Busara

Busara Blue Paper Series

The unreliability of value elicitation methods in valuing development interventions

No. 002

Jeremy Shapiro, Changing Jang, and Nicholas Owsley

June 2020

The unreliability of value elicitation methods in valuing development interventions

Jerem Shapiro[1], Changing Jang[2], Nicholas Owsley[3]

Author Affiliations:
[1] The Busara Center for Behavioral Economics. Corresponding author: nicholas.owsley@busaracenter.org
[2]-[3] The Busara Center for Behavioral Economics

Non-Author Contributions: We are indebted to all field officers who worked tirelessly on this project, and specifically to Esther Owelle for managing a host of the complex field procedures. We thank Akanksha Arora for contributing to the project design and data cleaning, to Aya Vang for assistance in data cleaning and analysis, and Chang Tang, Tommie Thompson, and the attendees of the SEEDEC 2019 conference in Berkeley for helpful comments on earlier drafts.
Conflicts of Interest: There are no Conflicts of Interest to declare for this study.
Open Science: The study was pre-registered at the AEA RCT registry: AEARCTR-0001371. All materials and data will be made public on OSF following journal publication.
IRB: This design was approved by the Kenyan Medical Research Institute (KEMRI); IRB Approval Date: 2016-06-15, IRB Approval Number: 531
Funding: The Funding for this study was provided by the Bill and Melinda Gates Foundation.

The Busara Blue Paper Series is a collection of working papers on topics relevant to experimental and behavioral economics in the global South. All papers comply with Busara's Open Science and research ethics policies, have IRB approval, and all study materials and data are to be published on the Open Science Framework www.osf.io. Papers are circulated for discussion and comment - feedback on the paper directly to the corresponding author is welcomed.

Busara is a research and advisory firm dedicated to advancing and applying behavioral science in the global south. Daykio Plaza, Ngong Road, Nairobi, Kenya | busaracenter.org | +254 713 474 296.

The unreliability of value elicitation methods in valuing development interventions

Abstract

This study assesses the relative reliability of the most common incentive-compatible value elicitation techniques, and compares valuations generated by each technique to those from a hypothetical question. Specifically, we collect valuations for 18 common aid interventions from 793 potential aid recipients using 6 randomly assigned elicitation methods. In a follow up survey, respondents were given a 'take-it-or-leave-it' (TIOLI) offer for an intervention - we measure reliability as whether the elicitation method predicts the respondent's choice at follow-up. Our results show that valuations are systematically overstated across methods and are generally not consistent with responses to a concrete TIOLI offer - only 40\% of valuations were consistent with TIOLI choices. Valuations are also sensitive to the elicitation method used and to framing. Overall, incentive-compatible techniques do not perform meaningfully better than a hypothetical question. We conclude that valuations can be obtained inexpensively using a hypothetical question, but that policy makers should use valuation outputs with caution and refrain from using them as 'point estimates' given the limitations to their content.

JEL Codes: C81, C83, D90, O12

Key Words: experimental economics; willingness-to-pay; field experiment; elicitation methods; Beck-DeGroot-Marschak; multiple price list; contingent valuation.

1 Introduction

Measuring individual valuations for tangible and intangible goods is central to the academic discipline of microeconomics, and is a widely-used tool for government policy decisions and resource allocation. In the academic domain, experimenter-induced valuations are a ubiquitous tool in experimental economics (Andersen et al., 2008). They have historically been used to derive demand curves and map game and market dynamics in controlled settings (Smith, 1982; Harrison et al., 2004). They have also been used to derive individual preferences, such as time and risk preferences, and to build models of economic behavior (Andersen et al., 2008; Harrison et al., 2004; Holt and Laury, 2002; Frederick et al., 2002).

In public policy, valuations provide critical inputs into pricing and subsidy policy, cost-benefit analysis, and other resource allocation decisions, specifically in health and environmental policy. For example, since the 1980s valuations have been used to guide compensation for environmental accidents (Boyle, 2017; Freeman III et al., 2014), and by the Environmental Protection Agency (EPA) in cost-benefit analyses and for pricing environmental services (Nyborg, 1996; Boyle, 2017; Freeman III et al., 2014).

More recently, tools for measuring valuation have been used in development economics to understand users' willingness to pay for development aid services and to derive demand curves for these services. There is a rich literature in development economics that deploys value elicitation methods in field experiments, commonly using methods such as contingent valuation, Becker-DeGroot-Marschak (BDM) lotteries, multiple price lists (MPL), and take-it-or-leave-it offers (TIOLI). Many studies have utilized these approaches to measure valuations for common development interventions (Dupas and Miguel, 2017; Shapiro, 2019a,b; Guiteras et al., 2016a; Berry et al., 2020; Kremer et al., 2011; Cohen et al., 2010; Whittington et al., 1992; Anjum Altaf and Hughes, 1994; Bohm et al., 1993; Swallow and Woudyalew, 1994; Cole and Fernando, 2016; Cole et al., 2013; Guiteras et al., 2016b).

This research has contributed heavily to development aid and developing country policy. Seminal studies on valuations of bed nets (Cohen et al., 2010), for example, precipitated a shift in pricing and subsidy policy for bed nets by a range of international organizations and governments. The World Health Organization (WHO), The UK's Department for International Development (DfID), and Population Services International (PSI) all support free bed net distribution, citing valuation studies in their policy justifications (Organization, 2014; J-Pal). Many governments across Africa and Asia now distribute fully-subsidized bed nets in collaboration with international aid organizations (Stevens et al., 2013; Leonard et al., 2014; Peletz et al., 2017). Similarly, deworming tablet distribution has been influenced by valuations (Kremer and Miguel, 2007) and studies using valuation elicitation methodologies have led to large-scale government and aid organization support of free deworming in schools across Africa and Asia (Action).

Experimenter-induced valuations therefore provide foundational inputs for both microeconomic modelling in the economics literature and in the allocation of scarce resources. This is specifically true of the development economics literature and policies related to development aid interventions. The veracity of value elicitation methods is therefore of keen scientific and policy interest. While a variety of studies have explored value elicitation methods, most existing research has been conducted with Western populations and few studies directly contrast alternative valuation techniques.

There is a relatively rich literature comparing alternative techniques within western populations for the valuation of consumer goods or lotteries; these studies typically find that valuations vary across techniques and are susceptible to biases and contextual factors (e.g. (Bohm et al., 1997; Andersen et al., 2006; de Meza et al., 2013; Brebner and Sonnemans, 2018)). Some studies extend these findings to assess these methods in development aid settings (Guiteras et al., 2016a; Bohm et al., 1993; Whittington et al., 1992). For example, Berry et al. (2020) measure willingness to pay for water filters in Ghana using both TIOLI and BDM, which is the most common method for eliciting individual indifference points (Dupas and Miguel, 2017). The authors find that BDM systematically under-predicts willingness to pay relative to TIOLI and the magnitude of this under-prediction increases with price. Similarly, Dupas and Miguel (2017) discuss that BDM is more likely to underpredict WTP when an aid intervention is relatively unknown compared to when it is well-known. de Meza et al. (2013) discuss how the BDM mechanism may not be incentive compatible due to reference-dependent preferences, price anchoring, and mechanism misunderstanding on the part of respondents. They conduct an experiment with university students using a modified BDM, and find that valuations are sensitive to the framing of the BDM method. The most common alternative to the BDM method, the Multiple Price List (MPL), has also demonstrated limitations, specifically that it produces interval responses and is susceptible to framing effects (Andersen et al., 2006; Brebner and Sonnemans, 2018). Evidence also suggests that some of the biases associated with these techniques could interact with the specific population group receiving aid interventions (Falk et al., 2018) and with the specific aid interventions offered (Dupas and Miguel, 2017).

However, in spite of this rich literature, there is no study that systematically compares the most common value elicitation techniques against each other in terms of valuation amounts and valuation reliability, and with aid recipients and across multiple different aid interventions.

In this study, we evaluate several of the most common incentive-compatible valuation techniques in terms of valuation amounts and the reliability of valuations and compare them to a non-incentivized hypothetical valuation. The sample population for this study comprises a low-income group of aid recipients in Kenya and we measure valuations for a variety of common aid interventions in Kenya. Specifically, we interviewed a sample of 793 low-income Kenyans to seek their valuation for 18 common development interventions. The specific elicitation mechanism we used to obtain these valuations was randomly determined for each respondent and applied for the full range of 18 possible interventions. We use these values to test for differences in valuation amounts across elicitation methods: BDM, BDM with illustrative training examples, MPL, probabilistic incentives (the respondent's valuation has a chance of determining if they receive a particular intervention) and certainty (the respondent knows for sure their valuation will determine whether they receive a particular intervention). Following the value elicitation survey, one program was chosen for the respondent by lottery. Respondents were then revisited and offered a choice between that program and a cash amount close to their indifference point: observing what percentage of respondents have actual choices under this TIOLI offer that are consistent with the predicted choices based on their indifference points across the different value elicitation methods allows us to identify the reliability of each method for obtaining respondent's valuations.

