Giving stated preference respondents “time to think”: results from four countries

Joseph Cook [a]
Marc Jeuland [b]
Brian Maskery [c]
Dale Whittington [d,e]

[a] Evans School of Public Affairs, University of Washington, USA
[b] Sanford School of Public Policy, Duke University, USA
[c] International Vaccine Institute, Seoul, Korea
[d] Departments of Environmental Sciences & Engineering, and City & Regional Planning,University of North Carolina at Chapel Hill, USA
[e] Manchester Business School UK

Abstract

Previous studies have found that contingent valuation (CV) respondents who are given overnight to reflect on a CV scenario have 30 - 40% lower average willingness-to-pay (WTP) than respondents who are interviewed in a single session. This “time to think” (TTT) effect could explain much of the gap between real and hypothetical WTP observed in experimental studies. Yet giving time to think is still rare in binary or multinomial discrete choice studies. We review the literature on increasing survey respondents’ opportunities to reflect on their answers and synthesize results from parallel TTT studies on private vaccine demand in four countries. Across all four countries, we find robust and consistent evidence from both raw data and multivariate models for a TTT effect: giving respondents overnight to think reduced the probability that a respondent said he or she would buy the hypothetical vaccines. Average WTP fell approximately 40%. Respondents with time to think were also more certain of their answers, and a majority said they used the opportunity to consult with their spouse or family. We conclude with a discussion of why researchers might be hesitant to adopt the TTT methodology.

DRAFT AUGUST 2011

Giving stated preference respondents "time to think": results from four countries

Introduction

Stated preference (SP) valuation methods are now widely used throughout the world to study individual and household preferences in various policy sectors (environment, transportation, health, social welfare programs, etc.). From its inception, stated preference research has been subject to the criticism that it suffers from hypothetical and other biases. The charge is that hypothetical responses do not predict actual behavior, and that welfare estimates arising from these studies overestimate true willingness-to-pay. There have now been over two decades of careful empirical work that has improved SP methodologies, including the well-known NOAA panel guidelines, the inclusion of "cheap talk" scripts, and the recent attention to the incentive compatibility of SP scenarios (Carson and Groves 2007)[1]. At virtually every turn, most SP researchers tend to make research design and analysis choices that will provide the most conservative estimates of willingness to pay (WTP) in order to guard against charges of inflated welfare estimates.

Hypothetical bias probably remains, however (Harrison 2006). The rise of behavioral economics may soon focus more criticism on this bias in stated preference studies. We discuss one SP design feature – "time to think" – that has not yet been widely used, yet may contribute to narrowing the gap between real and hypothetical WTP estimates. Many real-world situations, either in the context of voting for referenda or private purchasing decisions, potentially involve some degree of deliberation. Many consumers and voters will want to think carefully about the program or referendum with regards to their budget or their annual tax bill before reaching their decision. They may want to confer with partners, family members or friends. They may seek out alternative sources of information. Other people may do none of these things. Yet a dominant feature of nearly all existing in-person stated preference interviews is that respondents are given the scenario and are asked to quickly make their decisions during a single interview. [This is not true of mail surveys, of course, although most SP researchers would probably agree that in-person interviews remain the gold standard for high-quality SP studies because of the ability

---

[1] There is also evidence that stated preference studies do in fact predict actual behavior with a reasonable degree of accuracy. Griffin et al. (1995) found that a binary choice stated preference survey predicted well whether respondents in Kerala, India would connect to an improved water supply network that in fact become available a short time after the survey. In one stated behavior study (though not a SP study), Whitehead (2005) find that stated decisions in 1998 to evacuate from hypothetical hurricanes of varying severities in coastal North Carolina were a reasonably good predictor of actual evacuation decisions from real hurricanes in 1999.

of a well-trained interviewer to control the information and flow of the interview]. Beginning in the 1990's, a handful of split-sample studies gave some respondents overnight to think about the scenario before reaching a decision. Most studies found clear evidence that respondents who had a chance to deliberate were less likely to say yes to the hypothetical scenario, lowering average willingness-to-pay in the sample population. Some respondents who were initially given no time to think were then given the opportunity to revise their answers overnight. In many cases, they revised their answers downwards.

Despite this, very few SP studies have incorporated a time-to-think. Researchers may be wary of the logistical and financial difficulties of implementing TTT (because it requires two interviews). Some may feel that giving time to think does not in fact replicate real-world situations since many or most voters (in the public good context) may pay little attention to an issue until they enter the voting booth. Another valid theoretical objection is that giving a respondent time to think introduces the possibility of strategic behavior, especially if survey participants have opportunities to discuss their answers with each other.

We synthesize results from four parallel TTT studies conducted in India, Mozambique, Vietnam and Bangladesh between 2002 and 2006. The studies estimated private demand for improved, "next generation" cholera and typhoid vaccines that are unavailable in these countries. Although the research designs differed somewhat, all four studies used the same general split-sample approach to examine the effect of time to think on both single binary discrete choice (India, Bangladesh, Mozambique) and a sequence of multinomial choice (Vietnam) responses[2]. Although the full results from these studies have recently appeared in the literature (Cook et al. 2007, Lucas et al. 2007, Whittington et al. 2008 and Islam et al. 2008), these papers were not focused on the examination of the effect of TTT (with the exception of Cook et al. 2007). Furthermore, because of their public health and development focus, they may not have come to the attention of many SP practitioners, many of whom are environmental economists. We find consistent and robust results that a time-to-think treatment lowers average WTP on the order of 30-40%. This result is statistically-significant and robust to a correction which drops uncertain respondents. In fact, giving respondents' time to think increases the certainty they have in their answers.

The next section reviews the literature on giving respondents time to think, including related research on test-retest studies, group elicitation approaches, and the "drop-off" protocol. We then discuss seven hypotheses on the underlying mechanisms for why TTT may change responses. The fourth section

---

[2] There has been some confusion in the literature on the names for various stated preference approaches. We use Carson and Louviere's (2011) nomenclature. Carson and Louviere's phrase "multinomial discrete choice sequence" is used to describe the approach that has been called discrete choice experiments, choice models, applied choice analysis, choice experiments, stated choice experiments, and conjoint analysis.

describes the research design and the four sites where we gave respondents time to think. The fifth section describes our main time-to-think results. We conclude with the limitations of the study as well as reflections on the advantages and disadvantages of giving respondents time to think during stated preference surveys.

Literature

Since differences in preferences between respondents with and without time to think could reflect a lack of stability of preferences, we begin with a brief review of the large number of "test-retest" studies conducted since the mid-1980's. Test-retest studies measure the reliability of single or repeated binary choices by measuring willingness to pay at two or more points in time. The earliest of these studies were panels, administering exactly the same survey to the same respondents separated by a gap of 1 month (Jones-Lee et al. 1985) or 2 weeks (Kealy et al. 1988, Kealy et al. 1990) Both studies found no statistical evidence for differences in responses. Because of concerns that respondents simply anchored on their previous answers, subsequent studies lengthened the interval between interviews to several months or years. Loomis (1989, 1990) surveyed the same respondents about water quality improvements in Mono Lake 9 months after the initial survey and found that initial and follow-up WTP were significantly correlated. In a more recent example, Brouwer et al. (2008) found that 80 percent of Bangladeshi respondents who were contacted six months after the original in-person WTP survey (for flood risk reduction) said they would answer the WTP question in the same way. In addition to panel studies, four "test-retest" studies were conducted with independent samples from a single population at two time periods (Reiling et al. 1990, Teisl et al. 1995, Carson et al. 1997, Whitehead and Hoban 1999). All four studies found that WTP was stable over the time periods tested, ranging from several months to several years[3]. Many of the test-retest studies, as well as other early contingent valuation studies, were mail surveys, so that respondents could answer at their leisure and in effect had time to think if they wanted it. Among test-retest studies, only Carson et al. (1997), Jones-Lee et al. (1985), and Brouwer et al. (2008) have used in-person interviews.

A series of studies in the early 1990's first examined the effect of giving stated preference respondents more "time to think" about their answers to valuation questions. Whittington et al. (1992)

---

[3] One exception is Whitehead and Hoban (1999), who found that WTP for a generalized scenario of environmental protection was lower at during the follow-up survey five years after the initial survey. They also observed, however, that the percent of the underlying population with favorable views toward government as well as the percent who ranked environmental concerns highly had decreased over the same period. Controlling for these changes, preferences were temporally stable.

examined WTP for public water taps and private water connections in Anambra State (Nigeria). Half of respondents in the study heard the SP scenario during the first visit from interviewers but were given overnight to think about it before responding. The other half responded during the first interview; these respondents were subsequently given the opportunity to revise their answers overnight. Elicitation was a double-bounded (bidding game) question, analyzed using four different approaches. The researchers found that giving respondents time to think decreased WTP, program ratings and likelihood of connection for both public taps and private connections; the results were robust across the four modeling approaches. Average WTP fell by approximately 37% for public taps and 32% for private taps. Informal group interviews conducted after the interview uncovered no evidence that the extra time induced people to answer strategically. Similarly, Lauria et al. (1999) found that giving respondents time to think reduced bids for improved sanitation services in the Philippines. They found evidence of a starting point bias in the double-bounded referendum questions (higher starting bids increased WTP), but found that giving time to think erased that bias. In contrast, Whittington et al. (1993) did not find that time to think had a robust effect on WTP for sanitation services in Ghana. Time to think only reduced WTP for sewer connections within the subset of respondents with water connections.

