A game of two halves? Starting point bias within the bidding game method

Paul McNamee, Health Economics Research Unit, University of Aberdeen, UK

Laura Ternent, Department of Public Health, University of Aberdeen, UK

Adjima Gbangou, Centre de Recherche en Sante de Nouna, Nouna, Burkina Faso

Danielle Belemsaga, Centre Muraz, Bobo Dioulasso, Burkina Faso

David Newlands, Aberdeen Business School, University of Aberdeen, UK

Paper presented to the HESG Meeting, University of Birmingham, January 3-5 2007.

WORK-IN-PROGRESS
PLEASE DO NOT QUOTE

Introduction

Valuation of the effects of health care can help policy makers in resource allocation decisions concerning which health care programmes to fund. Contingent Valuation (CV) is a method developed to provide monetary valuation of benefits for use in cost-benefit-analysis (CBA). One argument for adopting this approach to valuation is that many of the benefits of health care are 'intangible' or unobservable. An example of this is the level of reassurance offered by health care, such as the information provided by screening for health status. In addition, some benefits may arise through 'externality' effects; e.g. one individual may derive benefit from another persons' receipt or potential receipt of health care.

One necessary but not sufficient condition for obtaining valid estimates of benefit is that respondents are required to assume that a market exists for a health care programme or health benefit, and are asked to reveal the maximum they would be willing to pay (WTP) to obtain such a programme or benefit (Drummond et al. 1997) if they were able to buy it alongside other goods and services. For a given level of income, higher WTP valuations indicate that individuals would derive greater benefit from the programme under consideration.

As many developing countries have less developed market structures and prices for goods and services, the use of a technique such as CV in these countries may lead to less robust estimates of benefit than other methods. On the other hand however, CV may be a more acceptable technique in these settings, as in many cases health care is funded through out-of-pocket payments. This is because the feasibility and validity of the method is partly determined by the extent to which respondents are familiar with paying for health care. In addition, in many developing nations, the prices of many goods and services, including health care, are not fixed. This suggests that iterative methods of obtaining WTP estimates may be better than non-iterative methods.

One type of iterative method is the Bidding Game (BG) technique. This uses a predetermined search algorithm to bid the respondent up or down, conditional upon responses to the prompted values. However, a well-known problem with the BG method is starting-point bias (for example, Desvousges et al 1987). Starting point bias is present if WTP estimates are influenced by the amounts used to begin the bidding game. Bias may arise through an anchoring effect: the respondent may think that the initial bid represents the value, or very close to the value, of the good or service in question. In addition, the bidding game may suffer from incentive incompatibility problems: the respondent may behave in a strategic fashion in the follow-up questions.

Most studies that have investigated starting point bias in health and health care find evidence of such bias (Bhatia 2005; Frew et al 2004; Phillips et al 1997; Stalhammer 1996), although three studies failed to detect bias (O'Brien & Viramontes 1994; O'Brien et al 1998; Onwujekwe & Nwagbo 2002). In the case of the first two studies, it is possible that small sample sizes may have produced a failure to reject the null hypothesis of no starting point bias. In the latter study, the regression analysis did not include income. This could have led to a similar error, as a recent paper reports that insufficient random assignment of bids across different income groups can significantly affect estimated WTP (Takemura et al 2005).

The study by Onwujekwe & Nwagbo (2002) raises the possibility however that the BG method may be less susceptible to starting point bias in a developing country. We sought to examine this hypothesis by conducting a study in Burkina Faso, West Africa that examined the extent to which the BG method suffered from starting point bias. This formed part of a larger study concerned with assessing the feasibility and validity of existing methods to measure and value the outcomes of maternal health programs in developing countries.

Methods

Data and survey design

A total of 1,236 male and female individuals resident in Nouna, Burkina Faso, West Africa participated in the survey. Nouna is located in North West Burkina Faso, about 300 km from the capital Ouagadougou, and is a largely rural area, with Nouna town being semi-urban. The town and surrounding area is dry orchard Savannah, with a large proportion of the population earning an income from subsistence farming. Burkina Faso is one of the poorest countries in the world, with GDP (PPP) per capita equal to $1285 in 2005.