With respect to the average valuations of aid programs, we find that using BDM or BDM with a detailed explanation does not generate valuations any different than simply asking. When respondents know the valuation they state will definitely determine whether they receive cash or a program (via the BDM), in contrast to doing so with some probability, they systematically report higher valuations for the intervention. We do not believe these reflect true valuations, as these respondents are more likely than others to reverse their expected decision when presented with a TIOLI offer. We speculate this increase in valuation is driven by framing effects. We find that MPL influences valuations since respondents skew towards the maximum offer in the MPL menu. How this impacts average valuations relative to other methods depends on where the bounds of the MPL menu are set.

We also test whether respondents stick to the choice their initial valuation implies when presented with a TIOLI offer. Across all valuation methodologies, valuations are largely inconsistent with choices under a TIOLI offer ($\sim$ 40% consistency) and are not explained by an acceptable degree of noise, nor by socio-economic or demographic characteristics that theoretically predict inconsistency. Respondents are systematically more likely to choose cash over the program under TIOLI, even when their initial program valuation far exceeds the cash offer. The MPL method appears to increase the rate of consistency, though still only generates 60% 'consistent' choices, and the effect is not robust to all specifications.

We add to the literature by being one of the first studies to test the relative reliability of different valuation techniques for valuations of development aid interventions, and the only (to our knowledge) to compare BDM, MPL and contingent valuation methods in the same study. Second, by comparing valuations to a TIOLI offer with the same participants, we are also the first to compare these techniques with regards to their reliability compared to a benchmark. Third, we test the sensitivity of the most common technique, the BDM method, to different formulations. Finally, we test the above methodological variations both for a range of the most common development aid interventions and with a group of actual aid recipients, thereby increasing the external validity of the methodological findings through robustness to aid intervention heterogeneity and through targeting the most policy-relevant population.

2 Study design

As the specific context for this study was the valuation of aid interventions by aid recipients, we began by identifying a relevant set of interventions and a population of likely aid recipients to participate in the valuation exercise. Subsequently, drawing on previous literature, we identified the most common incentive compatible valuation techniques. In an initial survey, respondents valued each intervention using one of the identified methodologies. In a follow up survey, respondents made a TIOLI choice between an intervention and a cash payment near their indifference point for the intervention. Finally, either the program or cash was delivered.

2.1 Program selection

We undertook a multistep process to select programs for benchmarking, with the goal of selecting aid programs that are relevant to recipients and policy makers. Additionally, given there may be idiosyncrasies driving the valuation for specific types of interventions, we sought to include a broad set of interventions across multiple sectors. This process is detailed in Shapiro (2019a). At a high level, we identified the sectors most commonly supported by the Government of Kenya, multilateral aid institutions and philanthropic foundations. We then identified common specific aid interventions within those sectors that we could feasibly deliver to recipients.

The final list of interventions included in the study is:

1. Agriculture:
   (a) Extension - group-based agricultural extension courses over the period of one cropping season.
   (b) Inputs - 50 kg fertilizer.
2. Water:
   (a) Water supply - an easily accessible water source such as a water tank for the community.
   (b) Hygiene / WASH education - a group-based single session on safe water practices.
   (c) Hygiene / WASH supplies - basic hygienic supplies (soap and chlorine for water treatment) for two months.
3. Health:
   (a) Family planning services - one free visit to a family planning clinic to receive family planning services with allowance for transportation.
   (b) Condoms - box of 50 condoms.
   (c) Bed nets - an insecticide treated bed net.
   (d) HIV research - a donation to HIV / AIDS research.
   (e) Malaria research - a donation to malaria research.
   (f) Mass deworming - a donation to support deworming programs.
4. Education
   (a) Teacher training - training for one teacher.
   (b) Computers in schools - computers provided to one government run school.
   (c) Out of school tutoring - weekly tutoring sessions for one child for one school term.
   (d) Vocational training - a vocational training course in computer skills.
5. Energy:
   (a) Solar power - a solar power system that allows one to power a lamp and recharge a cell phone.
6. Other
   (a) Access to stress management smartphone app - a smartphone and training on how to use stress and anxiety reduction tools available on that phone.
   (b) Financial literacy training - a group-based training session on financial management

2.2 Location selection

The aim of this study is to understand the preferences of current or potential recipients of development programs. We therefore selected areas with relatively high poverty. Beginning with a list of Kenyan counties, we filtered out all counties with less than a 40% poverty rate, or just below the national rate of 46% (World Bank, 2015). The one exception is Nairobi County, as we sought to include low-income households in urban centers as well. Due to logistical considerations, we then filtered out counties in the lower third based on household density. Remaining counties were then filtered out or prioritized based on the poverty rate, household density, fertilizer use, HIV, diarrhea and malaria prevalence, bed net use and secondary school enrollment rates (all data comes from Kenya Open Data). Ultimately we chose to collect data in three Kenyan counties: Nairobi, Nakuru and Makueni.

2.3 Value elicitation methods and assignment

In order to find an accurate, reliable, efficient and low-cost method to measure recipients' valuations of aid programs, we tested two of the most common incentive-compatible techniques used in valuation field experiments - The Multiple-Price-List (MPL) and the Becker-De Groot-Marschak (BDM) method (Dupas and Miguel, 2017; Andersen et al., 2006). We then applied several variations to these main methods: whether a detailed training example was provided, and whether the choice would definitely determine if the recipient receives a program or cash, or would do so with some probability.

The MPL is a relatively simple method for eliciting willingness to pay. It confronts the subject with a list of discrete choices between a given amount of money or some other offering (in the case of this study, receiving a development aid intervention). Respondents then indicate whether they prefer the "money" or "offering" at each price. The experimenter then selects one offer on this list at random, and executes the respondent's choice for that offer. In this way, respondents are incentivised to select their true preference for each offer, given they have equal probability of selection.

BDM is slightly more complicated for respondents. Subjects are asked to state the monetary payment at which they are indifferent between receiving the payment and some good or offering. A random amount is then drawn from a predetermined distribution; if this amount is higher than their indifference point, they get this amount. If the amount drawn is lower than their indifference point, they receive the good or offering. They therefore have an incentive to give their true valuation, or else risk receiving a monetary value lower than their value of the good or of receiving the good instead of some preferred amount of money.

Given that BDM has been cited as the most-common incentive compatible technique for valuation in field experiments and that methods studies find it sensitive to framing effects (de Meza et al., 2013), we test two cross-randomized variations of this method. Specifically, we vary whether or not an example of the method is provided to respondents (or alternatively an appeal for the respondent to have 'faith' that it is in their best interest to report truthful valuations), and whether the specific program (out of the 18 possible programs) selected for delivery at endline is known to the respondent at elicitation (or alternatively the program is selected randomly at endline, in the same way as other elicitation methods).

The hypothetical involves a non-incentivized hypothetical question simply asking for the respondent to state their indifference point between cash and the program.

Assignment to each elicitation method and the BDM variations was random at the individual level. The 'Certainty' variation of the BDM method was cross-randomized with the other two BDM variations. The assignment to the treatment conditions and details on their administration is described for each method below:

1. **Hypothetical (H)**: Respondents were administered 18 questions asking how much cash would make them as well off as each of the aforementioned aid programs, with no mention of receiving either. (N = 136)
2. **Becker-DeGroot-Marschak with example and probabilistic program (BDMe)**: Respondents were administered 18 questions asking how much cash would make them as well off as each of the aforementioned aid programs. Respondents are told that the particular program for which their valuation will actually be applicable is determined by lottery. The BDM mechanism is explained and respondents are told that they will receive either cash or the program by a lottery. An example is provided to aid explanation. (N = 142)
3. **Becker-DeGroot-Marschak on faith and probabilistic program (BDMf)**: Respondents were administered 18 questions asking how much cash would make them as well off as each of the aforementioned aid programs. Respondents are told that the particular program for which their valuation will actually be applicable is determined by lottery. The BDM mechanism is explained and respondents are told that they will receive either cash or the program by a lottery, which is designed by scientists in such a way that it is always in their best interest to report their true valuation. (N = 133)
4. **Multiple Price List (MPL)**: Respondents were administered 18 questions where they are asked to choose between a program and a given amount of cash. If they choose the program, they are asked the same choice for a larger amount of cash. This continues until the respondent selects the cash or until an upper bound of cash is reached. Respondents are told that the particular program for which their valuation will actually be applicable is determined by lottery. (N = 117)
5. **Certainty (c)**: The same method is applied for treatments 2 and 3 above (the BDMe and BDMf). However, instead of using a lottery to determine the particular program for which their valuation will actually be applicable, respondents are told that the valuation for WASH supplies (randomly chosen ex-ante) is the one which will determine their award. Thus, they have certainty they will receive this program or cash when asked how much cash would make them as well off as the program. (N = 265 - cross-randomized with the BDMe and BDMf treatments).