There has been little additional empirical investigation on the subject since the 1990's. Svedsater (2007) examined hypothetical donations to an environmental program among 111 students in London. He used a multiple-bounded discrete choice format with explicit levels of uncertainty (i.e. "I am 90% sure I would pay"; Welsh and Poe 1998), and gave one third of students one week to consider their responses. Svedsater found that TTT reduced respondents' uncertainty and lowered WTP. No split sample time-to-think studies have been conducted in the general population of an industrialized country. However, there have been two parallel lines of research that echo some of the same findings.

The first is the use of groups to elicit WTP. Alternately called the "market stall" (Macmillan et al. 2002, Lienhoop and Macmillan 2002) or "citizen jury" approach (Alvarez-Farizo and Hanley 2006), these surveys are typically split sample experiments where half of respondents complete a valuation exercise in one interview. The other half of respondents meet in groups to discuss the scenarios and ask clarifying questions of a moderator. At the end of the hour-long session they answer the valuation exercise individually and confidentially. They then attend another group session one week later after they have had a chance to discuss with family, seek out more information, consider budget constraints, etc., and answer the same valuation question(s). MacMillan et al. (2002) found that respondents who were interviewed individually without time to think had mean WTP 2 - 4 times higher than those who participated in this "market stall" group approach. Over one-third of group respondents changed their answers during the intervening week, and group participants were *less* certain of their responses than those interviewed individually.

The second line of research is the use of a “drop-off” protocol (Subade 2007, Labao et al. 2008, Nabangchang 2008). In a study valuing reef protection in the Philippines, Subade (2007) randomly assigned half of respondents to complete a standard in-person CV interview. Interviewers made personal contact with the other half of respondents, but left the survey with the respondent to complete. They made an appointment to retrieve the completed questionnaire and answer any questions. Subade found that WTP among the sample who completed the drop-off protocol was approximately 50% lower than WTP estimates from the sample respondents who completed the standard in-person interview. One drawback of this method is that, like a mail survey, it requires that respondents be literate, which is a problem in some populations in developing countries. Like mail surveys, the drop-off protocol also does not carefully control the flow of the information in the questionnaire, allowing respondents to jump ahead in the survey instrument in unobserved ways. Nevertheless, the design shares a common feature of TTT designs in that respondents have time to consider their responses and discuss with family.

Why might time to think change responses?

In this section we briefly explore seven hypotheses on the underlying mechanisms for why time to think (TTT) may change responses to valuation questions. They are not mutually-exclusive. Our intent is not to develop a formal model of how time to think enters into individuals’ decision-making processes, but rather to provide a structure for considering what we observe in our four study sites and to inform future research on the role and mechanism of the time-to-think effect.

First, TTT may attenuate enumerator bias, or the tendency for respondents to give enumerators the answers that they think enumerators want to hear. Good SP surveys, of course, have carefully-crafted language ensuring respondents that there are no right or wrong answers and providing socially-acceptable reasons to say no. They are also conducted by enumerators who are carefully trained to be neutral. Still, respondents may feel pressured to give the “right” answer, and giving TTT may allow respondents to reach their decision overnight (without the enumerator present). In our studies respondents still had to tell the enumerator their decision the next day, though it would be possible to pair TTT with a “ballot box” approach (for literate respondents, this would be very similar to Subade’s “drop-off” protocol).

Second, TTT may attenuate a salience bias. With in-person surveys, someone comes to the respondent’s home and spends 15-30 minutes describing a good which the respondent probably has not spent much time considering before. Despite a warning from the enumerator to consider other things that the respondent may spend their money on, they may not adequately consider their budget constraint. This phenomenon would be familiar to most used-car salesmen.

Third, the TTT approach may be eliciting group rather than individual preferences. Respondents are allowed, even encouraged, in our studies to consult with their household members and friends. Carlsson (2010) point outs that it is not clear in these circumstances whose preferences are being recorded. As he notes, this is may be desirable if household-level decisions are being studied (as is the case in our vaccine studies), and in fact may be more representative of how individuals make decisions are made in reality. Time to think may improve predictive validity.

Fourth, giving respondents the ability to confer with friends and neighbors could induce strategic behavior, compromising the incentive compatibility of the elicitation approach. SP surveys for currently unavailable private goods, like our vaccines, already have incentive compatibility problems as respondents may overstate their true WTP in order to ensure that the good is actually produced and made available in the market, even if they have little intention of actually buying it (Carson and Groves 2007). Giving respondents TTT may allow them to think strategically in this manner. Even for local public goods (like an improved piped water distribution system that a household need not connect to), TTT may increase the probability that survey respondents could confer, compare the referendum prices they were offered, and develop a strategy for answering the valuation question(s) that is not reflective of their true WTP. As noted above, though, Whittington et al. (1992) found little evidence of this from focus groups.

Fifth, TTT may relax an "attention budget" constraint. Cameron and DeShazo (2010) develop a theoretical model for stated preference and revealed preference data that acknowledges that respondents allocate scarce cognitive resources and time to understanding attributes and completing choice tasks in a rational way. They argue that economists may incorrectly infer a low marginal utility for an attribute that a respondent cares about but chooses to pay little attention to because of time constraints or because of small variations in that attribute's levels. Giving respondents time to think about multinomial discrete choice tasks may relax this attention or time constraint, leading to a different pattern of attention allocation across tasks than would be the case in the single-interview format.

Sixth, TTT may allow the respondent to learn more about an unfamiliar good by acquiring information that was not provided in the survey (e.g.,. from the Internet, friends, etc.). This obviously confounds the stated preference experiment. At a minimum, researchers can ask about, and control for, whether respondents sought additional information. For predictive validity of the archetypal referendum vote on a public good, however, one would expect some subset of voters to seek out additional information beyond that provided in a voter's ballot.

Seventh, for unfamiliar goods, TTT may help respondents "discover" their preferences. Thus, TTT may increase respondent's certainty in their answers. A number of studies (Welsh and Poe 1998, Blumenschein et al. 2008) have found that dropping respondents who say they are not certain eliminates much of the gap between real and hypothetical responses. We explore this "uncertainty correction" in our data below. Another approach that has not yet been tied to stated preference experiments might be to investigate respondents' subjective expectations with respect to different aspects of the good being valued (Delavande et al. 2010). For example, in the case of vaccines, one might explore whether there is a link between the distribution of subjective expectations about disease risk and WTP, before and after time-to-think.

Study Sites and Research Design

As part of the Diseases of the Most Impoverished (DOMI) program, researchers from the University of North Carolina at Chapel Hill and host-country partners conducted a series of studies on private demand for cholera and typhoid vaccines in six countries from 2002 to 2006.[4] Donors and policymakers were interested in understanding household demand for "next generation" vaccines which had been the target of scientific research to improve their efficacy and lower their cost. Although some respondents had experience with older, less effective versions of these vaccines, the improved "next generation" cholera or typhoid vaccines were generally unavailable in these countries except in limited trials designed to evaluate the vaccines' protective effectiveness, during which they were distributed free of charge (Jeuland et al. 2009c). Because of this unavailability of "next generation" vaccines, we relied on stated preference techniques to measure household demand (Canh et al. 2006, Lucas et al. 2007, Islam et al. 2008, Kim et al. 2008, Whittington et al. 2008). This information was then used to evaluate the economic attractiveness of investments in these vaccines using both benefit-cost and cost-effectiveness techniques (Cook et al. 2008, Cook et al. 2009b, Jeuland et al. 2009a, Jeuland et al. 2009b). It was also used to examine cross-subsidy schemes where sales to adults cross-subsidize free vaccines for children (Lauria et al. 2009) and optimal vaccine subsidies in the presence of a herd immunity externality (Cook et al. 2009a). We focus here only on the subset of these studies that provided stated preference respondents time to think about their answers. These were conducted in Hue (Vietnam), Kolkata (India), Beira (Mozambique) and Matlab (Bangladesh).

---

[4] The DOMI program was administered by the International Vaccine Institute (IVI) with support from the Bill and Melinda Gates Foundation.

Although the study team tailored surveys to local conditions and attitudes after careful focus groups and pretesting, the structure and content of the questionnaires in all four countries were similar. We begin by describing the research design in Kolkata (India) in some detail, and then discuss how the design in other sites differed.