Individuals were randomly assigned to one of three different starting bids (low, medium and high) for valuation of a 25% or 50% reduction in the absolute number of maternal deaths in one year. The values for the initial starting bids were determined following analysis of the results of a previous study that examined willingness to pay for community based health insurance (Dong et al 2003), group discussion with researchers involved in the previous study, and the results of pre-test and pilot studies. In the pre-test and pilot studies, using samples sizes of 30 and 80 separate respondents respectively, we sought to examine the applicability of the initial bid values to achieve a balanced bid design, with percentiles of 0.25 and 0.75 for the upper and lower initial starting values respectively. The starting bids were 2000, 5000 and 8000 CFA (approximately 1000 CFA = £1). Although the middle and upper bids were acceptable, with 40% and 25% (respectively) of people being willing to pay these initial amounts, only 25% of individuals said ‘yes’ to the initial low bid of 2000CFA. Thus, in the pilot study, the lower bid was revised to 1000 CFA. The proportion of respondents willing to pay the new lower initial bid rose to 81%. Thus, a bid vector of 1000, 5000 and 8000 CFA was used in the main survey.

Individuals were selected by identifying a random but representative sample of households from an on-going Demographic Household Survey (DHS), a subset of a larger Demographic Surveillance Survey (DSS) in Nouna, Burkina

Faso. Six hundred and eighteen households in total were identified, with one male head of household and one other female member of the household selected. Demographic and socioeconomic information on individual respondents were collected separately via the DHS immediately prior to the CV survey. Other questions relating to knowledge and experience of maternal morbidity and mortality were asked prior to the valuation questions. As there were 926 households in the DHS, approximately 67% of all households were selected. To achieve a sample that was representative of the Nouna district, a different number of households in each village and Nouna town were selected in order that each contributed 67% of their total number of households to the sample. The first house in each village to be interviewed was chosen randomly, with every n^{th} house chosen afterwards according to household survey number.

The survey was conducted during June and July 2005 by 18 interviewers in a local dialect of the French language, Dioula. All had received three days training in survey design and questionnaire administration by members of the Nouna research group, Centre de Recherche en Sante de Nouna (CRSN). Male headed households were asked to nominate an adult female member of the household to complete the questionnaire. They were instructed to do this by thinking about who they would have liked to answer on their behalf had they not been able to take part in the interview. Female headed household were excluded from the survey, as there were no adult males in such households. Interviewers were instructed to conduct both interviews separately. However, they were also informed that it was important to avoid undue delays in commencing the second interview, to minimise the opportunity for the first value to influence the second value. Common initial bids were used for both male and female members of the same household.

The bidding algorithm and the scenario that was valued by respondents are shown in Appendix 1, Boxes 1 and 2. Individual willingness to pay is recovered from responses to the bidding algorithm. For the 5000CFA and 8000CFA initial amounts, if the first response is ‘yes’, the interviewer increases the bid amount in increments of 500 CFA up to a maximum of four

times until the respondent says ‘no’. At that point, the WTP is estimated to be equivalent to the previous bid amount. If the respondent says ‘yes’ throughout, they are asked to state their maximum WTP in an open-ended question. If the first response is ‘no’, the interviewer reduces the amount by 500 CFA up to a maximum of four times until the respondent says ‘yes’. At that point, the WTP is estimated as equivalent to that bid amount. If the respondent says ‘no’ throughout, they are asked to report WTP in an open-ended question. The process was identical in the 1000CFA initial bid, except that three increments of 250CFA were used for ‘no’ responses.

It can be seen that the payment vehicle was a one-off out-of-pocket payment for a service re-organisation that would lead to a one time reduction in the number of maternal deaths. To minimise the potential for free-riding, individuals were informed that a threshold level of payment had to be reached for the re-organisation to occur.

Estimation strategy

Most previous studies that test for starting-point bias estimate a regression model that includes a set of dummy variables to represent the assigned bids, or the log of the initial bid amount, and examine the null hypothesis that the coefficient/s are equal to zero (i.e. no starting point bias).