2.4 Data and program delivery

2.4.1 Sample

The baseline survey was conducted with 806 individuals across the three locations mentioned above and the follow-up survey with 793 of these individuals. In Nairobi, eligible individuals include those over 18 years of age residing in low-income neighborhoods. In Nakuru and Makueni eligible individuals are those over 18 years of age residing in a home made of all or partially natural materials (e.g., wood, local stone or mud, excluding homes which include cement or cinder blocks). Eligible households were first identified, and later revisited for data collection if they met the screening criteria. Summary statistics for each treatment group and T-tests showing differences on each dimension relative to the H (Hypothetical) group are shown in Table 1. The mean age of respondents in each treatment group varied between approximately 36 and 41 years of age, and the county representation was roughly equal for each of the three counties across groups. Gender, The presence of a liquidity shock between baseline and endline, the index of total assets, and the log of monthly spending were not statistically different across groups. For some comparisons, age was significantly different between treatment groups. However, this applies for only 2 comparisons out of the full 40 - we add age as a control variable to check for robustness in the analysis.[1]

2.4.2 Recipient valuation survey

Each respondent was administered a baseline survey that elicits indifference points between cash and aid programs. The survey also measured a range of baseline characteristics. The survey was administered on tablets using SurveyCTO software built on the Open Data Kit platform. We had different survey versions where the questions on baseline characteristics were the same for all individuals but the value elicitation treatments differed as described above. As per our field protocol, each surveyor administered one of these versions to each household surveyed in turn, i.e. the first household surveyed was administered survey version 1 (where the Hypothetical elicitation treatment was used), the second household surveyed was administered survey version 2 (where BDM with Example was used), and so on. Further, within each survey version, the order in which different programs are asked about is randomized through SurveyCTO (with the exception of the last program, in order to complete the certainty treatment. The last program was randomly selected ex ante to be sanitation supplies).

[1] Age was not significantly predictive of any outcome. Overall, its inclusion made no statistical or practical difference to estimated treatment effects.

Table 1: Summary Statistics by Treatment Arm

Variable	(1) H Mean/SE	(2) BDMe Mean/SE	(3) BDMe+c Mean/SE	(4) BDMf Mean/SE	(5) BDMf+c Mean/SE	(6) MPL Mean/SE	T-test Difference (1)-(2)	(1)-(3)	(1)-(4)	(1)-(5)	(1)-(6)
County: Nairobi	0.31	0.31	0.31	0.34	0.37	0.38	0.00	0.00	-0.03	-0.05	-0.07
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)					
County: Makueni	0.33	0.30	0.32	0.33	0.38	0.36	0.02	0.00	-0.00	-0.06	-0.04
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)					
County: Nakuru	0.36	0.39	0.37	0.33	0.25	0.26	-0.02	-0.00	0.03	0.11*	0.10*
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)					
Male	0.47	0.42	0.46	0.37	0.40	0.47	0.05	0.01	0.11*	0.08	0.01
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)					
Age	36.21	35.85	36.96	37.75	39.70	41.56	0.36	-0.75	-1.54	-3.49*	-5.35***
	(1.20)	(1.20)	(1.20)	(1.29)	(1.39)	(1.47)					
Index of total assets	0.35	0.35	0.34	0.33	0.35	0.37	0.01	0.01	0.03	0.00	-0.02
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)					
Log Monthly Spending	11.41	11.44	11.36	11.37	11.32	11.36	-0.03	0.04	0.03	0.09	0.05
	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)					
Liquidity shock	0.30	0.31	0.21	0.26	0.30	0.24	-0.01	0.09	0.03	-0.00	0.05
	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)					
N	135	142	142	133	123	116					

Notes: The value displayed for t-tests are the differences in the means across the groups. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

2.4.3 Valuation method reliability survey

The purpose of the follow-up visit was to verify the initial survey visit was conducted according to protocol as well as to offer respondents a TIOLI choice between an aid program and a cash amount around the indifference point derived from their initial stated valuation. The endline offer was drawn from a random distribution with the respondent's stated indifference point as the expected mean. However, for each intervention, there was a threshold above which cash offers were capped (or 'top coded') due to budgetary restrictions.[2] This allowed us to derive predictions of respondents' endline choice. If the cash offer is greater than the respondent's previously stated indifference point, the respondent should rationally choose cash (and if it is lower, they should rationally choose the program). Respondents are therefore assigned a 'predicted choice' of either cash (30% of the sample) or program (70%) conditional on their indifference point and endline offer. If at follow-up the respondent chooses the program when cash is predicted (or vice-versa), the elicitation mechanism yields an "inconsistent" valuation. By measuring the relative proportions in each group whose choice coincides with the prediction from their valuation, this procedure allows us to test the reliability of the method in eliciting accurate valuations. These respondents then received either the program (20% of sample) or the cash offer amount (80%), depending on their choice at follow-up.

The aid programs randomly chosen for delivery included insecticide-treated bed nets, a supply of condoms, fertilizer, hygiene supplies, and donations to various programs (such as malaria or HIV research). Respondents were informed that no matter what choice they made, the cash or the program would be delivered to them at the same time as the program would be. As life events could cause the need for liquidity, we also inquired whether the household received a liquidity shock since our initial visit. Finally, the interviewer confirmed the names and ID numbers of primary male and female household members, the location of residence and phone numbers listed for MPesa transfers. This information was used to ensure transfers reach the intended recipient. These surveys were conducted by field officers other than those who conducted the initial recipient surveys. If discrepancies were found between this survey and the initial survey, transfers were delayed until issues were resolved.

2.4.4 Program delivery

If respondents chose to receive the program, the goods or services were delivered in-person by an individual not involved in the initial data collection. At that visit, the respondent's name and other details were verified. For respondents choosing cash, a transfer was sent through the MPesa digital payment platform. This platform allows the researchers to confirm the name from the survey matches the name associated with the mobile money account. Finally, we followed up with recipients (by phone or in person) to confirm receipt of goods, services or cash. Out of the 116 program recipients that were contacted, 6 reported not receiving the program; whereas for cash respondents, we were able to verify through the receipt confirmation survey and MPesa details that all received the transfer amount.

[2] For each intervention selected at the endline, top coding was at some value significantly higher than intervention cost. However, this still allows for the final choice to be assessed in relation to the original indifference point; where top coding applied, the significant majority of offers were below the indifference point.

2.4.5 Primary Outcomes

We use two outcomes to compare elicitation methods: 1) the valuation amounts from the initial elicitation, and 2) the reliability of the initial valuation relative to endline TIOLI choice, which we call 'consistency'.

Valuation amounts (Logval): our primary outcome for valuation is the natural log of the valuation stated by the recipient divided by the cost of the program in question. The log transformation was chosen given the distribution in absolute amounts is severely skewed to the right, and the natural log distribution contains the properties required for inference. For the multiple price list treatment (MPL) amounts were coded as follows before the log transformation: interval responses were coded as the switch point[3]; lower 'corner responses' (where participants always select cash) were coded as the median valuation of the hypothetical treatment group conditional on the valuation being lower than the lower bound of the MPL; upper corner responses (participant always selects program) were coded according to the same logic, at the median valuation of the hypothetical treatment group conditional on the valuation being *higher* than the *upper bound* of the MPL. We apply a range of alternative codings of the corner-responses to check for robustness (results in the Appendix).[4] This outcome is at a program by individual level.

Consistency (Consistent): This is a binary outcome, taking on 1 if the choice at endline is consistent with the predicted choice from the indifference point and the TIOLI cash offer, and taking on 0 otherwise. This applies the same MPL corner-response coding indicated above. This outcome is at an individual level.

3 Empirical Strategy

The full regression model for each outcome is specified below in Equations 1 and 2 respectively, where $Logval_{ij}$ in Equation 1 represents the natural log of valuation amount for respondent i and intervention j, and $Consistent_i$ in Equation 2 represents the dummy variable outcome for consistency for respondent i. In both equations, $BDMe_i$, $BDMf_i$, and MPL_i represent dummy variables for each of the elicitation methods for respondent i, and c_i represents the cross-randomized certainty treatment for respondent i. The hypothetical question serves as the reference category. Therefore, in Equation 1 β_1, β_2, β_3 and β_4 represent the difference in valuation due to each elicitation method relative to the valuation from a hypothetical question. In Equation 2 they represent the difference in consistency for each method (i.e., whether the choice in a TIOLI offer is predicted by the valuation elicited by that method) relative to the consistency of a hypothetical valuation.

[3]The value at which participants indicated that they would select the cash amount instead of the program.

[4]There were 4 alternative codings of corner-responses, representing more extreme assumptions. The first two exploit different extremes in the range of responses: "MPL - Bounds/Low-range" takes lower and upper bound valuations at the lower and upper bound of the MPL respectively; "MPL - High-range" takes these as the 25th percentile for hypothetical valuations conditional on being lower than the lower bound, and the 75th percentile for hypothetical valuations conditional on being higher than the upper bound, respectively. The second two exploit extremes in lower or higher valuation amounts: For the lower and upper bound responses respectively, "MPL - Low values" takes the 25th percentile for hypothetical valuations conditional on being lower than the lower bound, and takes the upper bound itself; "MPL - High values" takes the lower bound itself and the the 75th percentile for hypothetical valuations conditional on being higher than the upper bound, respectively

$$Logval_{ij} = \beta_0 + \beta_1 BDMe_i + \beta_2 BDMf_i + \beta_3 MPL_i + \beta_4 c_i + \beta_x Xi + \epsilon_{ij} \quad (1)$$

In Equation 1, X_i represents the vector of control variables X for respondent i, and e_{ij} is the error term for respondent i and intervention j, and is clustered by respondent.

$$Consistent_i = \beta_0 + \beta_1 BDMe_i + \beta_2 BDMf_i + \beta_3 MPL_i + \beta_4 c_i + \beta_p Pi + \epsilon_i \quad (2)$$

In equation 2, P_i represents the vector of predictors of consistency (which also serve as control variables) for respondent i; these include aid intervention fixed effects, field officer fixed effects, location fixed effects, gender, age, and several socio-economic variables: the presence of a liquidity shock, an index of the value of household assets, and a logged indicator of spending the month of the interview. e_i is the error term for respondent i.

