CE. Therefore, the actual behaviour measured in Area 6 can be directly compared to the CE estimates, thus allowing us to test the accuracy of the SPs.

To implement the interventions, 145 households in the Area 6 sample were visited by an enumerator (in most cases the same person who conducted the interview previously) and were invited to participate in delivering urine to a community collection point. The requirements (walking) and benefits (payment) of participation were explained verbally and an information sheet was left with the family to review; households were informed that participation was purely voluntary and that that they could leave or join at any time.

The interventions tested were designed based on the results from the CE. Specifically, we made cash payments (rather than payments in the form of items or fertilizer), installed 20 L tanks (to minimize the delivery frequency), and placed three collection points within the community so that the walking time was about 15 minutes (based on an average distance of 340 m between the households and the collection point). Although the CE results indicate that a lower payment frequency is preferred, we found in a pilot that the data about the accumulated payments were easy to corrupt and so we were forced to implement weekly payments. The CCT programme implemented was therefore based on a weekly delivery frequency.

Table 5. *Estimated and measured participation numbers (and rates) for different incentive values*

	Estimated		*Measured*		
	Incentive value				
	0.24R	*1.71R*	*5R*	*10R*	*20R*
Participation	50%*	50%*	0	48%	74%
N (TTT)	626	–	0	17 (43%)	31 (78%)
N (NTTT)	–	177	0	52 (50%)	105 (73%)

Notes: *Indicates an estimated value based on the mean WTA value.

We first conducted a pilot study and offered 5R (in cash) per 20 L container of urine delivered to a collection point within the community, where the urine volume was measured and where the payment was given. In the first full intervention, we offered 10R per 20 L of urine, and in a second intervention among the same population we offered 20R for 20 L of urine. The pilot study was operated for a month, and each of the full-scale interventions lasted four months. In this way, we collected three sets of data from the 145 households over a nine-month period, resulting in a total of 435 observations.

The revealed preferences of the population painted a very different picture from the preferences stated in the CE. A summary of the estimated and the observed participation rates for different payment values is presented in table 5.

In the previous sections, we estimated a mean WTA value for the TTT subsample to be 0.24R (for a weekly payment, 30-minute walk with 20 L, cash payment). On average, respondents would be willing to participate in the programme for 0.24R (left side of table) (i.e., about half would accept the offer and half would not). Similarly, the average estimated participation rate among the No-TTT subsample based on the SP would be 1.71R for the same work scenario. The actual (measured) participation rate when 5R was offered per 20 L was zero, despite the fact that the walking time was on average shorter than 30 minutes and, in some cases, only a few minutes. In other words, the same group of people who, on average, stated that they would participate for 0.24 and 1.7R, respectively, did not once accept a 5R payment. We then offered 10R per 20 L container (first intervention) and eventually increased the payment to 20R per 20 L (second intervention). The actual participation rate at 10R per 20 L was 48 per cent; at 20R per 20 L the participation rate reached 74 per cent (right side of table).

The actual WTA value can be approximated based on the percentage of the sample that participated in the intervention for a given price. For example, when offered 10R for 20 L, 48 per cent of the targeted households participated in the intervention, implying that their actual WTA value is less than or equal to 10R for the collection and transport involved, while the WTA value for the remaining 52 per cent of the households is higher than 10R. The participation rates are calculated based on the number of households that made at least one visit to a collection point for payment. That is, each household had approximately four months to participate in each of the CCT programmes. If at the end of an intervention a household had made one or more visits to a collection point, it was considered as a participant household; otherwise it was treated a non-participant household.

Of the 145 households in the sample, 40 of them had previously completed the TTT CE and the remaining 105 had completed the No-TTT CE. The participation rates from each of these subsamples were relatively similar for each intervention, which is an expected result given that TTT and No-TTT questionnaires were randomly distributed across households. The type of CE completed should not have any relevance on the decision to participate in the programme since every household had an equal and (compared to CE) much longer time to consider the actual offer (up to four months).

The lack of actual participation in the CCT programme that offered the highest payment amount that was proposed in the CE (5R, as in table 2), and the moderate levels of participation in the real programme that offered payments higher than the stated WTA value (10R), indicate that respondents significantly underestimated their stated WTA values. Recall that the stated WTA estimates from the TTT subsample were even lower than from the No-TTT subsample. This finding implies that giving respondents TTT resulted in a larger discrepancy between stated WTA values and actual behaviour, and therefore in a larger hypothetical bias associated with SPs compared with the No-TTT sample.