This paper adopts this approach and also employs the estimation strategy used by Whitehead (2002), in which shift, anchoring model, and shift and anchoring models were used to explain WTP values. All models assume that the response to the first question is an indicator of true WTP, and that the second and subsequent response of an individual *i* is dependent on the first bid amount (b_{1i}). A Bayesian framework is employed through the assumption that the response to the initial bid is based on a prior point or interval estimate of unobserved WTP (WTP_{pi}) of an individual *i*, and that this is transformed to observed WTP (WTP_{si}) based on new information provided by the second and subsequent bid amounts (b_{2i}).

In the shift model (Alberini et al (1997)), WTP is held to systematically shift according to the following process:

$$WTP_{si} = WTP_{pi} + \delta \qquad [1]$$

where δ is estimated as a dummy variable = 0 for an initial yes (no) response to the first question, followed by a no (yes) response to the second question, and δ = 1 otherwise. Initially developed for double-bounded dichotomous choice data to provide a test for incentive incompatibility, an extension to bidding game data involves the same test of whether individuals use different decision rules when answering the first and subsequent questions. It might be expected for instance that an initial 'yes' response induces respondents to say 'no' to the second question, due to a perception that an agreed price has already been reached. In addition, an initial 'no' response might induce respondents to say 'no' in the second and subsequent questions, as the respondent would now have reason to believe that the price is not fixed, because they have been asked a follow-up question. An individual could then respond to further follow-up questions according to usual market behaviour, such as seeking to obtain the good for at least half the stated price. In both situations, incentives to say 'no' to the second question cause a downward shift of WTP.

In Burkina Faso, like many other West African countries, the first price posted by the seller is often high and is used as a signal by the buyer to engage in bidding to lower the price by a substantial margin. As a result, for an initial 'no' response, we would expect to observe a high proportion of 'no' responses to follow-up questions. An initial 'yes' response to the first question would not be expected to lead to any further 'yes' responses, as the respondent may consider that an agreed price has already been set. Thus, we expect the shift parameter to have a negative sign.

In the anchoring model (Herriges & Shogren 1996), WTP is assumed to be an average of WTP_{pi} and the initial starting amount b_{1i} :

$$WTP_{si} = (1 - \gamma)\, WTP_{pi} + \gamma\, b_{1i} \qquad [2]$$

where γ is estimated as an interaction term between the bid value (or log of the bid value) and δ. If $\gamma = 0$ there is no anchoring and starting point bias is absent, and as $\gamma \rightarrow 1$ there is strong anchoring, with the final WTP value approaching the initial starting bid. If $\gamma < 0$, the final WTP value is not an average of WTP_{pi} and b_{1i} , suggesting that respondents' answering behaviour is not consistent with such an 'averaging' hypothesis. The model assumes therefore that respondents perceive the initial bid amount as an indicator of expected value; a second price will be perceived to be too high or low, and so an initial 'yes' answer will lead to a 'no' response to the second question, and vice versa.

In the presence of anchoring, the effect on final WTP depends on response to the initial bid amount, which is likely to be affected by the size of the initial starting bid. Thus, where starting points are below (above) true WTP, WTP estimates from follow-up questions will be biased downwards (upwards). In the context of the CV survey employed here, it might be expected that higher initial bids would be more likely to trigger greater haggling to reduce the price, relative to lower initial bids where the scope for haggling is more limited. Thus, we expect there to be no 'averaging' between WTP_{pi} and b_{1i} and the anchor parameter to have a negative sign.

Finally, in the shift and anchoring model (Whitehead 2002), WTP is assumed to be dependent on both shift and anchoring effects:

$$WTP_{si} = (1 - \gamma)\, WTP_{pi} + \gamma\, b_{1i} + \delta \qquad [3]$$

In this model, it is assumed that respondents are not only more likely to provide a 'no' response to the follow-up questions, but in the event that they also anchor valuations to the first bid, there will be an even stronger incentive to provide a 'no' response to follow-up questions. Similar to the anchoring model, a negative value for γ would suggest that respondents' answering behaviour is not consistent with an 'averaging' hypothesis.

Results

A total of 1236 individuals received the CV survey, with 414 receiving the 1000CFA initial bid, 412 the 5000CFA bid and 410 the 8000CFA bid. Of these, thirty one respondents stated that they would not be prepared to pay anything. A further eleven respondents stated an annual WTP value that was higher than their annual income, and were excluded from further analysis. In addition, individual income data were missing for 269 respondents. A total of 956 respondents therefore produced feasible WTP responses (77%), of which 925 respondents provided non-zero values.