India

Stated preference surveys were conducted in 2004 in two areas of Kolkata (formerly Calcutta): a densely-crowded low income slum (Tiljala), and a neighborhood with more diverse incomes and living conditions (Beliaghata)[5]. The split-sample TTT experiment was conducted only in Beliaghata (n=559). We randomly selected households that had children under the age of 18 and interviewed either the mother or father of the children. We begin by describing the structure of the survey instrument for the no-time-to-think (NTTT) sample, where the entire interview was done in one session (Table 1). As in all the four countries, interviews were done in person with a team of well-trained local enumerators. The beginning of the survey elicited views on vaccines in general, including the households' vaccination histories. Enumerators then described either a hypothetical cholera or a hypothetical typhoid vaccine in detail. They then reminded respondents of their budget constraint and emphasized reasons for and against purchasing the vaccine[6]. Enumerators also handed respondents a card that read in Bengali:

---

[5] More details on the Kolkata field site, the BC design and results can be found in Whittington et al. 2008. The full bilingual questionnaire used in Kolkata is also available at http://faculty.washington.edu/jhcook/research.html. Similarly, details on the Bangladesh, Mozambique and Vietnam studies can be found in Islam et al. 2008, Lucas et al. 2007, and Cook et al. 2007.

[6] The script in Kolkata was: "Now I'd like to know whether you would buy the vaccine if it was available at a specified price. Some people say they cannot afford the price of the vaccine or that they are actually not at risk of getting this disease. Other people say that would buy the vaccine because the protection is really worth it to them. In other studies about vaccines, we have found that people sometimes say they want to buy the vaccine. They think: 'I would really like as much protection from this disease as possible.' However, they may forget about other things they need to spend their money on. Please try to think carefully about what you would actually do if you had to spend your own money. There are no right or wrong answers. We really want to know what you would do…. When you give your answer about whether you would or would not buy the vaccine, please consider the following: yours and your family's income and economic status compared with the price of the vaccine, and your risk of getting cholera. Apart from the vaccine, remember that we still have other ways to treat cholera such as oral rehydration solution. Also, remember that the benefit of the vaccine in preventing cholera is [50% effective for 3 years]. Again, the cholera vaccine cannot be used by children under 1 year and pregnant women."

> <u>"In thinking about whether you might want to purchase this vaccine, please keep in mind</u>:
> - Yours and your family's income and economic status
> - Your risk of getting cholera;
> - We still have other ways to treat cholera such as oral dehydration solution; and
> - The vaccine is 50% effective in preventing cholera for 3 years."

The card for typhoid vaccine had the same information except it substituted typhoid-specific treatment options in place of those for cholera, and used the typhoid vaccine's effectiveness (70%).

Respondents were then asked to respond to a single binary choice (BC) question about whether they would buy the vaccine for *themselves only* (what we refer to below as "respondent demand"). They were randomly assigned one of four vaccine prices[7]. Enumerators asked respondents for the reasons for their answers and asked how certain they were of their answers (very certain, somewhat certain, not certain/unsure, or no response). Respondents who said no to the initial vaccine price were also asked if they would take the vaccine if free, and if so, whether they would pay anything for it, though we did not ask what that amount would be. Respondents were then asked how many of the vaccines they would buy for their household members at the randomly-assigned vaccine price (what we call "household demand"), and for which household members they would buy a vaccine. We label this count elicitation mechanism "CT" below.

All respondents who either said they would purchase the vaccine for themselves at the assigned price or that they would pay something for the vaccine completed a payment card (PC) exercise. The exercise used the analogy of a traffic light where "green" prices were those that a respondent was completely sure she would pay, "red" prices were those that a respondent was completely sure she would not pay, and "yellow" prices were those over which the respondent was uncertain (see Figure 1). Respondents were asked about fifteen prices[8]. In summary, we elicited respondent demand using binary choice (BC) and payment card (PC) approaches, and demand for vaccines for other household members using a count (CT) approach.

"Time-to-think" respondents were presented with the same information about the hypothetical vaccine but were not immediately given the price. Instead, enumerators said the following:

> "We are almost at the end of our first interview, and I want to thank you very much for your time. I would like to return again tomorrow to ask you more questions. I will ask you whether you would want to buy this vaccine for yourself as well as for other members of your household if it were sold at a certain price. I would encourage you to think overnight about how much this new vaccine is worth to you, and the range of prices

---

[7] These were Indian Rs. 10, 25, 50 and 500. At the time of the studyUS1$ = Rs. 45.

[8] These were Indian Rs. 0, 1, 5, 10,15, 25, 50, 75, 100, 200, 300, 400, 500, 1000, 5000.

> you might be willing to pay for this vaccine for yourself and for your household members. You may also want to discuss these decisions with other members of your household."

The card shown to NTTT respondents (see above) was given to respondents as a "reminder card" to examine overnight. During the second interview, the interviewer asked the exact same sequence of stated preference questions as for NTTT respondents. The interviewer also asked several time-to-think debriefing questions, including how long respondents spent thinking about the decision, who they discussed the decision with, and other sources of information they consulted.

Bangladesh

Fieldwork was conducted in Matlab, a rural area southeast of Dhaka, in the summer of 2005. The survey focused only on cholera vaccines. The time-to-think research design is nearly identical to the one used in Kolkata. Because the PC exercise was completed for the TTT subsample only, we do not analyze that data here. The final sample size was 591 respondents split among six vaccine prices, though nine respondents rejected the vaccine scenario and were dropped from the analysis. We report results for respondents who lived within the service delivery area of the International Centre for Diarrhoeal Disease Research (ICDDR-B) separately from those who live in the area served by the government. As Islam et al. (2008) note, free access to the ICDDR-B's nationally-renowned diarrheal disease treatment center may have lowered stated demand for cholera vaccines.

Mozambique

This study, conducted in the coastal city of Beira in the summer of 2005, also focused on cholera vaccines. The study design was somewhat different than Kolkata or Matlab. First, respondent demand was elicited only with the PC exercise. Household demand was elicited, as in the other two sites, with a count (CT) question with five randomly-assigned prices. There were four subsamples on time-to-think. The first subsample was given no time to think, and completed the PC exercise for respondent demand before answering the CT household demand question. The second subsample was also given no time to think, but were asked the CT question before the PC exercise. The third subsample completed the PC exercise in the first interview, were shown the single randomly-chosen price for household demand, and then given overnight to think about the CT question. The fourth subsample were given the single randomly-chosen price for household demand and then given overnight to think. During the second interview they answered the CT question and completed the PC exercise. Finally, the starting point on the PC exercise (i.e. whether one starts at a high price first and moves down, or starts at the lowest price first and moves up) was randomized. A figure summarizing the research design in Beira is provided in the supplementary appendix (see Figure A1 in the appendix). The final sample size was 991 households.

Vietnam

Fieldwork was conducted in Hue in the summer of 2003[9]. Respondents were asked to complete a series of six multinomial discrete choice (MDC) tasks about purchasing a vaccine for themselves. Each task required the respondent to compare a hypothetical cholera vaccine and a hypothetical typhoid vaccine. The attributes of the vaccines were its effectiveness, duration, and price. Respondents could choose a cholera vaccine, a typhoid vaccine or neither. The tasks were designed using a standard fractional factorial research design and blocks were randomly assigned to participants. Half of these MDC respondents were given the set of six choice tasks (in the form of laminated cards) to examine and mark overnight (Cook et al. 2007). The other half completed the cards in the first interview, but were given the opportunity to revise their answers overnight if they wished, following the approach of Whittington et al. (1992) in Nigeria. The final sample size was 400 households.

Table 2 provides summary statistics for key socioeconomic characteristics by TTT treatment in each study site.

Results

Analysis of respondent demand

We begin with the raw results from the binomial choice (BC), payment card (PC), and multinomial discrete choice (MDC) exercises for respondent demand. In the Kolkata and Matlab BC task, respondents were less likely to say they would buy a cholera or typhoid vaccine for themselves if they were given time to think (Table 3). In all but four cases, the percentage of NTTT respondents who said yes is higher than the percentage of TTT respondents who said yes. Time to think reduced the percentage of respondents who said yes by 3 – 25%. Using a two-tailed Fisher exact test, however, the

[9] We also used a BC approach to elicit WTP for typhoid vaccines in 2002 (Canh et al. 2006) and cholera vaccines in 2003 (Kim et al. 2008). These interviews were all conducted during one interview (i.e. no time to think) and are thus not discussed further in this paper.

differences in percent saying yes is statistically significant at the 10% level for only two of the comparisons.