We estimate each equation using OLS. We additionally estimate two more parsimonious variations of the above models. First by omitting the control variables and second omitting the certainty treatment conditions (which were cross-randomized across other conditions).

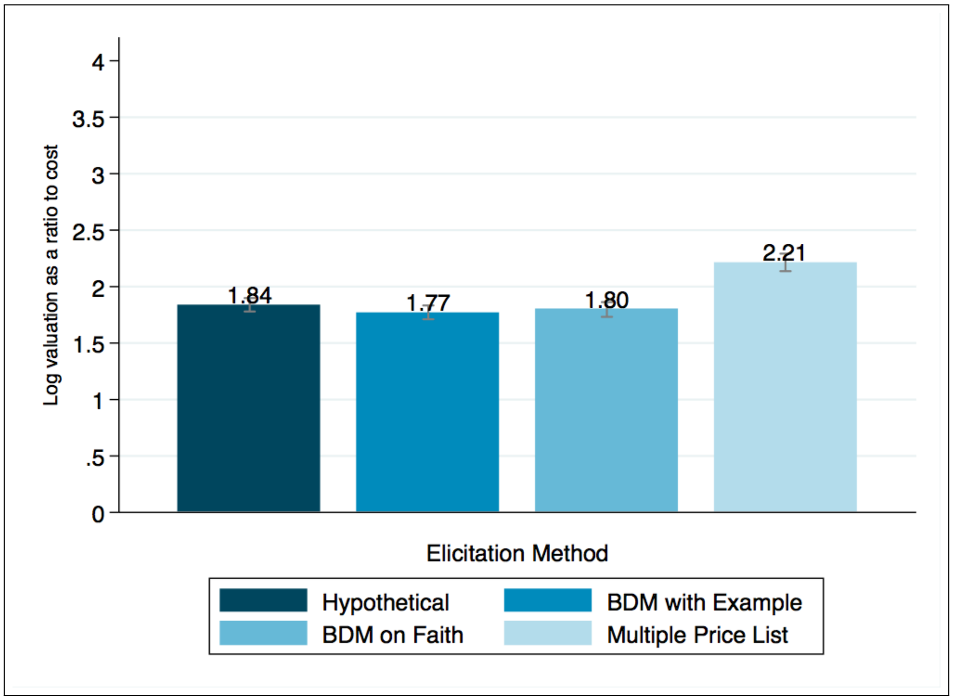
4 Results

4.1 Valuation amounts

Figure 1 shows the mean of the (log) ratio of recipient's valuations for interventions relative to the cost of the interventions across elicitation techniques. This figure shows that valuations elicited by the incentivized BDM are no different than those from a simple hypothetical question. Moreover, there is no difference in valuations when the BDM mechanism is explained in a detailed example compared to asking the respondent to take the experimenter's word it is in their interest to report their truthful valuation (p-value = 0.575).

As seen in Figure 2, knowing that the valuation response will definitely determine whether a recipient receives the program or cash (rather than knowing there is a chance the valuation will do so) increases valuation. This increase is statistically significant (Table 2). We speculate two possible explanations for this result. First, this might be due to demand effects. Given that we indicate to respondents this is the selected intervention out of 18 possible interventions, it is plausible that this serves as a signal from the experimenter that this intervention has relatively higher value - it appears researchers are 'pushing this program'. A second explanation is that some misunderstanding of the method led to a perceived strategic incentive to overstate true valuations. The certainty of this being the intervention rolled out might magnify this perceived strategic incentive to overstate. It is also important to note that the certainty sub-treatment was only evaluated for one good (WASH supplies), whereas other treatments included 18 interventions. Thus it is possible the result of the Certainty treatment is specific to this

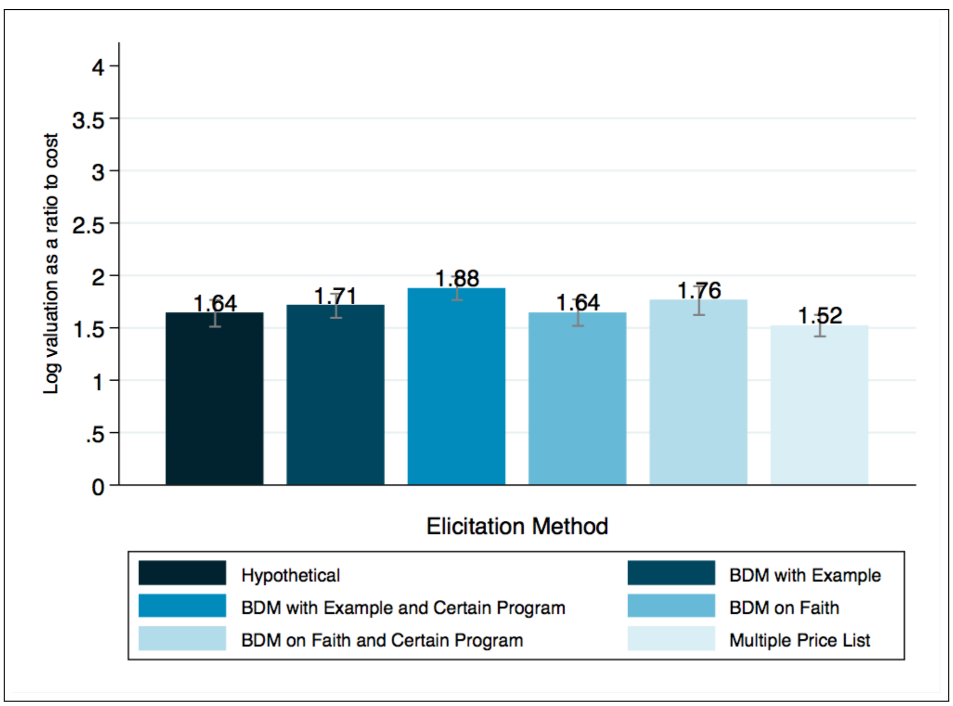
Figure 1: Natural-log of valuation amounts by treatment group (certainty excluded)



Bars represent means of the natural log of valuation amount as a ratio of program cost for each treatment. Lines represent 95% confidence intervals

particular intervention, and may not generalize.

Figure 2: Natural-log of valuation amounts by treatment group (WASH supplies)



Bars represent means of the natural log of valuation amount as a ratio of program cost for each treatment. Lines represent 95% confidence intervals

Table 2: Regression Output for Valuation of Programs

	No Certainty	**No Certainty + Controls**	**WASH only + Certainty**	**WASH only + Certainty + Controls**
Multiple Price List	0.38***	0.38***	-0.12	-0.12
	(0.05)	(0.05)	(0.08)	(0.08)
BDM with Example	-0.07	-0.05	0.08	0.08
	(0.04)	(0.04)	(0.08)	(0.08)
BDM on Faith	-0.04	0.01	-0.01	-0.01
	(0.04)	(0.04)	(0.09)	(0.09)
Certain Payoff			0.14**	0.14**
			(0.06)	(0.06)
Constant	1.84***	1.06***	1.64***	1.64***
	(0.03)	(0.30)	(0.07)	(0.07)
Certainty Treatments	No	No	Yes	Yes
Control Variables	No	Yes	No	Yes
Observations	8,774	8,774	793	793

Table shows output for regressions on the natural log of valuation amount as a ratio of intervention cost. Standard errors are in parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level. Standard errors are clustered at he respondent level.

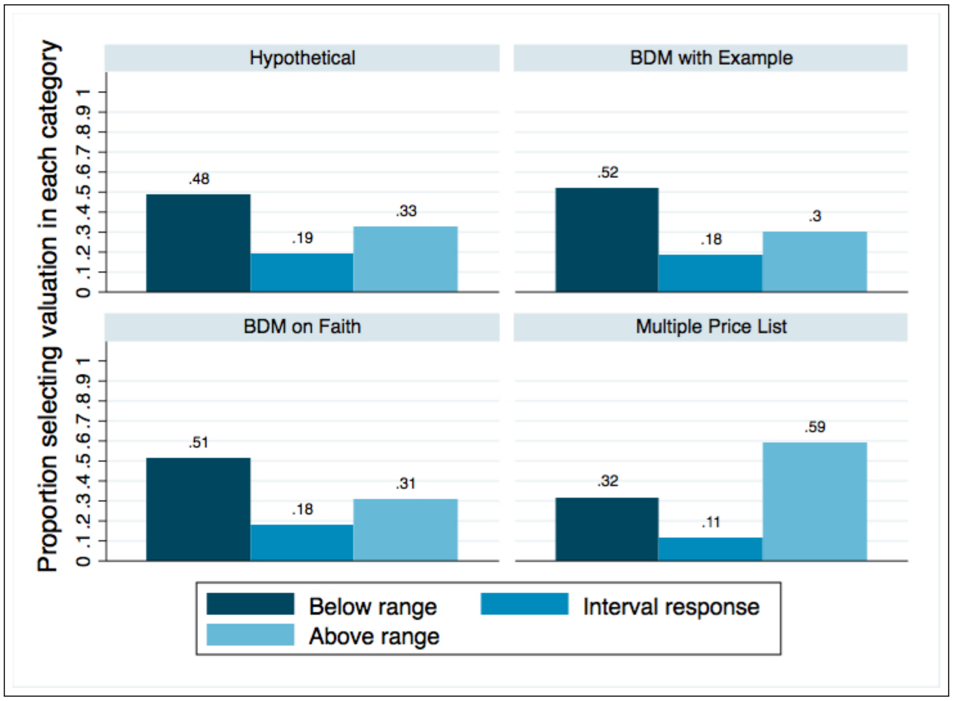
Finally, Figure 1 and Table 2 indicate that the MPL methodology leads to higher valuations on average.[5] We believe this is a result of the specific interventions we selected and the range of the MPL menu presented for those interventions rather than a universal feature of the MPL method. The evidence suggests that MPL tends to push valuations towards the upper end of the MPL menu: Figure 3 shows that 60% of respondents select the highest MPL option, which is higher than the percentage of respondents in other treatments that report a valuation above the highest MPL option. However, Figure 4 illustrates that whether this phenomenon results in a higher valuation compared to other methods depends on where the bounds of the MPL menu are set. Only if the MPL menu includes the full range of valutions from other methodologies would MPL always generate higher valuations.