Table 1 describes the background characteristics of the 956 respondents. There was approximate balance in respondent characteristics across the three starting bids, with no statistically significant differences. However, respondents who received the lowest bid had a slightly higher mean annual income than the other groups. In addition, there were some differences in the proportion of respondents who had knowledge of a previous maternal death in their community.

Table 2 reveals that there was a positive relationship between willingness to pay and the starting bid. Individuals who received the 5000CFA bid were prepared to pay on average approximately 1400CFA more than those who received the 1000CFA bid. Relative to the 5000CFA bid, respondents who received the 8000CFA bid were prepared to pay on average an additional 570CFA.

Figure 1 provides an indication of the distribution of WTP by initial bid. There is a fairly consistent pattern in the responses across all three formats. There is a spike at the initial starting bids, with further spikes occurring at approximately 50% of the opening bid, and 50% of the final bid. With the 8000CFA bid, the second spike occurs a little above the 50% point, at 5000CFA, with further occurrences at 2500CFA, 1000CFA and 500CFA. With the 5000CFA opening bid, the second spike occurs at the 50% point,

with occurrences at the 1000CFA and 500CFA levels. With the 1000CFA bid, there is one occurrence at the 500CFA level.

Table 3 presents the results of ordinary least squares regression estimates of log WTP for those with non-negative WTP values. In the base model, no account is taken of whether WTP differs according to whether follow-up questions are asked. When the parameter δ is included in the shift model to account for this effect, the coefficient on the log bid amount increases by 18% from 0.34 to 0.40. The shift parameter δ is negative and statistically significant, indicating that those with follow-up questions had a lower WTP. In the anchoring model, the anchor parameter is negative and statistically significant. This indicates that the response to the follow-up questions was not anchored to the first bid. In the shift and anchoring model, the regression coefficient for the shift parameter δ is now positive, and the anchor parameter has a larger negative value relative to the anchoring model.

Income was found to have a positive relationship with WTP, whilst age had a small negative relationship. Experience of a previous maternal complication in the household and knowledge of a maternal death in the community both lead to increased estimates of WTP. A higher level of education was associated with a higher WTP. Living in Nouna town and being married were associated with a non-significant reduction in WTP. Surprisingly, experience of a previous maternal death in the household had no significant effect on WTP, and females had a significantly lower willingness to pay than men.

Discussion

There was a statistically significant relationship between willingness to pay for a reduction in maternal deaths and the initial starting bid, but the final values were not anchored to the starting amounts. This suggests that the method generates incentive incompatible responses. The main implication of such responses is that individuals who received the lowest starting point amount would be predicted to provide a higher valuation had they received a higher starting point bid, and vice versa. However, as equal proportions of individuals were randomised to receive one of the three bids, which were subject to pre-test and piloting to obtain optimal estimates of bid design, the effect on aggregate mean willingness to pay values is likely to have been minimised. The results also suggest however that the use of higher starting points throughout for all three groups would have resulted in a higher aggregate willingness to pay value. Ideally, incentive incompatibility and starting point bias could be more fully examined in future studies using a test-retest design within respondents, with some respondents receiving a different starting bid and others receiving an identical bid to the first test.

There are several potential weaknesses associated with the study. First, the description of the good that was valued could have induced strategic bias. That is, if respondents perceived the exercise as playing an important role in future policy making, there may have been an incentive to uplift values in order to ensure that the re-organisation took place. It is possible that our respondent sample may have behaved in this way, as they were identified from the same sampling frame used in a previous CV study conducted to estimate the out-of-pocket charge for community based health insurance (Dong et al 2003). This effect would lead to biased mean estimates of willingness to pay.

Second, it would be desirable to repeat the analysis using double-bounded probit models. As this paper reports work-in-progress, there is some uncertainty over whether the shift, anchoring and shift and anchoring models can also be applied to the final responses from bidding game data. In the

paper by Whitehead (2002), only random effects probit models were estimated from the second and subsequent responses.