We use information from the payment card (PC) exercise to define lower and upper bounds on WTP, where the lower bound is the highest price that respondents were certain they would pay (the highest "green" price in the stoplight analogy), and the upper bound is the lowest price they were certain they would not pay (lowest "red" price). Figure 2 shows the percentage of Kolkata respondents who said they were completely certain they would buy the vaccine over a range of prices. Giving time to think clearly reduced this percentage for typhoid vaccines, though somewhat less so for cholera vaccines. For typhoid vaccines in Kolkata, the lower bound, upper bound, and the midpoint of the interval were all lower among TTT respondents. The midpoint was 40% lower, and all differences were statistically significant using two-sample t-tests (see Table A6 in the appendix). Although the direction of the effect is the same for cholera vaccines in Kolkata, none of the differences are statistically significant. Using a nonparametric Kolmogorov-Smirnov test of whether the distribution of values among NTTT respondents is different than the distribution of values among TTT respondents, we reject the null of no difference at the 10% level for both the lower and upper bound when the cholera and typhoid data are pooled[10]. The size of the uncertainty interval, or the interval between the upper bound and lower bound, is also smaller for TTT respondents, a result to which we return later in the paper. In Beira, the midpoint, lower bound and upper bound were all lower among TTT respondents who did the PC exercise after the CT question on household demand (Appendix Table A6). The midpoint is 30% lower. For respondents who did the PC exercise during the first interview, however, only the differences in midpoint and upper bound are statistically significant, and the midpoint is 12% lower (and weakly statistically significant). A Kolmogorov-Smirnov test on pooled Beira data rejects the null hypothesis of no difference of both the lower and upper bounds at the 1% level[11].

Finally, Figure 3 presents raw MDC responses in Vietnam. Because we used exactly the same choice tasks in the NTTT and TTT subsamples, we are able to directly compare raw responses. Each data point plotted in the graph reflects the percentage of respondents who said they would buy neither vaccine at the offered price in the task. The solid line denotes responses where the percentage of respondents who

---

[10] The p-values for the pooled dataset are p=0.058 for a difference in the distribution of lower bounds, and p=0.03 for a difference in the distribution of lower bounds. For the cholera subsample only, p=0.107 (lower bounds) and p=0.894 (upper bounds). For the typhoid subsample only, p=0.183 (lower bounds) and p=0.017 (upper bounds).

[11] For the pooled dataset, p=0.000 (lower bounds) and p=0.000 (upper bounds). Among those who did the PC exercise before the CT, p=0.437 (lower bounds) and p=0.045 (upper bounds). Among those who did the PC exercise after the CT exercise, p=0.000 (lower bounds) and p=0.000 (upper bounds).

said "neither" on a given card is exactly the same. Points lying above this line indicate that a higher percentage of TTT respondents bought "neither" vaccine than NTTT respondents. A higher percentage of TTT respondents opted for "neither vaccine" in 16 of 18 choice tasks.

Although exact multivariate modeling specifications differed somewhat between field sites (see Cook et al. 2007, Lucas et al. 2007, Islam et al. 2008, Whittington et al. 2008 for more details), our approaches were similar. We analyzed respondent demand from the BC exercise (Kolkata and Matlab) using a probit model, and analyzed PC data using both OLS and an interval regression model (Cameron and Huppert 1989). We used a random-parameters logit model (RPL) for the MDC data in Hue. Our multivariate results confirm the pattern seen in the raw data (the full results are available in Table A1-A5 in the appendix). After controlling for a number of covariates for vaccine demand, a dummy variable for whether a respondent was given time to think was negative and statistically significant in all models, including probit, OLS, interval regression, and RPL models of respondent demand.

We find that mean respondent willingness-to-pay is consistently lower in the TTT subsamples (Table 4). Estimates from the nonparametric Kristrom midpoint model as well as parametric model estimates (probit, interval, OLS and RPL) consistently show these reductions. The smallest difference in WTP is for cholera respondents in Kolkata using the Kristrom estimator of mean WTP (12% less). This is largely driven by the fact that the percentage of cholera respondents who said they would buy a vaccine for themselves was very similar at the highest price in the two subsamples (Table 3). The largest difference in WTP is for demand for cholera vaccines in Hue from the MDC exercise. [In fact, the RPL model predicts a negative mean WTP for a cholera vaccine that is 50% effective for 3 years, leading to a very large difference between the NTTT and TTT subsamples. This negative WTP estimate is driven by a large coefficient on vaccine effectiveness in the model; respondents were shown vaccines with 50, 70, and 99% effectiveness and valued increasing effectiveness highly. We present in Table 4 the WTP for a 50% effective vaccine because that is the best estimate of the effectiveness of the real next-generation cholera vaccine.] A simple average of the percent reduction in all respondent WTP estimates in Table 4, not weighted by sample size, is 41%. Omitting the 105% reduction from Hue, the average reduction in WTP is 36%.

The differences are not, however, all statistically significant. We construct CIs for the Kristrom WTP around standard errors calculated using the method in Vaughn and Rodriguez (2001). None of the differences in Kristom WTP are statistically significant (90% CIs overlap). We use a simulation approach to calculate CIs around mean WTP calculated from the probit model[12]; 90% intervals not overlap in Kolkata and in Matlab's government service area, though they do (barely) for the ICDDRB study area[13]. Difference in mean WTP is, however, statistically significant for PC responses in both Kolkata and Beira. In addition to differences in predicted WTP from the interval regression models[14], a two-sided t-test shows a statistically significant difference at the 10% level in the midpoint of the uncertainty interval from the PC exercise for cholera in Beira and typhoid in Kolkata, although not for cholera in Kolkata (see Table A6 in the appendix) . WTP distributions from the RPL models in Hue were not statistically different[15].

Analysis of household demand

Like the results for respondent demand, our raw response count (CT) data consistently show that demand for vaccines for household members is lower when respondents are given time to think. In the three sites where we conducted a split-sample TTT experiment with household demand, the percentage of all household members vaccinated (including the respondent) is lower for TTT respondents at nearly every price level (Table 5). Using a two-tailed t-test, we find that these differences are statistically significant at the 10% level or better in 11 of 21 comparisons. Using a Kolmogorov-Smirnov test and pooling across prices, we find that the distribution of the percentage of household members vaccinated is statistically different in all sites except for cholera in Kolkata[16].

---

[12] These distributions were calculated in Stata by drawing 10,000 sets of coefficient estimates from a multivariate normal distribution with means set to the vector of parameter coefficients estimated by the probit model and the covariance structure set to the model's estimated variance-covariance matrix. WTP was then calculated for each draw at the sample means of independent variables. The empirical distribution of those 10,000 WTP estimates provides the 90% confidence intervals shown. A similar approach was used for confidence intervals of household WTP from the negative binomial model.

[13] Because our simulation approach (for probit and negative binomial) draws the WTP estimates for NTTT and TTT from the same underlying model, the covariance of the two measures is likely nonzero, so non-overlapping confidence intervals are not equivalent to a statistically-significant difference.

[14] The prediction from the interval model is the estimate of mean WTP, so here we simply predict WTP for the NTTT and TTT subsamples separately, constructing confidence intervals using Stata's post-estimate "ci" command. See Table

[15] Because we used a mixed logit-hierarchical Bayes approach, we were able to estimate an individual-level WTP estimate for each of the 200 NTTT respondents and 200 TTT respondents. The confidence intervals in Table 4 are constructed around empirical standard errors among each of the subsamples (i.e standard deviation / sqrt(200)). We tested statistical significance with a two-tailed t-test comparing the 200 NTTT and 200 TTT WTP estimates.

[16] See Table A10 in the appendix for more detailed test results. The Kolmogorov-Smirnov p-value comparing NTTT with TTT in Matlab gov't service area is p=0.002; Matlab ICDDR-B area p=0.08; Beira p=0.000; Kolkata Cholera p=0.15; Kolkata typhoid p=0.02.

We modeled the CT data in Kolkata, Beira and Matlab with a negative binomial model, a variant of a Poisson count model appropriate for situations where data are over-dispersed (conditional variance is larger than conditional mean). The full multivariate results are shown in Tables A1, A2 and A4 in the appendix. A dummy for giving respondents time to think was again statistically significant in all models. The effect was also economically significant: giving respondents time to think lowered the average number of vaccines a household said it would purchase by 0.6 vaccines at the mean price and 1.1 vaccines at the lowest price in Beira (average marginal effects with other covariates set to their sample means). The average number of household vaccines fell by 1.3 vaccines at US$0.50 and 1.1 vaccines at US$1.00 in Matlab. Similarly, we find total household WTP is consistently lower among those given time to think (Table 4). A simple average of the percent reduction in all household WTP estimates in Table 4, not weighted by sample size, is 28%. We use a simulation approach (identical to the one described above for the probit model) to calculate confidence intervals around mean WTP calculated from the negative binomial model. 90% confidence intervals do not overlap in Beira and Matlab, though they do in Kolkata.

Interactions of time to think with bid prices, income and education

To test whether time to think had a differential effect on respondents who were randomly assigned higher bid prices, we interacted the offered hypothetical price with time to think and re-ran the probit and negative binomial models. The interaction term was negative and significant in models of Kolkata and Beira household demand, but was not statistically significant for Kolkata respondent demand, Matlab respondent demand and Matlab household demand. We also interacted time to think with income, both continuous income and income quartile dummies. These interaction terms were only significant for household demand in Beira: an interaction with the highest income quartile and time to think was positive (0.69) and statistically-significant at the 5% level. Finally, we interacted time to think with dummy variables for education levels in respondent and household models. In Matlab, the interactions were positive and statistically significant for having either an education level of 1-9 years or having 12 years or vocational training. This would suggest that, compared to respondents with no formal schooling who were also given TTT, respondents with intermediate levels of education said they would purchase more vaccines for household members if given time to think. They were not statistically-significant in the other models.