4.2 Results for "Consistency"

Whether a particular valuation method influences stated valuations at the margin may not matter if the prediction based on that response is valid. We assess whether this may be the case by considering consistency of choices between cash and programs with previously stated valuations for those programs. What is most striking from this analysis overall is the small proportion of people who make consistent choices (see Figure 5). Specifically, we find that only 40% of people make the choice between cash and a program that was predicted by their stated valuation for the program.

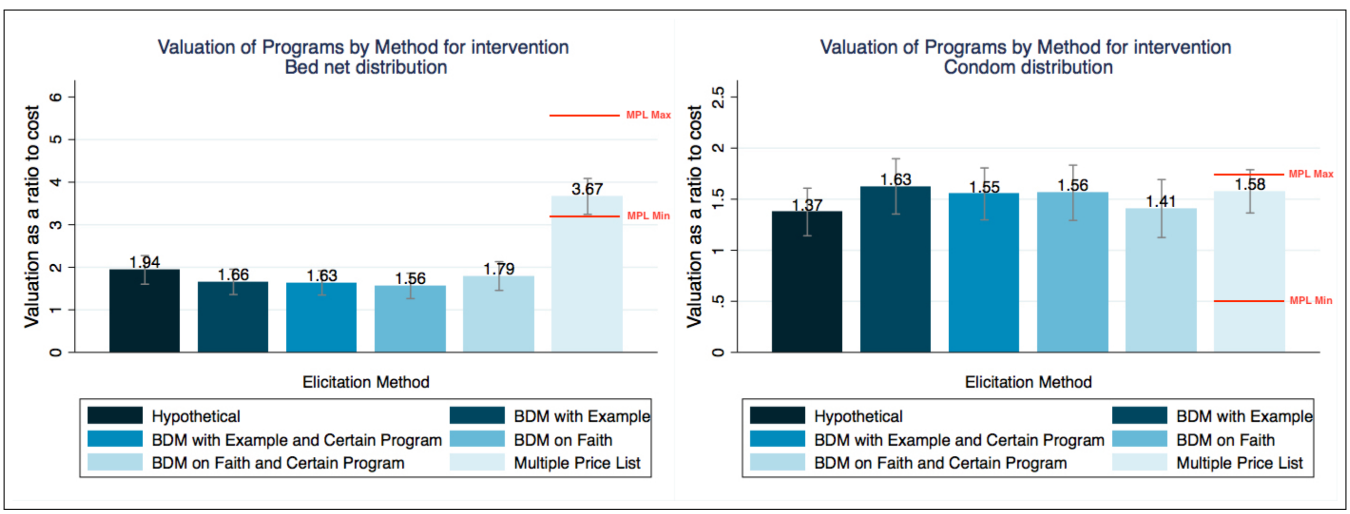
[5]this result is robust to most alternative codings of the MPL corner responses, see Appendix Figure A1 and Table A1

Figure 3: Percentage of MPL corner and interior responses by methodology



This graphs shows the proportion of valuations by elicitation method that fall below, within, or above the range of interior responses on the MPL for each aid program. That is, it collapses responses for non-MPL treatments to the MPL distribution.

Figure 4: Valuation amounts by Treatment Condition for 2 individual interventions



Bars represent means of the log of valuation amount as a ratio of program cost for each treatment. Lines represent 95% confidence intervals

At first pass, one might not perceive 40% consistent choices as problematic: respondents were presented with a TIOLI offer between the programs and a cash offer near their indifference point, so simple noise could generate "inconsistency" rates of 50%. However, inconsistency is not explained by pure noise. Three factors suggest that random variation is not sufficient to explain the inconsistency we find. First, people are ~ 60 percentage points more likely to be consistent when cash is the predicted endline choice

Table 3: Regression Output for Consistency

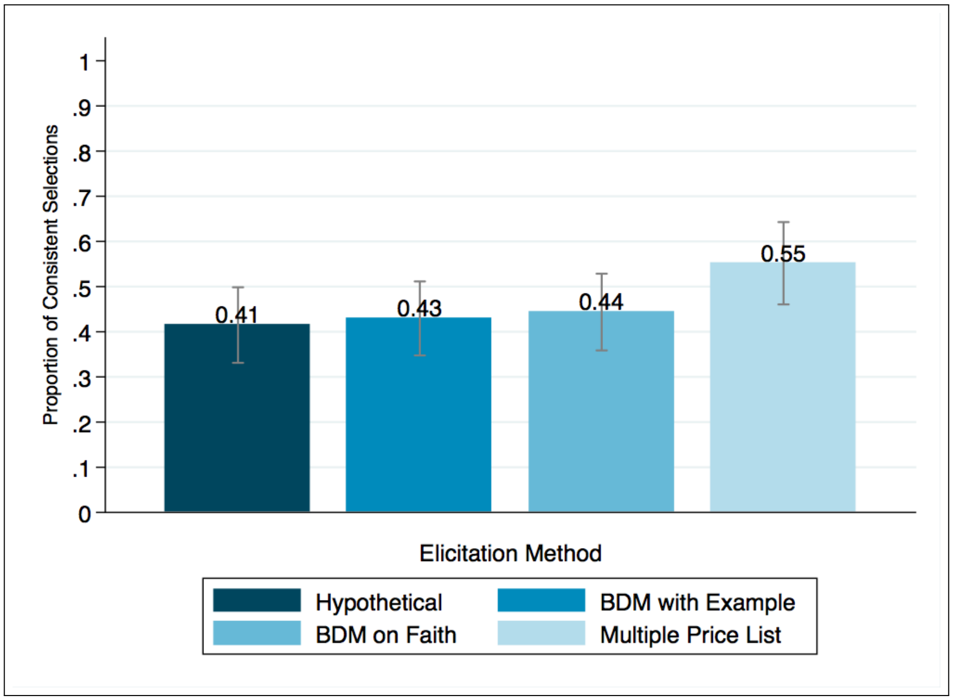
	No Certainty	No Certainty + Predictors	WASH only + Certainty	WASH only + Certainty + Predictors
Multiple Price List	0.14**	0.20***	0.09	0.10
	(0.06)	(0.06)	(0.15)	(0.14)
BDM with Example	0.01	0.04	0.04	0.13
	(0.06)	(0.05)	(0.14)	(0.13)
BDM on Faith	0.03	0.03	0.05	0.12
	(0.06)	(0.05)	(0.13)	(0.12)
Certain Payoff			-0.27***	-0.20**
			(0.09)	(0.09)
Predictors				
Gender = Female		-0.00		-0.01
		(0.04)		(0.05)
Liquidity shock		-0.02		0.04
		(0.04)		(0.05)
Index of total assets		0.16		0.36*
		(0.15)		(0.20)
Log of Monthly Spending		0.04*		-0.01
		(0.03)		(0.03)
Cash was predicted choice at endline		0.65***		0.61***
		(0.04)		(0.06)
Absolute difference between offer and valuation		-0.10***		-0.01
		(0.03)		(0.07)
Constant	0.41***	-0.38	0.48***	0.39
	(0.04)	(0.37)	(0.10)	(0.50)
Certainty Treatments	No	No	Yes	Yes
Control Variables	No	Yes	No	Yes
Observations	526	526	327	327

Table shows output for regressions on the 'consistent' variable (dummy variable indicating whether endline choice is consistent with predicted choice from valuation). Standard errors are in parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

than when the program is the predicted choice (Figure 6). If it were noise explaining inconsistency, we would expect this inconsistency in 'both directions'. Second, the absolute difference between the offer and the valuation does not increase consistency (see Table 3): if there were some noise in initial valuations, we would expect offers that deviate more from the initial valuation to lead to greater consistency.[6] Third, in

[6]In Table 3 the effect of larger differences between offers and valuations is actually *negative*, though this is not robust to alternative codings of the MPL - see the Appendix

Figure 5: Consistency by treatment group (certainty excluded)