In the face of mounting evidence for the existence of incentive incompatibility and starting point bias, it may be natural to ask the question of why researchers continue to undertake CV studies using the bidding game format. To address this question, it is important to consider the extent of biases associated with other methods. For example, the single and double-bounded dichotomous choice (DBDC) formats may suffer from "yea-saying" bias; a tendency to say "yes", out of politeness or other reasons, irrespective of the offer price. This effect however would be expected to vary across the range of initial bid amounts and so a similar anchoring problem would be expected to emerge. With regard to other methods, the payment card technique is known to be prone to range bias (Whynes et al 2004). Therefore, as different kinds of biases and errors remain possible within all methods, it is important that reliability and validity issues continue to be investigated in any survey.

The acid test that would help inform the choice between the available methods is one based on a comparison with actual purchasing behaviour based on revealed preferences. We are aware of two studies that have conducted such tests with the bidding game method in health care. Onwujekwe et al (2005) found that there was a divergence of 69% between actual and stated WTP with the bidding game method, compared with 79% in the DBDC format and 49% in the structured haggling format. Bhatia (2005) also found a divergence between stated and actual behaviour: 75%, 67% and 29% of households randomised to high, middle and low starting points were expected to buy treated mosquito nets, yet when the product was subsequently offered for sale at the middle price, all households were equally as likely as each other (approximately 50% of households) to buy the nets. Further studies are clearly required to gather more information about the sources of divergence. For now, it can only be concluded that this study adds to the existing evidence that it is possible to manipulate willingness to pay through selection of the initial bid.

References

Alberini A, Kanninen B, Carson R. Modeling response incentive effects in dichotomous choice contingent valuation data. *Land Economics* 1997; 73: 309-324.

Bhatia M. From evidence to calibration for starting point bias: willingness to pay for treated mosquito nets in Gujarat, India. *Applied Economics* 2005; 37: 1-7.

Desvousges W, Smith V, Fisher A. Option price estimates for water quality improvements: a contingent valuation study for the Monongahela River. *Journal of Environmental Economics and Management* 1987; 14: 248-267.

Dong H, Kouyate B, Cairns J, Sauerborn R. A comparison of the reliability of the take-it-or-leave-it and the bidding game approaches to estimating willingness-to-pay in West Africa. *Social Science & Medicine* 2003; 56: 2181-2189.

Drummond MF, O'Brien B, Stoddart GL, Torrance GW. Methods for the Economic Evaluation of Health Care Programmes, 2nd edition. OU Press, New York, 1997.

Frew E, Wolstenholme J, Whynes D. Comparing willingness to pay: bidding game format versus open ended and payment scale formats. *Health Policy* 2004; 68: 289-298.

Herriges J, Shogren J. Starting point bias in dichotomous choice valuation with follow-up questioning. *Journal of Environmental Economics and Management* 1996; 30: 112-131.

O'Brien B, Viramontes J. Willingness to pay: a valid and reliable measure of health state preference? *Medical Decision Making* 1994; 14: 289-297.

O'Brien B et al. Assessing the value of a new pharmaceutical: a feasibility study of contingent valuation in managed care. *Medical Care* 1998; 36: 370-384.

Onwujekwe O, Nwagbo D. Investigating starting-point bias: a survey of willingness to pay for insecticide-treated nets. *Social Science & Medicine* 2002; 55: 2121-2130.

Onwujekwe O, Hanson K, Fox-Rushby J. Do divergences between stated and actual willingness to pay signify the existence of bias in contingent valuation surveys? *Social Science & Medicine* 2005; 60: 525-536.

Phillips K et al. Willingness to pay for poison control centers. *Journal of Health Economics* 1997; 16: 343-357.

Stalhammer N-O. An empirical note on willingness to pay and starting point bias. *Medical Decision Making* 1996; 16: 242-247.

Takemura S, Ohida T, Sone T, Fukuda T, Takemura Y. Influence of the absence of random assignment of bids on estimating willingness to pay using a discrete choice question. *Health Economics* 2005; 14: 209-213.

Whitehead J. Incentive incompatibility and starting point bias in iterative valuation questions. *Land Economics* 2002; 78: 285-297.

Whynes D, Wolstenholme J, Frew E. Evidence of range bias in contingent valuation payment scales. *Health Economics* 2004; 13: 183-190.