Certainty and Preference Errors

We find evidence that providing time to think increases the certainty that respondents have in their responses. In Kolkata, 86% of TTT respondents felt "very certain" about their answer about buying a cholera vaccine for themselves, compared to 68% of NTTT respondents. This difference was statistically significant at the 1% level (paired t-test). There was, however, no statistical difference in certainty among respondents who were asked about a typhoid vaccine. In Beira, 75% of the "uncertain" respondents were in the NTTT subsample. Giving TTT also reduced the length of the uncertainty interval from the PC exercise for the respondents in Beira and the cholera typhoid respondents in Kolkata (see Table A3 and A6 in the appendix). We re-ran all the multivariate models of respondent demand (probit) and household demand (negative binomial) dropping a) all respondents who reported being "uncertain" and b) all respondents who reported being "uncertain" or "somewhat uncertain" [17]. In all cases, a dummy variable for TTT was still statistically significant and negative. In most cases, the magnitude of the coefficient actually increased when we dropped uncertain respondents (see Table A7 in appendix).

We also found from the MDC data in Hue that giving TTT reduced the number of respondents who made preference errors (violations of stability, monotonicity, or transitivity). Of the 200 respondents in NTTT subsample, 22 (11%) made some type of preference errors. Fourteen of 200 (7%) respondents in the TTT subsample made a preference error, a statistically significant difference (one-tailed t-test, p=0.05). Furthermore, when NTTT respondents were given the opportunity to revise their answers overnight, their revised responses showed fewer preference errors and were no longer statistically different than the TTT responses. A probit model confirmed that giving TTT reduced the probability of making a preference error (at the 10% level). When TTT was interacted with education, the model showed that TTT reduced preference errors in respondents with secondary school education by 35% but only 6% for respondents with a primary school education. All TTT variables and interaction terms were highly significant in this interaction model.

Debriefing questions: time spent thinking and discussion with others

Most respondents who were given the opportunity to think about the valuation question did so. In Vietnam, Kolkata, Matlab and Beira, only 2%, 11%, 5%, and 5% of TTT respondents respectively

[17] In Beira, respondents were not asked how certain they were about the number of vaccines they wanted to purchase for other household members, but rather how confident they were that they could afford that number. This is a different question, so we did not test for the effect of dropping uncertain respondents on Beira household demand.

reported spending no time on the task. Figure 4 shows the distribution of time spent thinking across all four sites, with a significant fraction reporting either 30 minutes or one hour. The average time spent across all four sites was 37 minutes, with a median of 20 minutes (Table 6).

A strong majority in all four sites reported discussing the decision with their spouse (Table 6); across all four sites, 69% of respondents consulted with their spouse. Fewer consulted with someone outside the household (23% across all sites), and, in Kolkata and Hue, even fewer reported using other information (we did not ask questions about use of other information in Beira and Matlab).

We used these debriefing questions to explore which elements of the time-to-think protocol might be influencing responses. First, we replaced the continuous variable of minutes spent thinking (imputing a zero for all NTTT respondents) for the time-to-think dummy variable in the multivariate models of respondent and household demand. We did not find a statistically significant relationship between minutes spent on the task and responses to the BC question (probit models). We did find that more minutes spent thinking lowered the number of vaccines purchased for the household in Kolkata and Matlab (though not in Beira) and lowered the midpoint of the uncertainty interval in OLS models of respondent demand from the PC exercise in Kolkata. We also explored whether, among only TTT respondents, those who spent longer on the task reported lower respondent or household demand. We did not find any statistically significant relationships.

Next, we added a variable tracking whether the respondent had discussed the question with anyone inside or outside the household along with the TTT dummy[18]. Since such a high percentage of TTT respondents reported discussing the decision (and because NTTT respondents did not by definition have this opportunity to discuss), this analysis does not have much statistical power. The coefficient on discussion was generally positive while the TTT dummy variable became more negative, suggesting that discussing the decision with someone actually increased respondent and household demand and counter-balanced other elements of TTT which reduced demand. None of the discussion coefficients, however, were statistically significant. A parallel approach examined responses among only the TTT subsample. Here the coefficient on discussion was positive in seven of nine models, and statistically significant at the 10% level in two models (Kolkata probit and OLS). Finally, we looked at the effect of TTT among only those respondents who did <u>not</u> discuss the decision (dropping the TTT respondents who said they conferred). The coefficient on the TTT dummy remained negative, statistically significant, and generally

[18] We also replaced the TTT dummy with this discussion dummy, and found consistent effects on the negative binomial models in all sites, but mixed significance for the respondent data (see "Model B" in Table A8).

had a higher magnitude than the models including those TTT respondents who did discuss the decision. Interested readers can find the results on discussion in Table A8 of the appendix.

Discussion

In summary, we find that giving respondents overnight to think about their responses has a large and consistent effect across all four study sites. This was true for both the BC and MDC methods. Time-to-think lowered the percentage of respondents who said they would purchase a cholera or typhoid vaccine for themselves across a range of bid prices, and lowered the percentage of household members for whom the respondent said they would purchase a vaccine. It decreased average response WTP by 41% and household WTP by 28%, though the differences were not all statistically significant. It reduced the likelihood that a MDC respondent made a preference error. We find that, given the opportunity, over half of respondents consulted with their spouse about their vaccine purchase decision, and spent 20 – 30 minutes thinking about the decision. We also find suggestive evidence that consultation with others actually increased demand for vaccines, counteracting other components of TTT which reduced demand.

In most cases, time-to-think also increased the certainty that respondents had in their answers. Our result stands in contrast to MacMillan *et al.* (2002), who find that respondents who were given time and who completed a group elicitation approach were less certain than respondents who completed the survey in one session individually. Group interactions, however, could have been the driver of increased uncertainty in MacMillan et al.'s (2002) results, rather than the effect of time to think. Our finding is similar, however, to Svedsater (2007), who found that TTT reduced respondents' uncertainty. Comparing real versus hypothetical (stated) purchases of a diabetes management program, Blumenschein et al. (2008) find that dropping respondents who said they were less than "absolutely sure" from the analysis reduced nearly all of the gap between real and hypothetical choice (a standard cheap talk script had no significant effect). We find that our TTT treatment effect remains, however, even after we similarly drop respondents who are "uncertain" or "somewhat certain" of their answers.

One limitation of our study is that although we speculate the reduction in WTP is a reduction in "hypothetical bias", we do not test for hypothetical bias directly. We only offered respondents hypothetical vaccines and were unable to give a comparison group the opportunity to buy the real vaccine. A more direct test of the effect of time-to-think in a real vs. hypothetical laboratory experiment would be useful. A second limitation, noted above, is that stated preference surveys for private goods may have higher levels of hypothetical bias than those for public goods. Respondents may say yes in order to increase the probability of the hypothetical product being provided in the marketplace. It is possible that TTT has the effect of reducing this potential inherent bias in using SP to elicit demand for hypothetical private goods, and may not be more generally applicable to public good contexts.

Why have so few TTT studies been conducted? For many researchers who are aware of the TTT protocol, the first response to this question may be cost. In-person stated preference surveys are already expensive endeavors. Most of this expense derives from the fact that skilled interviewers must be employed and respondents compensated for their time in completing the interview. Although there is a substantial fixed cost to developing the survey instrument and training enumerators, adding a second follow-up interview will certainly increase survey expenses. It is also probably no coincidence that all existing time to think studies have been conducted in developing countries where the cost of implementing a high-quality stated preference survey is much lower because both interviewer salaries and any household compensation are much lower than in industrialized countries. Here is another way of viewing this dilemma: for the same total cost, a researcher might implement a TTT study but be forced to reduce the sample size. The researcher might then be concerned with the decrease in statistical precision surrounding estimation of WTP. This should be weighed against the risk (for policy analysis) that welfare estimates from conventional single-visit survey methodologies might be 30-40% too high.

Conducting a survey with a time-to-think protocol is also logistically complicated. As described earlier, the study in Kolkata interviewed households in one of the poorest slums in Kolkata called Tiljala. Because of safety concerns (our interviewers were not comfortable interviewing households after dark, and there was a large riot during our fieldwork), we decided to finish the survey work quickly and did not give respondents time to think. To avoid the potential for strategic bias and for households to receive prior (and unobserved) information about the survey from their neighbors, we also completed the sites where respondents were given no time to think first and then completed interviews with the TTT subsample. This sort of logistical problem, however, exists for all split-sample studies, and might be less of a concern for studies in which all respondents are given time to think (though researchers may still want to ask households if they have heard about the study and what information they have already learned).

Another reason why time-to-think designs are still rare may be the fact that in-person interviews for valuation work in industrialized countries are now the exception rather than the norm. The need for such a design decreases considerably for more commonly used mail surveys, which allow respondents to reflect longer on their choices. For other survey modes in which response rates and selection problems are a major concern, such as telephone surveys, it may be even more difficult to contact respondents multiple times. Researchers would have difficulty in estimating population WTP if a sizeable number of respondents completed only the first half of the interview. We did not find this type of sample selection problem in our four TTT studies[19]. Representative Internet panels are also increasingly being used in developed countries. They have the flexibility to allow respondents time to reflect, though we are unaware of any internet-mode studies that have examined the role of reflection specifically.