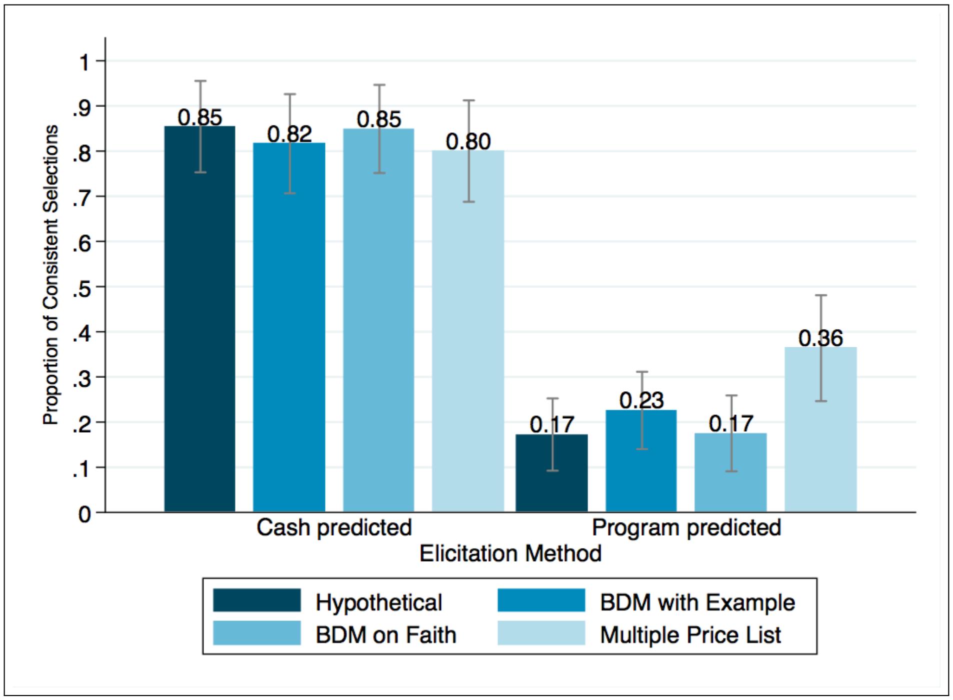
Bars represent the proportion of respondents whose choices at endline are consistent with the predicted choice conditional on valuations. Lines represent 95% confidence intervals

a high number of cases, respondents choose cash offers substantially below their stated valuations instead of choosing the program: the mean accepted cash offer is Ksh 1655, while the mean valuation for those who accepted cash instead of the program is Ksh 3500. This suggests that respondents are biased towards cash and systematically overstate their valuations.

Finally, the bias towards cash is not explained away by factors that might indicate an immediate need for liquidity. While people with greater assets and spending are slightly more consistent, those experiencing a recent liquidity shock are not any more or less consistent than others (see Table 3).

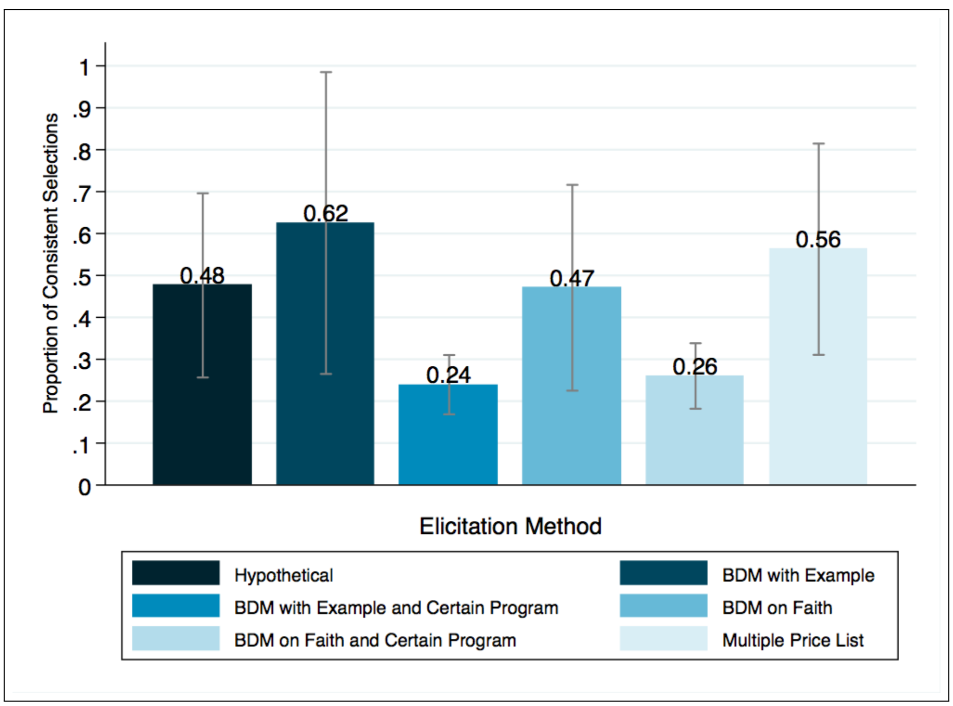
Turning to the relative consistency across value elicitation methods, again we find that incentivized BDM does not generate any more valid responses (gauged by the consistency of the choices the responses imply) than a hypothetical question (Table 3). In some specifications, it appears that the MPL method generates more consistent choices than other methods, but this result is not robust to alternative approaches to coding corner responses in the MPL method (See Appendix Figure A2 and Table A2). Finally, Figure 7 and Table 3 suggest that the Certainty treatment leads to choices that are less consistent with the elicited valuation. This result is due to the fact that the Certainty treatment leads to higher valuations, and most respondents choose cash instead of the program, which leads to higher rates of inconsistency.

Figure 6: Consistency by treatment group by predicted choice



Bars represent the proportion of respondents whose choices at endline are consistent with the predicted choice conditional on valuations. Lines represent 95% confidence intervals

Figure 7: Consistency by treatment group (WASH supplies)



Bars represent the proportion of respondents whose choices at endline are consistent with the predicted choice conditional on valuations. Lines represent 95% confidence intervals

5 Discussion and Conclusion

Overall, our results suggest that valuations of development aid programs elicited using common incentive compatible techniques are generally not consistent with responses to a concrete TIOLI offer and are not different to valuations elicited by hypothetical questions. They also show that valuations are sensitive to both the elicitation method and to framing. Taken together, consumers of valuation outputs should read results with caution and take account of the specific elicitation method and framing of the measurement tools. In particular, the BDM method performs no differently to a hypothetical question in both valuation amounts and consistency; this is a relatively novel and surprising result, given the range of studies that rely on this method to derive important parameters and its virtues of incentive compatibility and precision in parameter estimates (Cohen et al., 2010; Guiteras et al., 2016a; Berry et al., 2020; Dupas and Miguel, 2017). The failure of the example intervention in the BDMe sub-treatment further suggests that mechanism misunderstanding may not be the explanation for the lack of consistency in this method. (de Meza et al., 2013) describe mechanism misunderstanding as one explanation for systematic overstatement of valuations (which seems to be at the core of the inconsistency observed in this study); yet we don't see evidence of this variation affecting valuation amounts or consistency.

With regards to framing effects of various elicitation methods, we find that the Multiple Price List (MPL) method must be approached carefully. In particular, this method seems to nudge recipients towards the upper edge of the MPL menu, and consequently the valuation obtained by this method will be highly sensitive to the range selected for the menu. Similar framing effects are well-established in the literature (Andersen et al., 2006; de Meza et al., 2013; Plott and Zeiler, 2005), and our results reinforce their validity and importance.

The Certainty treatment leads to the highest valuations and the lowest consistency rates - this suggests that valuations under this treatment condition include the highest 'overstatements' in initial valuations. We speculate two possible explanations for this result. First, this might be due to demand effects. Participants might have perceived the selection of this specific intervention to indicate the experimenter's belief that it is more valuable. A relatively extensive literature supports the existence of demand effects in development economics experiments (Zizzo, 2010; De Quidt et al., 2018; Cilliers et al., 2015) and in valuation studies specifically (Dupas and Miguel, 2017). A second explanation is that this is due to some misunderstanding of the method, leading to a strategic incentive to overstate true valuations. Cason and Plott (2014) indicate that overstated valuations in BDM arise from perceived strategic incentives to do so. The presence of systematic overstatements across our interventions suggest this could be the case. This is a somewhat surprising result, as making the choice more likely to be realized would in theory incentivize respondents to pay more attention to generating truthful responses. Given our results pertaining to certainty apply to a single valuation, additional research should test whether this is a general phenomenon.

In spite of these differences between treatments, overall we observe a high level of inconsistency between valuations by any method and a TIOLI offer (60% average inconsistent choices across treatments), and specifically a bias towards cash in favour of the program, suggesting systematic overstatement of valuations. This finding aligns with much of the literature, which finds that valuations are frequently overstated through contingent valuation (Kremer et al., 2011), which approximates our hypothetical treatment, and through BDM (Peletz et al., 2017; de Meza et al., 2013). However, the degree of overstatement is more pronounced in this study than in this similar work. This result is also consistent with recent work in experimental economics that finds that hypothetical questions perform as well as incentive compatible techniques at predicting other incentivised behaviours in a lab and in predicting real-life behaviours (Falk et al., 2016, 2018)

Our findings suggest that valuations for development programs can be obtained cheaply - since incentivized methods do not generate meaningful differences compared with a hypothetical question. However, policy makers should consider subtle influences on valuation from the design of the instrument, and select appropriate methods to minimize these. It is further important to realize that while there may be some content in valuations, they are highly unreliable relative to a concrete TIOLI choice. Consumers of valuations should use caution if they intend to use valuations for parameter estimates, as a means to price goods, or as a means to determine whether something is valued more or less than cost (for a 'go-no-go' decision). Rather, these methods can serve as means to compare valuations across goods as overstatement should apply for each in a similar way.