5. Conclusions

Using a CE we illustrated the relative importance of different CCT programme attributes, estimated the WTA payments for participation in different CCT programmes, tested the impact of giving respondents TTT on choice behaviour and welfare measures in a split-sample approach, and compared the degree of hypothetical bias between the two samples by comparing stated WTA values derived from the CE with actual behaviour in the implemented CCT programme.

The largest difference found between the TTT and No-TTT subsamples was in terms of the payment vehicle. The No-TTT respondents did not express a strong preference for receiving payments in cash compared to a household item, while the TTT respondents significantly preferred cash payments to items. Moreover, the No-TTT respondents preferred item-based payments much more than those in the form of fertilizer, while this effect is not observed among the TTT respondents.

The stated willingness to work with urine in the survey was a strong predictor of choosing one of the non-status quo options, i.e., the stated willingness to participate in the CCT programme. Indeed, models that do not include this covariate produced highly significant constant terms, indicating a strong presence of unobserved utility, unexplained by the choice attributes. Respondents who were willing to work with urine derived a high degree of utility from any work-payment options, regardless of the attributes and attribute levels presented.

The TTT respondents stated lower mean WTA values than those who answered the survey immediately. While we had expected differences between the two subsamples, we assumed that the TTT sample would state higher WTA values. Based on the results, we are inclined to conclude that the extra TTT had the effect of increasing the perceived value of the payments offered. However, once the CCT programme was implemented, the actual behaviour was quite different from that stated in the survey, as people were not willing to participate in the programme for 5R, which was more than two times higher than the stated WTA value for a similar hypothetical scenario (1.7R). Most interestingly, the TTT respondents, who had lower stated WTA values than the No-TTT group, overestimated their actual participation in the programme more than the No-TTT sample. The discrepancy between their stated WTA values and actual behaviour was thus larger for the TTT sample, indicating a higher degree of hypothetical bias among these respondents. We expected that the extra TTT would allow respondents to realize how much work was required for the given amount of payment; instead, they seemed to think that any payment was better than no payment, as long as the payment was received in cash.

In the estimation of the WTA values for the design of our CCT programme we excluded the ASC, but our results reveal the important implications that including or excluding the ASC can have on the outcomes of the choice models. They showed that unobserved utility as captured by the ASC was high when we included it in the models, but could be reduced substantially when controlling for respondent socio-economic characteristics. We believe that this aspect of WTA estimation merits more attention in future research.

Despite the large degree of hypothetical bias associated with the SPs obtained from the CE, the results were useful in informing the design (i.e., attributes) of the real intervention, even if the values were not completely accurate. Specifically, the TTT group correctly predicted the importance of the walking time and the importance of cash payments. The results from the CE all point in the direction of the actual behaviour, although we were surprised at the low levels of actual participation, especially given the very low levels of employment and the payment values that, when converted, represented average, hourly salaries.

There are two possible explanations for the discrepancy between the stated WTA values and actual behaviour. First, the range of values for the price (i.e., value of payment) attribute in the CE design might have been too narrow, i.e., values higher than 5R could have been included in the CE. Secondly, and not surprisingly, the reality of participating in the programme might not have been as simple as the stylized images in the survey may have indicated. Notwithstanding the toilet use and urine collection, the work of transporting 20 kg over rough terrain in hot temperatures was likely too daunting in reality to be worth the payment.

References

Banerjee, A.V., E. Duflo, R. Glennerster, and D. Kothari (2010), ‘Improving immunisation coverage in rural India: clustered randomised controlled evaluation of immunisation campaigns with and without incentives', *British Medical Journal* **340**: c2220.

Bateman, I.J., R.T. Carson, B. Day, *et al*. (2002), *Economic Valuation with Stated Preference Techniques: A Manual*, Cheltenham: Edward Elgar.

Beaudry, R. (2002), 'Workfare and welfare: Britain's new deal', CCGES Working Papers Series, Canadian Centre for German and European Studies, Toronto.

Besley, T. and S. Coate (1992), 'Workfare versus welfare: incentive arguments for work requirements in poverty-alleviation programs', *American Economic Review* **82**(1): 249–261.

Bhat, C.R. (2001), 'Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model', *Transportation Research Part B: Methodological* **35**: 677–693.