Acknowledgements

We would like to thank the International Vaccine Institute and its director John Clemens, along with a number of the DOMI project staff, including Jacqueline Deen, Lorenz von Seidlein, R. Leon Ochiai, and Andrew Nyamete. We are also thankful for the assistance of our partners in each of the four countries, including: India's NICED (Dr. Susmita Chatterjee, Dr. S.K. Bhattacharya, and Dr. Dipika Sur), Vietnam's NIHE (Dr. Do Gia Cahn), Mozambique's Center for Environmental Hygiene and Medical Examination (Dr. Marcelino Lucas, Dr. Avertino Barreto, Dr. Francisco Songane, Arnaldo Cumbane, Arminda Macamuale, and Nivalda Nazaro) and Bangladesh's ICDDR-B (Dr. Ziaul Islam). All four country studies were funded with support from the Bill and Melinda Gates Foundation. We thank the Associate Editor Christian Vossler for very helpful and detailed suggestions that have demonstrably improved the paper, as well as Sonia Akter and two anonymous reviewers for thoughtful comments and suggestions.

[19] In Kolkata, Matlab and Vietnam all TTT respondents completed the second interview at least through the main valuation questions (and nearly all completed the whole survey). In Beira, all TTT respondents began a second interview, though one did not complete the main valuation questions.

References

Alvarez-Farizo, B. and N. Hanley (2006). "Improving the process of valuing non-market benefits: combining citizens' juries with choice modeling." manuscript.

Blumenschein, K., G. C. Blomquist, M. Johannesson, N. Horn and P. Freeman (2008). "Eliciting Willingness to Pay Without Bias: Evidence from a Field Experiment*." The Economic Journal **118**(525): 114-137.

Brouwer, R., S. Akter, L. Brander and E. Haque (2008). "Economic valuation of flood risk exposure and reduction in a severely flood prone developing country." Environment and Development Economics **14**: 397-417.

Cameron, T. A. and J. R. DeShazo (2010). Differential attention to attributes in utility-theoretic choice models. Manuscript version Feb 7 2010.

Cameron, T. A. and D. D. Huppert (1989). "OLS versus ML estimation of non-market resource values with payment card interval data." Journal of Environmental Economics and Management **17**(3): 230-246.

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

Carlsson, F. (2010). "Design of Stated Preference Surveys: Is There More to Learn from Behavioral Economics?" Environmental and Resource Economics **46**(2): 167-177.

Carson, R. and T. Groves (2007). "Incentive and informational properties of preference questions." Environmental and Resource Economics **37**(1): 1573-1502.

Carson, R. T., W. M. Hanemann, R. J. Kopp, J. A. Krosnick, R. C. Mitchell, S. Presser, P. A. Ruud, V. K. Smith, M. Conaway and K. Martin (1997). "Temporal reliability of estimates from contingent valuation." Land Economics **73**(2): 151-163.

Carson, R. T. and J. J. Louviere (2011). "A common nomenclature for stated preference elicitation approaches." Environmental and Resource Economics **forthcoming**.

Cook, J., M. Jeuland, B. Maskery, D. Lauria, D. Sur, J. Clemens and D. Whittington (2009a). "Re-visiting socially-optimal vaccine subsidies: an empirical application in Kolkata, India." Journal of Policy Analysis and Management **28**(1): 6-28.

Cook, J., M. Jeuland, D. Whittington, C. Poulos, J. Clemens, D. Sur, D. D. Anh, M. Agtini, Z. Bhutta and D. T. E. S. Group (2008). "The cost-effectiveness of typhoid Vi vaccination programs: Calculations for four urban sites in four Asian countries." Vaccine **26**: 6305-6316.

Cook, J., D. Sur, J. Clemens and D. Whittington (2009b). "Evaluating investments in typhoid vaccines in two slums in Kolkata, India." Journal of Health, Population and Nutrition **27**(6): 711-724.

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

Delavande, A., X. Giné and D. McKenzie (2010). "Measuring subjective expectations in developing countries: A critical review and new evidence." Journal of Development Economics **94**(2): 151-163.

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.

Harrison, G. W. (2006). "Experimental evidence on alternative environmental valuation methods." Environmental and Resource Economics **34**: 125-162.

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

Jeuland, M., J. Clemens, J. Cook, D. Whittington and D. C. E. S. Group (2009a). "Incorporating herd protection into cost-effectiveness calculations of new-generation oral cholera vaccines: a multi-site analysis." Value in Health **12**(6): 899-908.

Jeuland, M., M. Lucas, J. Clemens and D. Whittington (2009b). "A cost-benefit analysis of cholera vaccination programs in Beira, Mozambique." World Bank Economic Review **23**(2): 235-267.

Jeuland, M., M. Lucas, J. Deen, N. Lazaro and D. Whittington (2009c). "Estimating the private benefits of vaccination against cholera in Beira, Mozambique: A travel cost approach." Journal of Development Economics **91**: 310-322.

Jones-Lee, M. W., M. Hammerton and P. R. Phillips (1985). "The value of safety: Results of a national sample survey." The Economic Journal **95**(377): 49-72.

Kealy, M. J., J. F. Dovidio and M. L. Rockel (1988). "Accuracy in valuation is a matter of degree." Land Economics **64**(2): 158-173.

Kealy, M. J., M. Montgomery and J. F. Dovidio (1990). "Reliability and predictive validity of contingent values: Does the nature of the good matter?" Journal of Environmental Economics and Management **19**: 244-263.

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

Labao, R., H. Francisco, D. Harder and F. Santos (2008). "Do Colored Photographs Affect Willingness to Pay Responses for Endangered Species Conservation?" Environmental and Resource Economics **40**(2): 251-264.

Lauria, D. T., B. Maskery, C. Poulos and D. Whittington (2009). "An optimization model for reducing typhoid cases in developing countries without increasing public spending " Vaccine **27**: 1609-1621.

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

Lienhoop, N. and D. Macmillan (2002). "Valuing wilderness in Iceland: New approaches in contingent valuation." Icelandic Journal of Geography **18-19**(1): 29-40.

Loomis, J. B. (1989). "Test-retest reliability of the contingent valuation method: a comparison of general population and visitor responses." American Journal of Agricultural Economics **February**: 76-84.

Loomis, J. B. (1990). "Comparative reliability of the dichotomous choice and open-ended contingent valuation techniques." Journal of Environmental Economics and Management **18**: 78-85.

Lucas, M., M. Jeuland, J. Deen, N. Lazaro, M. MacMahon, A. Nyamete, A. Barreto, L. V. Seidlein, A. Cumbane, F. F. Songane and D. Whittington (2007). "Private demand for cholera vaccines in Beira, Mozambique." Vaccine **25**(14): 2599-2609.

Macmillan, D. C., L. Philip, N. Hanley and B. Alvarez-Farizo (2002). "Valuing the non-market benefits of wild goose conservation: a comparison of interview and group-based approaches." Ecological Economics **43**: 49-59.

Nabangchang, O. (2008). "Mobilizing resources for marine turtle conservation in Asia: A cross-country perspective." ASEAN Economic Bulletin **25**(1): 60-69.

Reiling, S. D., K. J. Boyle, M. L. Phillips and M. W. Anderson (1990). "Temporal reliability of contingent values." Land Economics **66**(2): 128-134.

Subade, R. F. (2007). "Mechanisms to capture economic values of marine biodiversity: The case of Tubbataha Reefs UNESCO World Heritage Site, Philippines." Marine Policy **31**(2): 135-142.

Svedsater, H. (2007). "Ambivalent statements in contingent valuation studies: inclusive response formats and giving respondents time to think." Australian Journal of Agricultural and Resource Economics **51**(1): 91-107.

Teisl, M. F., K. J. Boyle, D. W. McCollum and S. D. Reiling (1995). "Test-retest reliability of contingent valuation with independent sample pretest and posttest control groups." American Journal of Agricultural Economics **77**(613-619).

Vaughan, W. J. and D. J. Rodriguez (2001). "Obtaining welfare bounds in discrete-response valuation studies: comment." Land Economics **77**(3): 457-465.

Welsh, M. P. and G. L. Poe (1998). "Elicitation Effects in Contingent Valuation: Comparisons to a Multiple Bounded Discrete Choice Approach." Journal of Environmental Economics and Management **36**(2): 170-185.

Whitehead, J. (2005). "Environmental Risk and Averting Behavior: Predictive Validity of Jointly Estimated Revealed and Stated Behavior Data." Environmental and Resource Economics **32**(3): 301-316.

Whitehead, J. C. and T. J. Hoban (1999). "Testing for temporal reliability in contingent valuation with time for changes in factors affecting demand." Land Economics **75**(3): 453-465.

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.