References

Evidence Action. Deworm the world. Available at url https://www.evidenceaction.org/dewormtheworld-2/ (2020/06/12).

Steffen Andersen, Glenn W Harrison, Morten Igel Lau, and E Elisabet Rutström. Elicitation using multiple price list formats. *Experimental Economics*, 9(4):383–405, 2006.

Steffen Andersen, Glenn W Harrison, Morten I Lau, and E Elisabet Rutström. Eliciting risk and time preferences. *Econometrica*, 76(3):583–618, 2008.

Mir Anjum Altaf and Jeffrey A Hughes. Measuring the demand for improved urban sanitation services: results of a contingent valuation study in ouagadougou, burkina faso. *Urban studies*, 31(10):1763–1776, 1994.

James Berry, Greg Fischer, and Raymond Guiteras. Eliciting and utilizing willingness to pay: Evidence from field trials in northern ghana. *Journal of Political Economy*, 128(4):1436–1473, 2020.

Peter Bohm, Johan Lindén, and Joakim Sonnegård. Eliciting reservation prices: Becker–degroot–marschak mechanisms vs. markets. *The Economic Journal*, 107(443):1079–1089, 1997.

Robert A Bohm, Timothy J Essenburg, and William F Fox. Sustainability of potable water services in the philippines. *Water Resources Research*, 29(7):1955–1963, 1993.

Kevin J Boyle. Contingent valuation in practice. In *A primer on nonmarket valuation*, pages 83–131. Springer, 2017.

Sarah Brebner and Joep Sonnemans. Does the elicitation method impact the wta/wtp disparity? *Journal of behavioral and experimental economics*, 73:40–45, 2018.

Timothy N Cason and Charles R Plott. Misconceptions and game form recognition: Challenges to theories of revealed preference and framing. *Journal of Political Economy*, 122(6):1235–1270, 2014.

Jacobus Cilliers, Oeindrila Dube, and Bilal Siddiqi. The white-man effect: How foreigner presence affects behavior in experiments. *Journal of Economic Behavior & Organization*, 118:397–414, 2015.

Jessica Cohen, Pascaline Dupas, et al. Free distribution or cost-sharing? evidence from a randomized malaria prevention experiment. *Quarterly journal of Economics*, 125(1):1, 2010.

Shawn Cole, Xavier Giné, and James Vickery. How does risk management influence production decisions. *Evidence from a field experiment. World Bank Policy Research Working Paper*, 6546, 2013.

Shawn Allen Cole and A Fernando. 'mobile'izing agricultural advice: Technology adoption, diffusion and sustainability. *Harvard Business School Finance Working Paper*, (13-047), 2016.

David de Meza, Diane Reyniers, et al. Debiasing the becker–degroot–marschak valuation mechanism. *Economics Bulletin*, 33(2):1446–1456, 2013.

Jonathan De Quidt, Johannes Haushofer, and Christopher Roth. Measuring and bounding experimenter demand. *American Economic Review*, 108(11):3266–3302, 2018.

Pascaline Dupas and Edward Miguel. Impacts and determinants of health levels in low-income countries. In *Handbook of economic field experiments*, volume 2, pages 3–93. Elsevier, 2017.

Armin Falk, Anke Becker, Thomas J Dohmen, David Huffman, and Uwe Sunde. The preference survey module: A validated instrument for measuring risk, time, and social preferences. 2016.

Armin Falk, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde. Global evidence on economic preferences. *The Quarterly Journal of Economics*, 133(4):1645–1692, 2018.

Shane Frederick, George Loewenstein, and Ted O'donoghue. Time discounting and time preference: A critical review. *Journal of economic literature*, 40(2):351–401, 2002.

A Myrick Freeman III, Joseph A Herriges, and Catherine L Kling. *The measurement of environmental and resource values: theory and methods*. Routledge, 2014.

Raymond Guiteras, David I Levine, Thomas Polley, and Brian Quistorff. Credit constraints, discounting and investment in health: Evidence from micropayments for clean water in dhaka. *North Carolina State University*, 2016a.

Raymond P Guiteras, David I Levine, Stephen P Luby, Thomas H Polley, Kaniz Khatun-e Jannat, and Leanne Unicomb. Disgust, shame, and soapy water: tests of novel interventions to promote safe water and hygiene. *Journal of the Association of Environmental and Resource Economists*, 3(2):321–359, 2016b.

Glenn W Harrison, Ronald M Harstad, and E Elisabet Rutström. Experimental methods and elicitation of values. *Experimental economics*, 7(2):123–140, 2004.

Charles A Holt and Susan K Laury. Risk aversion and incentive effects. *American economic review*, 92 (5):1644–1655, 2002.

J-Pal. Preventing-malaria: Free insecticide-treated bednets. Available at nurlhttps://www.povertyactionlab.org/case-study/preventing-malaria-free-insecticide-treated-bednets (2020/06/12).

Michael Kremer and Edward Miguel. The illusion of sustainability. *The Quarterly journal of economics*, 122(3):1007–1065, 2007.

Michael Kremer, Jessica Leino, Edward Miguel, and Alix Peterson Zwane. Spring cleaning: Rural water impacts, valuation, and property rights institutions. *The Quarterly Journal of Economics*, 126(1): 145–205, 2011.

Lori Leonard, Samba Diop, Seydou Doumbia, Aboubacar Sadou, Jules Mihigo, Hannah Koenker, Sara Berthe, April Monroe, Kathryn Bertram, and Rachel Weber. Net use, care and repair practices following a universal distribution campaign in mali. *Malaria journal*, 13(1):435, 2014.

Karine Nyborg. Environmental valuation, cost-benefit analysis and policy making: A survey. *Statistics Norway, Research Department. Documents*, 96:12, 1996.

World Health Organization. Achieving universal coverage with long-lasting insecticidal nets in malaria control. Technical report, Geneva, 2014.

Rachel Peletz, Alicea Cock-Esteb, Dorothea Ysenburg, Salim Haji, Ranjiv Khush, and Pascaline Dupas. Supply and demand for improved sanitation: results from randomized pricing experiments in rural tanzania. *Environmental science & technology*, 51(12):7138–7147, 2017.

Charles R Plott and Kathryn Zeiler. The willingness to pay-willingness to accept gap, the" endowment effect," subject misconceptions, and experimental procedures for eliciting valuations. *American Economic Review*, 95(3):530–545, 2005.

Jeremy Shapiro. Exploring recipient preferences and allocation mechanisms in the distribution of development aid. *The World Bank Economic Review*, 2019a.

Jeremy Shapiro. The impact of recipient choice on aid effectiveness. *World development*, 116:137–149, 2019b.

Vernon L Smith. Microeconomic systems as an experimental science. *The American Economic Review*, 72(5):923–955, 1982.

Elizabeth R Stevens, Abigail Aldridge, Yawo Degbey, Akou Pignandi, Monique A Dorkenoo, and Justin Hugelen-Padin. Evaluation of the 2011 long-lasting, insecticide-treated net distribution for universal coverage in togo. *Malaria journal*, 12(1):162, 2013.

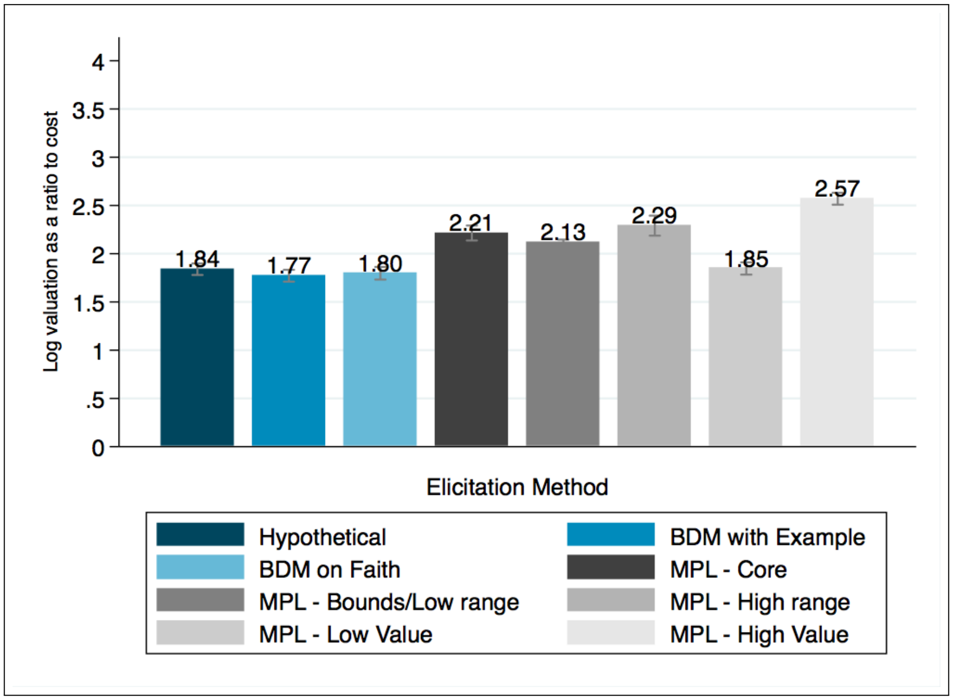
Brent M Swallow and Mulatu Woudyalew. Evaluating willingness to contribute to a local public good: application of contingent valuation to tsetse control in ethiopia. *Ecological Economics*, 11(2):153–161, 1994.