Briscoe, J. (1984), 'Technology and child survival: the example of sanitary engineering', *Population and Development Review* **10**: 237–253.

Carlsson, F. and P. Martinsson (2001), 'Do hypothetical and actual marginal willingness to pay differ in choice experiments?', *Journal of Environmental Economics and Management* **41**(2): 179–192.

Chung, C., T. Boyer, and S. Han (2010), 'How many choice sets and alternatives are optimal? Consistency in choice experiments', *Agribusiness* **27**(1): 114–125.

Cook, J., D. Whittington, D.G. Canh, F.R. Johnson, A. Nyamete, and A. Nyam (2006), '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 (2011), 'Giving stated preference respondents "time to think": results from four countries', *Environmental Resource Economics* **51**(4): 473–496.

Cueto, S. (2009), 'Conditional cash-transfer programmes in developing countries', *The Lancet* **374**(9706): 1952–1953.

Cummings, R.G., D.S. Brookshire, and W.D. Schulze (1986), *Valuing Environmental Goods: An Assessment of the Contingent Valuation Method*, Totowa, NJ: Rowman & Allanheld.

Donfouet, H.P.P., J. Cook, and P.W. Jeanty (2014), 'The economic value of improved air quality in urban Africa: a contingent valuation survey in Douala, Cameroon', *Environment and Development Economics* **20**(5): 630–649.

Fernald, L.C., P.J. Gertler, and L.M. Neufeld (2009), '10-year effect of Oportunidades, Mexico's conditional cash transfer programme, on child growth, cognition, language, and behaviour: a longitudinal follow-up study', *The Lancet* **374**(9706): 1997–2005.

Gounden, T., B. Pfaff, N. Macleod, and C.A. Buckley (2006), 'Provision of free sustainable basic sanitation: the Durban experience', Paper presented at the *32nd WEDC International Conference*, 13–17 November, Colombo, Sri Lanka.

Grau, M.G., S.L. Rhoton, C.J. Brouckaert, and C.A. Buckley (2015), 'Evaluation of an automated struvite reactor to recover phosphorus from source-separated urine collected at urine diversion toilets in eThekwini', *Water SA* **41**(3): 383–389.

Griffin, C.C., J. Briscoe, B. Singh, R. Ramasubban, and R. Bhatia (1995), 'Contingent valuation and actual behavior: predicting connections to new water systems in the state of Kerala, India', *World Bank Economic Review* **9**(3): 373–395.

GWANET (Gender and Water in Central Asia) (2015), 'Facts about women and water', [Available at] http://www.gender.cawater-info.net/what_is/facts_e.htm.

Hensher, D.A., J.M. Rose, and W.H. Greene (2005), *Applied Choice Analysis: A Primer*, Cambridge: Cambridge University Press.

Hidrobo, M., J. Hoddinott, A. Peterman, A. Margolies, and V. Moreira (2014), 'Cash, food, or vouchers? Evidence from a randomized experiment in northern Ecuador', *Journal of Development Economics* **107**: 144–156.

Hoddinott, J., S. Sandström, and J. Upton (2014), 'The impact of cash and food transfers: evidence from a randomized intervention in Niger', International Food Policy Research Institute (IFPRI) Discussion Paper No. 01341, Washington, DC.

Islam, Z., B. Maskery, A. Nyamete, M.S. Horowitz, M. Yunus, and D. Whittington (2008), 'Private demand for cholera vaccines in rural Matlab, Bangladesh', *Health Policy* **85**(2): 184–195.

Jehan, K., K. Sidney, H. Smith, and A. de Costa (2012), 'Improving access to maternity services: an overview of cash transfer and voucher schemes in South Asia', *Reproductive Health Matters* **20**(39): 142–154.

Kataria, M. (2009), 'Willingness to pay for environmental improvements in hydropower regulated rivers', *Energy Economics* **31**(1): 69–76.

Klinglmair, A., M.G. Bliem, and R. Brouwer (2015), 'Exploring the public value of increased hydropower use: a choice experiment study for Austria', *Journal of Environmental Economics and Policy* **4**(3): 315–336.

Krinsky, I. and A.L. Robb (1986), 'On approximating the statistical properties of elasticities', *Review of Economics and Statistics* **68**(4): 715–719.

Loomis, J. (2011), 'What's to know about hypothetical bias in stated preference valuation studies?', *Journal of Economic Surveys* **25**(2): 363–370.