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

Whittington, D., D. Sur, J. Cook, S. Chatterjee, B. Maskery, M. Lahiri, C. Poulos, S. Boral, A. Nyamete, J. Deen, L. Ochiai and S. K. Bhattacharya (2008). "Rethinking cholera and typhoid vaccination policies for the poor: Private demand in Kolkata, India." World Development **37**(2): 399-409.

TABLES AND FIGURES

Table 1. Summary of time-to-think studies on cholera and typhoid vaccines

	Kolkata, India	**Matlab, Bangladesh**	**Beira, Mozambique**	**Hue, Vietnam**
Vaccine	Cholera, Typhoid	Cholera	Cholera	Cholera, Typhoid
Respondent demand elicitation method	Single binary choice (BC) , payment card (PC)	BC	PC	Sequence of six multinomial discrete choices (MDC)
Household demand elicitation method	Open-ended response on quantity demanded based on randomly-assigned price (CT)	CT	CT	None
Given price overnight?	No	No	Yes	Yes, given valuation tasks overnight
Sample sizes	559	591	996	400

Notes: BC = single binary choice, with follow-up for whether the respondent would take a free vaccine and would pay anything for it, PC=payment card, MDC=multinomial discrete choice, CT = count of vaccines for household members. Other elicitation approaches were used in these sites (see text); this table refers only to data collected as part of a TTT split-sample experiment.

Figure 1. Diagram explaining the traffic light analogy to Kolkata respondents for the payment card exercise

Table 2. Socioeconomic characteristics by study site and TTT treatment

	Kolkata		Matlab		Beira		Hue	
	NTTT	TTT	NTTT	TTT	NTTT	TTT	NTTT	TTT
Male (%)	55%	48%	47%	51%	47%	47%	39%	38%
Age	36 (8.1)	35 (7.7)**	41 (9.8)	40 (9.6)	37 (11)	37 (12)	46 (8.7)	46 (9.9)
Percent with 1-9 years of school	42%	46%	53%	53%	77%[a]	78%[a]	42%	54%**
Percent with 12 years of school or vocational	28%	31%	10%	10%	15%[a]	9%[a]***	29%	28%
Percent with university or professional degree	18%	16%	2%	1%	14%	12%	26%	11%***
Monthly per capita income (US$)	$29 (51)[b]	$26 (31)[b]	$14 (10)[b]	$14 (10)[b]	$21 (25)[b]	$20 (25)[b]	$25 (15)[b]	$20 (14)[b]***
Household size	5.4 (2.4)	5.2 (2.2)	5.6 (2.0)	5.9 (2.0)	6.1 (2.4)	6.0 (2.3)	5.3 (2.1)	5.4 (2.0)
N=	278	273	312	279	499	495	200	200

Notes: Unless noted, summary statistics are mean (standard deviation). Stars refer to a two-tailed t-test for difference in means across TTT subsamples (within sites), ***=1%, **=5%, *=10%. Per capita income is defined as total household income divided by the number of household members. [a]The education categories used in Beira were 6-7 years and 8-10 years, so this percentage reports those with 1-10 years of education. [b]This variable has many missing observations; sample sizes for this variable, starting with Kolkata NTTT and reading to the right are: 278 , 273, 312, 279, 258, 159, 176, 180.

Table 3. Percent of respondents who said they would purchase a vaccine for themselves (BC), by price and time to think.

Matlab Respondent BC

Gov't Service Area	US$0.15	US$0.37	US$0.74	US$1.1	US$4.5	US$9.0
NTTT- % yes	89%	81%	68%***	38%	15%	7%
TTT- % yes	85%	68%	36%	33%	8%	8%

ICDDR-B Area	US$0.15	US$0.37	US$0.74	US$1.1	US$4.5	US$9.0
NTTT- % yes	65%	65%	54%	67%***	22%	13%
TTT- % yes	78%	59%	50%	31%	10%	13%

Kolkata Respondent BC

Cholera	US$0.22	US$0.56	US$1.11	US$11.1
NTTT - % yes	89%	68%	63%	18%
TTT- % yes	82%	60%	49%	18%

Typhoid	US$0.22	US$0.56	US$1.11	US$11.1
NTTT- % yes	91%	86%	63%	20%
TTT- % yes	88%	69%	52%	9%

Notes: GSA=government service area, ICDDRB = area served by ICDDR-B (see Islam et al 2008 for more details). *, **, and *** refers to statistical significance at the 10%, 5% and 1% levels of a difference using a two-tailed Fisher's exact test (for detailed test results, including p-values as well as results from one-tailed Fisher's exact tests and t-tests, see Table A9 in the appendix).

Figure 2. Percent of respondents in Kolkata who were certain that they would purchase the vaccine for themselves, by price and time to think (from PC exercise).

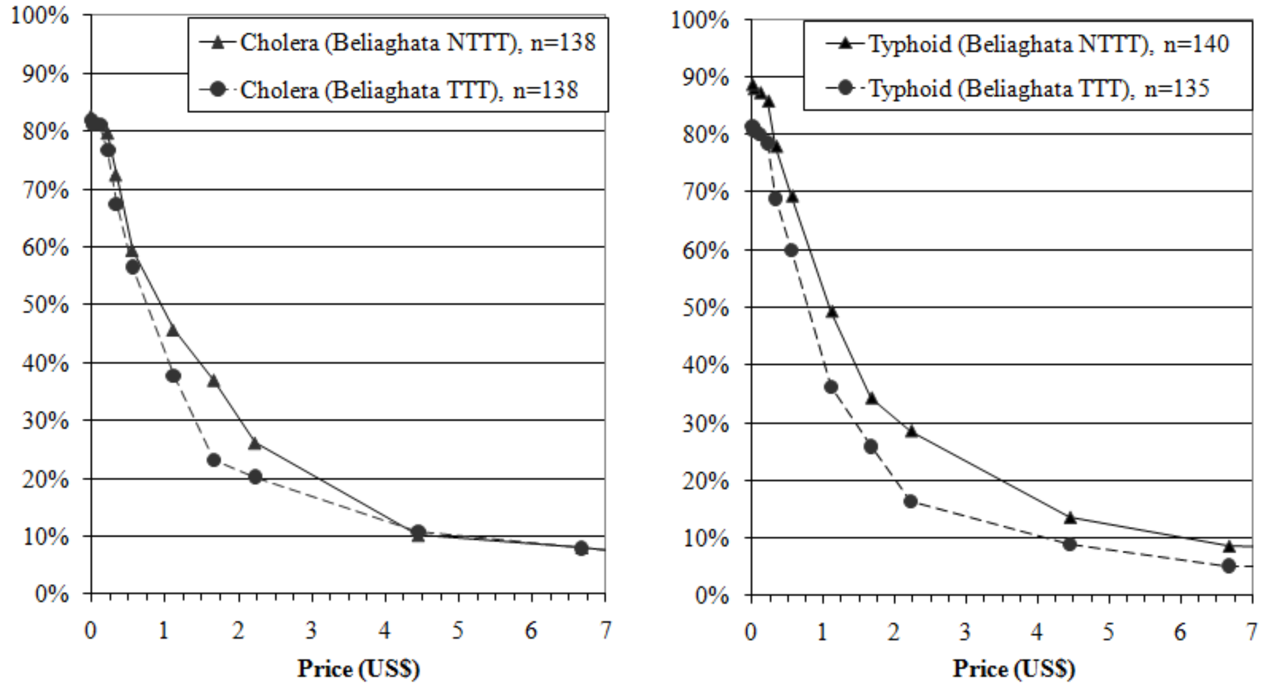
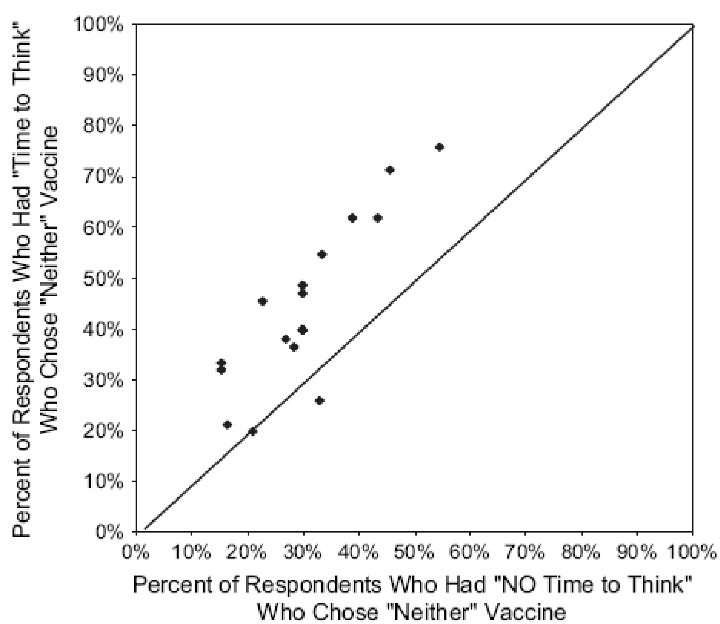


Figure 3. Percent of respondents in Hue (Vietnam), with and without time to think, who chose neither vaccine (18 tasks) (reprinted from Cook et al. 2007)



Note: Each data point reflects the answers of ~132 respondents on each task (66 with time to think, and 66 without).