Dale Whittington, V Kerry Smith, Apia Okorafor, Augustine Okore, Jin Long Liu, Alexander McPhail, et al. Giving respondents time to think in contingent valuation studies: a developing country application. *Journal of Environmental Economics and Management*, 22(3):205–225, 1992.

Daniel John Zizzo. Experimenter demand effects in economic experiments. *Experimental Economics*, 13 (1):75–98, 2010.

Appendix

Figure A1: Log Valuation by treatment group - Alternative MPL codings Included



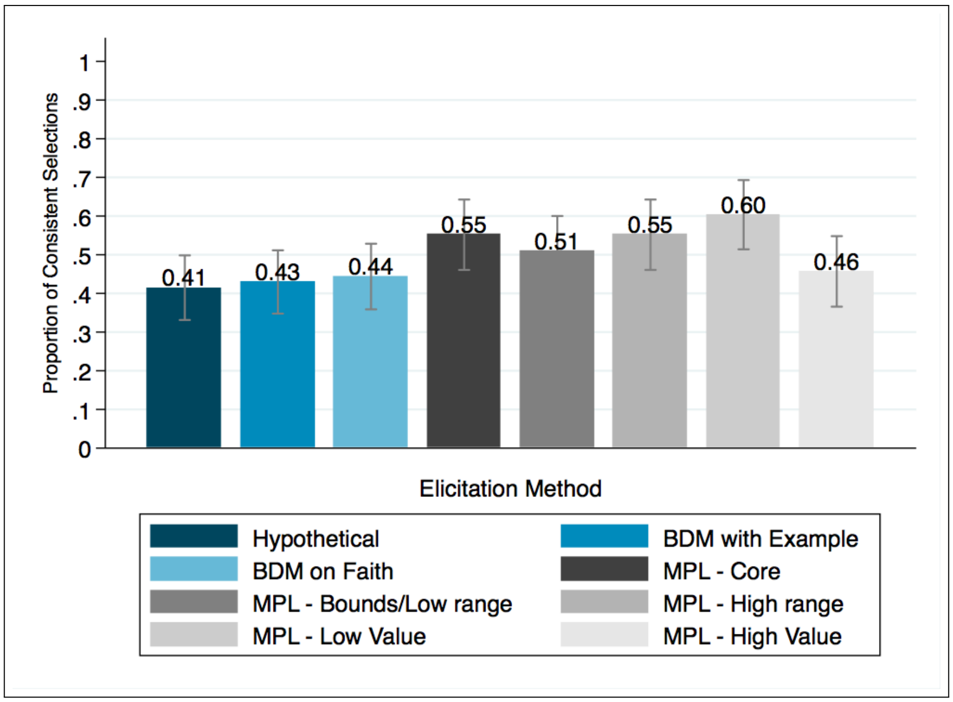
Bars represent means of the natural log of valuation amount for each treatment. Lines represent 95% confidence intervals. "MPL - core" takes the valuations as in the core coding used in the primary analysis in this paper. "MPL - Bounds/Low-range" takes lower and upper bound valuations at the lower and upper bound of the MPL respectively; "MPL - High-range" takes these as the 25th percentile for hypothetical valuations conditional on being lower than the lower bound, and the 75th percentile for hypothetical valuations conditional on being higher than the upper bound, respectively. "MPL - Low values" takes the 25th percentile for hypothetical valuations conditional on being lower than the lower bound, and takes the upper bound itself; "MPL - High values" takes the lower bound itself and the the 75th percentile for hypothetical valuations conditional on being higher than the upper bound, respectively

Table A1: Regression Analysis of Log Valuation with Alternative MPL Coding

	MPL - Bounds		MPL - High range		MPL - Low value		MPL - High value	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Multiple Price List	0.29***	0.30***	0.45***	0.46***	0.01	0.02	0.73***	0.74***
	(0.03)	(0.03)	(0.06)	(0.06)	(0.04)	(0.04)	(0.04)	(0.04)
BDM with Example	-0.07	-0.05	-0.07	-0.05	-0.07	-0.04	-0.07	-0.05
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
BDM on Faith	-0.04	0.01	-0.04	0.01	-0.04	0.01	-0.04	0.01
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Constant	1.84***	1.15***	1.84***	1.02***	1.84***	1.10***	1.84***	1.07***
	(0.03)	(0.26)	(0.03)	(0.34)	(0.03)	(0.28)	(0.03)	(0.29)
Observations	8774	8774	8774	8774	8774	8774	8774	8774
Control Variables	No	Yes	No	Yes	No	Yes	No	Yes
Certainty Treatment	No	No	No	No	No	No	No	No

Table shows regression output for the dependent variable natural log of valuation amount as a ratio of intervention cost for the alternative codings of MPL corner responses. Standard errors are in parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level. Standard errors are clustered at he respondent level. "MPL - core" takes the valuations as in the core coding used in the primary analysis in this paper. "MPL - Bounds/Low-range" takes lower and upper bound valuations at the lower and upper bound of the MPL respectively; "MPL - High-range" takes these as the 25th percentile for hypothetical valuations conditional on being lower than the lower bound, and the 75th percentile for hypothetical valuations conditional on being higher than the upper bound, respectively. "MPL - Low values" takes the 25th percentile for hypothetical valuations conditional on being lower than the lower bound, and takes the upper bound itself; "MPL - High values" takes the lower bound itself and the the 75th percentile for hypothetical valuations conditional on being higher than the upper bound, respectively

Figure A2: Consistency by treatment group - Alternative MPL codings Included



Bars represent the proportion of respondents whose choices at endline are consistent with the predicted choice conditional on valuations. Lines represent 95% confidence intervals. "MPL - core" takes the valuations as in the core coding used in the primary analysis in this paper. "MPL - Bounds/Low-range" takes lower and upper bound valuations at the lower and upper bound of the MPL respectively; "MPL - High-range" takes these as the 25th percentile for hypothetical valuations conditional on being lower than the lower bound, and the 75th percentile for hypothetical valuations conditional on being higher than the upper bound, respectively. "MPL - Low values" takes the 25th percentile for hypothetical valuations conditional on being lower than the lower bound, and takes the upper bound itself; "MPL - High values" takes the lower bound itself and the the 75th percentile for hypothetical valuations conditional on being higher than the upper bound, respectively

Table A2: Regression Analysis of Consistency with Alternative MPL Coding

	MPL - Bounds		MPL - High range		MPL - Low value		MPL - High value	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Multiple Price List	0.09	0.11*	0.14**	0.22***	0.19***	0.12*	0.04	0.19***
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)
BDM with Example	0.01	0.03	0.01	0.03	0.01	0.04	0.01	0.03
	(0.06)	(0.05)	(0.06)	(0.05)	(0.06)	(0.05)	(0.06)	(0.05)
BDM on Faith	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
	(0.06)	(0.05)	(0.06)	(0.05)	(0.06)	(0.05)	(0.06)	(0.05)
Female		0.05		-0.00		0.02		0.02
		(0.04)		(0.04)		(0.04)		(0.04)
Liquidity shock		0.01		-0.03		-0.00		-0.01
		(0.04)		(0.04)		(0.04)		(0.04)
Index of total assets		0.22		0.15		0.14		0.24
		(0.16)		(0.16)		(0.16)		(0.15)
Log Monthly Spending		0.04*		0.04		0.04		0.04*
		(0.02)		(0.03)		(0.03)		(0.02)
Cash predicted choice		0.64***		0.66***		0.64***		0.66***
		(0.04)		(0.04)		(0.04)		(0.04)
Absolute offer 'gap'		-0.06		-0.07***		-0.05*		-0.07
		(0.04)		(0.03)		(0.02)		(0.04)
Constant	0.41***	-0.57*	0.41***	-0.37	0.41***	-0.46	0.41***	-0.48
	(0.04)	(0.34)	(0.04)	(0.34)	(0.04)	(0.34)	(0.04)	(0.33)
Observations	9468	526	9468	526	9468	526	9468	526
Control Variables	No	Yes	No	Yes	No	Yes	No	Yes
Certainty Treatment	No	No	No	No	No	No	No	No

Table shows regression output for the dependent variable 'consistent' for the alternative codings of MPL corner responses. Standard errors are in parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

"MPL - core" takes the valuations as in the core coding used in the primary analysis in this paper. "MPL - Bounds/Low-range" takes lower and upper bound valuations at the lower and upper bound of the MPL respectively; "MPL - High-range" takes these as the 25th percentile for hypothetical valuations conditional on being lower than the lower bound, and the 75th percentile for hypothetical valuations conditional on being higher than the upper bound, respectively. "MPL - Low values" takes the 25th percentile for hypothetical valuations conditional on being lower than the lower bound, and takes the upper bound itself; 'MPL - High values" takes the lower bound itself and the the 75th percentile for hypothetical valuations conditional on being higher than the upper bound, respectively