Lucas, M.E.S., M. Jeuland, J. Deen, *et al*. (2007), 'Private demand for cholera vaccines in Beira, Mozambique', *Vaccine* **25**(14): 2599–2609.

Martins, N., P. Morris, and P.M. Kelly (2009), 'Food incentives to improve completion of tuberculosis treatment: randomised controlled trial in Dili, Timor-Leste', *British Medical Journal* **339**: b4248–b4248.

Morrison, M., J. Bennett, R. Blamey, and L. Louviere (2002), 'Choice modeling and tests of benefit transfer', *American Journal of Agricultural Economics* **84**(1): 161–170.

Murphy, J.J., P.G. Allen, T.H. Stevens, and D. Weatherhead (2005), 'A meta-analysis of hypothetical bias in stated preference valuation', *Environmental Resource Economics* **30**(3): 313–325.

Nguyen, T.C., J. Robinson, J.A. Whitty, S. Kaneko, and T.C. Nguyen (2015), 'Attribute non-attendance in discrete choice experiments: a case study in a developing country', *Economic Analysis and Policy* **47**: 22–33.

Okem, A.E., S. Xulu, E. Tilley, C.A. Buckley, and E. Roma (2013), 'Assessing perceptions and willingness to use urine in agriculture: a case study from rural areas of eThekwini municipality, South Africa', *Journal of Water, Sanitation and Hygiene for Development* **3**(4): 582–611.

Peal, A.J., B.E. Evans, and C. van der Voorden (2010), *Hygiene and Sanitation Software: An Overview of Approaches*, Geneva: Water Supply and Sanitation Collaborative Council.

Poe, G.L., K.L. Giraud, and J.B. Loomis (2005), 'Computational methods for measuring the difference of empirical distributions', *American Journal of Agricultural Economics* **87**(2): 353–365.

Ravallion, M. (2015), *The Economics of Poverty: History, Measurement, and Policy*, Oxford: Oxford University Press.

Rawlings, L.B. and G.M. Rubio (2005), 'Evaluating the impact of conditional cash transfer programs', *World Bank Research Observer* **20**(1): 29–55.

Roma, E., K. Philp, C.A. Buckley, S. Xulu, and D. Scott (2013), 'User perceptions of urine diversion dehydration toilets: experiences from a cross-sectional study in eThekwini Municipality', *Water SA* **39**(2): 302–312.

Saavedra, J. and S. Garcia (2013), 'Educational impacts and cost-effectiveness of conditional cash transfer programs in developing countries: a meta-analysis', CESR Working Paper No. 2013-004, Center for Economic and Social Research, Los Angeles, CA.

Schläpfer, F. and B. Fischhoff (2012), 'Task familiarity and contextual cues predict hypothetical bias in a meta-analysis of stated preference studies', *Ecological Economics* **81**: 44–47.
Sutton, W.R., D.M. Larson, and L.S. Jarvis (2008), 'Assessing the costs of living with wildlife in developing countries using willingness to pay', *Environment and Development Economics* **13**(4): 475–495.
Swait, J. and J. Louviere (1993), 'The role of the scale parameter in the estimation and comparison of multinomial logit models', *Journal of Marketing Research* **30**(3): 305–314.
Tilley, E. (2015), 'Acceptance, impact and feasibility of incentives for increasing toilet use: a case study in eThekwini, South Africa', PhD thesis, ETH Zurich, Zurich.
Tilley, E., L. Ulrich, C. Luethi, P. Reymond, and C. Zurbrugg (2014), *Compendium of Sanitation Systems and Technologies*, 2nd edn, Dübendorf: Eawag.
Train, K. (2009), *Discrete Choice Methods with Simulation*, New York: Cambridge University Press.
Vivalt, E. (2015), 'How much can we generalize from impact evaluations?', Unpublished manuscript, New York University, New York.
Vossler, C.A., M. Doyon, and D. Rondeau (2012), 'Truth in consequentiality: theory and field evidence on discrete choice experiments', *American Economic Journal: Microeconomics* **4**(4): 145–171.
Whittington, D. (2010), 'What have we learned from 20 years of stated preference research in less-developed countries?', *Annual Review of Resource Economics* **2**(1): 209–236.
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**(3): 205–225.
Whittington, D., D.T. Lauria, A.M. Wright, K. Choe, J.A. Hughes, and V. Swarna (1993), 'Household demand for improved sanitation services in Kumasi, Ghana: A contingent valuation study', *Water Resources Research* **29**(6): 1539–1560.
WHO/UNICEF (2014), *Progress on Drinking Water and Sanitation*, Geneva: World Health Organization.
Zepeda, E., S. McDonald, M. Panda, G. Kumar, and C. Sapkota (2013), *Employing India: Guaranteeing Jobs for the Rural Poor*, Washington, DC: Carnegie Endowment for International Peace.