Table 4. Respondent and household willingness-to-pay estimates (US$) for cholera and typhoid vaccines

	Respondent WTP (mean [90%CI])				**Household WTP** (mean [90% CI])
	Kristrom mid-point (BC)	Probit (BC)	Interval (PC)	Mixed logit (MDC)	Negative binomial count (CT)
Kolkata Cholera NTTT	5.2 [4.2 – 6.2]	5.4 [4.4 – 6.3]*	2.7 [2.6 – 2.9]***	n/a	37 [30 - 45]
Kolkata Cholera TTT	4.6 [3.4 – 5.7]	3.2 [2.3 – 4.0]	2.2 [2.1 – 2.4]	n/a	28 [23 – 35]
% reduction	*12%*	*41%*	*19%*		*22%*
Kolkata Typhoid NTTT	5.5 [4.5 – 6.6]	5.4 [4.6 – 6.2]*	2.8 [2.5 – 3.1]***	n/a	31 [26 – 36]
Kolkata Typhoid TTT	3.9 [3.0 – 4.8]	3.7 [3.0 – 4.4]	2.0 [1.8 – 2.2]	n/a	23 [20 – 27]
% reduction	*29%*	*31%*	*28%*		*23%*
Beira Cholera NTTT	n/a	n/a	0.76 [0.74 – 0.78]***	n/a	11 [10 – 12]***
Beira Cholera TTT	n/a	n/a	0.54 [0.52 – 0.56] [a]	n/a	7.2 [6.5 – 8.0]
% reduction			*32%*		*26%*
Matlab (Govt Area) Cholera NTTT	2.3 [1.7 – 2.9]	2.5 [2.1 – 2.9]*	n/a	n/a	13 [12 – 15]**
Matlab (Govt Area) Cholera TTT	1.5 [1.0 – 2.0]	1.1 [0.5 – 1.8]	n/a	n/a	8.7 [7.6 – 9.8]
% reduction	*35%*	*54%*			*34%*
Matlab (ICDDRB) Cholera NTTT	3.3 [2.1 – 4.5]	2.0 [1.2 – 2.8]	n/a	n/a	11 [9.0 – 12]**
Matlab (ICDDRB) Cholera TTT	1.6 [1.1 – 2.2]	0.6 [0.0 – 1.3]	n/a	n/a	6.9 [5.9 – 8.0]
% reduction	*50%*	*68%*			*34%*
Hue Cholera NTTT [b]	n/a	n/a	n/a	1.9 [-0.1 – 3.9]	n/a
Hue Cholera TTT	n/a	n/a	n/a	-0.09 [-0.4 – 0.2]	n/a
% reduction				*105%*	
Hue Typhoid NTTT	n/a	n/a	n/a	4.4 [1.8 – 7.0]	n/a
Hue Typhoid TTT	n/a	n/a	n/a	1.7 [0.2 – 3.2]	n/a
% reduction				*60%*	

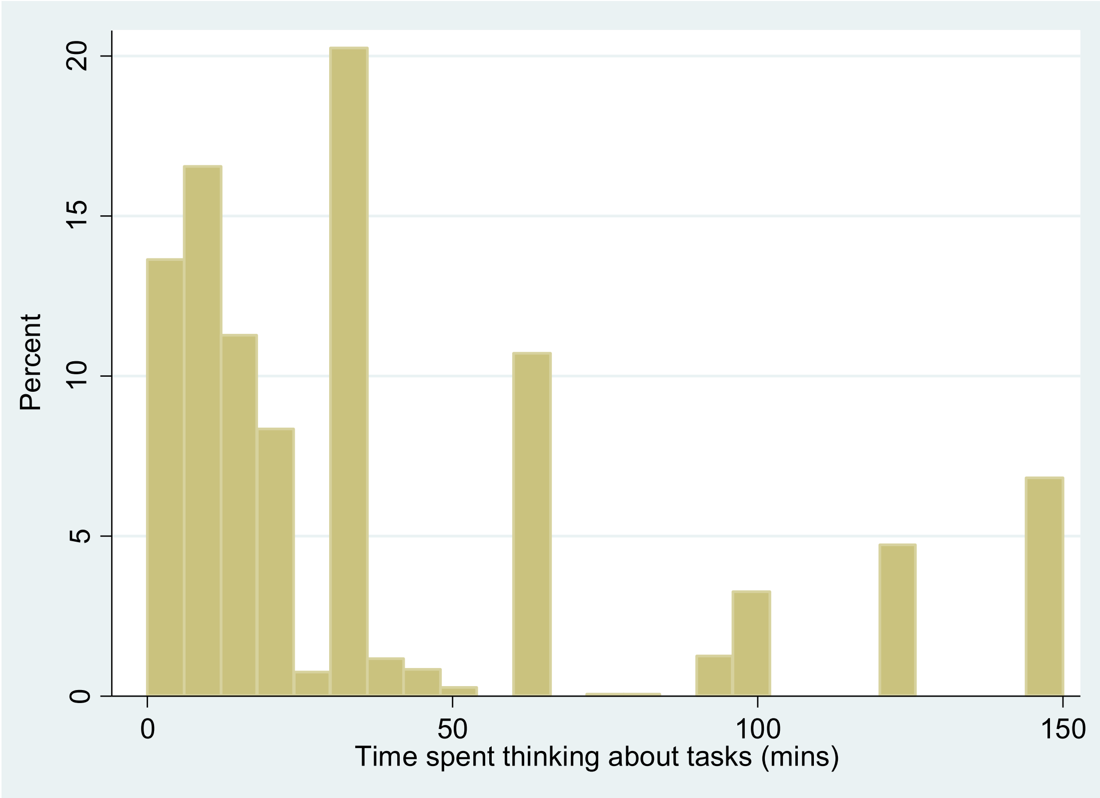
Notes: *, **, and *** refer to differences in mean WTP which are statistically significant at the 10%, 5%, and 1% levels. [a] Estimates shown for the Beira subsample averages together the groups who did the "traffic light" exercise after the single-price household question and vice-versa (i.e groups "TL1" and "TL2"). [b] The effectiveness and duration of the vaccine offered to respondents in Hue varied. Results here are for the characteristics most similar to the real "next-generation" vaccine: 50% effective for 3 years for cholera, and 70% effective for 3 years for typhoid.

Table 5. Average percent of household members who would be vaccinated at specified price (from CT exercise), by time to think (TTT) in Kolkata, Matlab and Beira.

Matlab Household CT (cholera)						
Gov't Service Area	US$0.15	US$0.37	US$0.74	US$1.1	US$4.5	US$9.0
NTTT- % hh vaccinated	82%**	83%***	61%***	39%	17%**	5%
TTT- % hh vaccinated	67%	58%	27%	31%	3%	8%
ICDDR-B Areas	US$0.15	US$0.37	US$0.74	US$1.1	US$4.5	US$9.0
NTTT- % hh vaccinated	58%	61%	53%	58%***	15%	1%
TTT- % hh vaccinated	57%	45%	39%	22%	7%	12%
Beira Household CT (cholera)	US$0.2	US$0.82	US$1.64	US$2.86	US$4.08	
NTTT- % hh vaccinated	76%**	57%***	43%***	30%**	23%	
TTT- % hh vaccinated	63%	36%	24%	19%	15%	
Kolkata Household CT						
Cholera	US$0.22	US$0.56	US$1.11	US$11.1		
NTTT- % hh vaccinated	88%*	58%	55%*	27%		
TTT- % hh vaccinated	74%	57%	40%	22%		

Notes: GSA=government service area, ICDDRB = area served by ICDDR-B (see Islam et al 2008 for more details). *, **, and *** refers to a statistically significant difference at the 10%, 5% and 1% levels with a two-tailed t-test (for detailed test results, including p-values and Kolmogorov-Smirnov tests of equality of distributions, see Table A10 in the appendix).

Figure 4. Amount of time that time-to-think respondents reported spending on the task, all four sites pooled (n=1449)



Note: the rightmost bar collapses all times longer than 150 minutes.

Table 6. Debriefing questions for time to think respondents.

	Kolkata	**Hue**	**Beira**	**Matlab**	**All sites**
Time spent thinking:					
Mean (SD)[a]	28 (37)	32(38)	52 (52)	30 (35)[b]	37 (43)
Median[a]	15	20	30	20	20
Percent who consulted with a household member	74%	56%	81%	64%	69%
Percent who consulted with someone outside the household	25%	4%	36%	27%	23%
Percent who used other information	6%	2%	n/a	n/a	3%
N=	*272*	*400*	*487*	*279*	*1438*

[a] Drops 11 records where reported time spent thinking was over 5 hours, including one respondent who reported thinking about the task for 24 hours.
[b] This data based on n= 436 for this question in Beira