Appendix

Table A1. *Mixed logit estimation results with additional covariates*

	Full	*TTT*	*No-TTT*
Random parameters			
Delivery frequency	−0.196***	−0.381***	−0.168***
	(0.035)	(0.108)	(0.040)
Payment frequency	−0.086***	−0.158*	−0.074**
	(0.028)	(0.088)	(0.031)
Cash payment	0.120	0.957**	−0.016
– relative to item	(0.125)	(0.429)	(0.139)
Fertilizer payment	−0.454***	−0.523	−0.418***
– relative to item	(0.131)	(0.355)	(0.147)
Walking time	−0.010**	−0.02171	−0.00453
	(0.005)	(0.015)	(0.005)
Non-random parameters			
Constant	0.676	−4.037	2.547**
	(0.982)	(4.354)	(1.084)
Value of payment	0.215***	0.398***	0.182***
	(0.022)	(0.068)	(0.024)
Age	0.006	0.062	−0.012
	(0.010)	(0.042)	(0.011)
No. of UDDT	0.259	0.440	0.185
	(0.490)	(1.793)	(0.558)
Family size	−0.117	0.047	−0.045
	(0.094)	(0.340)	(0.101)
Willing to work w/urine	2.740***	6.074***	2.222***
	(0.403)	(1.485)	(0.429)
Benefits (R/100)	0.021	0.050	0.019
	(0.015)	(0.069)	(0.016)
Proportion of female respondents	0.543*	0.629	0.229
	(0.325)	(1.262)	(0.350)
Number of adults (over 18)	−0.177	−0.125	−0.214
	(0.154)	(0.488)	(0.162)
% of adults with a job	−0.009*	−0.008	−0.011*
	(0.005)	(0.019)	(0.006)
Proportion of households with electricity	−0.033	−0.104	0.113
	(0.469)	(2.125)	(0.481)
Proportion of adults with bank account	−0.519	−0.064	−0.879**
	(0.372)	(1.205)	(0.413)
% of adults who finished university	0.009	0.027	0.004
	(0.012)	(0.030)	(0.013)
% of adults who completed primary school	0.009	0.006	0.002
	(0.005)	(0.016)	(0.006)
Area 1	−1.067	−2.854	−1.034
– relative to Area 6	(0.748)	(1.957)	(0.964)
Area 2	−3.381***	−7.091***	−3.108***
– relative to Area 6	(0.507)	(2.041)	(0.567)
Area 3	−3.110***	−0.805	−3.531***
– relative to Area 6	(0.757)	(1.789)	(0.784)

(*continued*)

https://doi.org/10.1017/S1355770X16000280 Published online by Cambridge University Press

Table A1. *Continued*

	Full	*TTT*	*No-TTT*
Area 4	−3.672***	−2.817*	−4.393***
– relative to Area 6	(0.518)	(1.490)	(0.605)
Area 5	0.475	4.368	0.148
– relative to Area 6	(0.469)	(3.251)	(0.502)
Standard deviation of random parameters			
SD Delivery frequency	0.399***	0.596***	0.387***
	(0.049)	(0.160)	(0.053)
SD Payment frequency	0.120	0.366**	0.116
	(0.096)	(0.151)	(0.102)
SD Cash	0.850	3.128**	0.154
	(0.705)	(1.253)	(1.425)
SD Fertilizer	2.293***	2.497***	2.281***
	(0.263)	(0.877)	(0.290)
SD Walking time	0.051***	0.084***	0.044***
	(0.007)	(0.020)	(0.008)
N	3212	708	2504
Log likelihood	−2390	−498	−1856
Restricted LL	−3128	−725	−2403
McFadden pseudo R^2	0.236	0.324	0.228

Notes: *, ** and *** denote $p < 0.1$, $p < 0.05$ and $p < 0.01$, respectively.

https://doi.org/10.1017/S1355770X16000280 Published online by Cambridge University Press