Table 1 Respondent characteristics

Bid	1000	5000	8000	p
Mean annual income (000)	304	222	257	.49
Mean age	41	42	43	.28
% female	44	47	48	.66
% mat complication	31	27	29	.53
% mat death household	4	4	4	.94
% mat death community	87	91	93	.09
Mean years schooling	1	1	1	.82
% married	94	94	92	.43
% resident in Nouna town	34	35	36	.87

Table 2 WTP estimates by initial starting bid

Bid	1000	5000	8000
Mean WTP (% of annual income)	1348 (4)	2761 (7)	3331 (8)
Standard deviation	1097	2458	3919
95% confidence interval	1227 – 1469	2495 - 3027	2890 - 3771

Figure 1 Distribution of WTP by initial starting bid

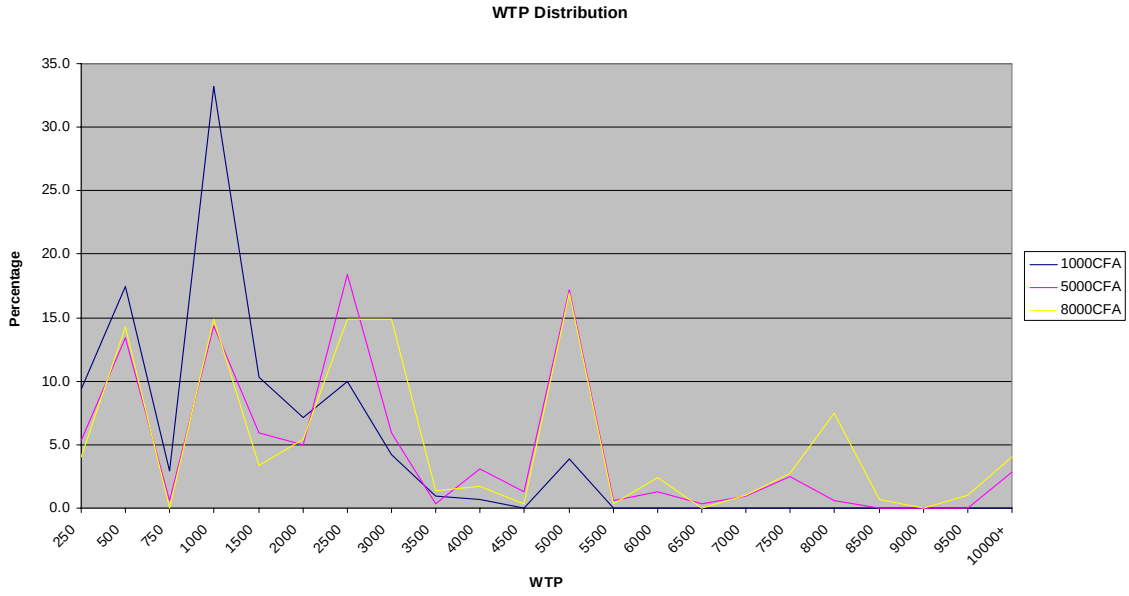


Table 3 Regression coefficient estimates of log WTP

	Base	Shift	Anchoring	Shift & Anchoring
Constant	4.11***	4.08***	3.61***	-0.20
Log annual income	0.08**	0.08**	0.08**	0.07**
Age (years)	-0.01***	-0.01***	-0.01***	-0.01***
Female	-0.22**	-0.19*	-0.19*	-0.20**
Maternal complication (=1 if yes)	0.26***	0.23**	0.23**	0.21**
Maternal death household (=1 if yes)	0.02	0.06	0.06	0.05
Maternal death community (=1 if yes)	0.22*	0.22*	0.22*	0.16
Years schooling	0.03*	0.02	0.02	0.01
Married (=1 if yes)	-0.18	-0.19	-0.19	-0.16
Resident in Nouna town (=1 if yes)	-0.10	-0.10	-0.10	-0.11
Logbid	0.34***	0.40***	0.47***	0.98***
δ		-0.54***		5.03***
Logbid * δ			-0.08***	-0.72***
Adjusted R^2	0.14	0.18	0.19	0.24
F statistic	16.62***	19.74***	21.13***	24.93

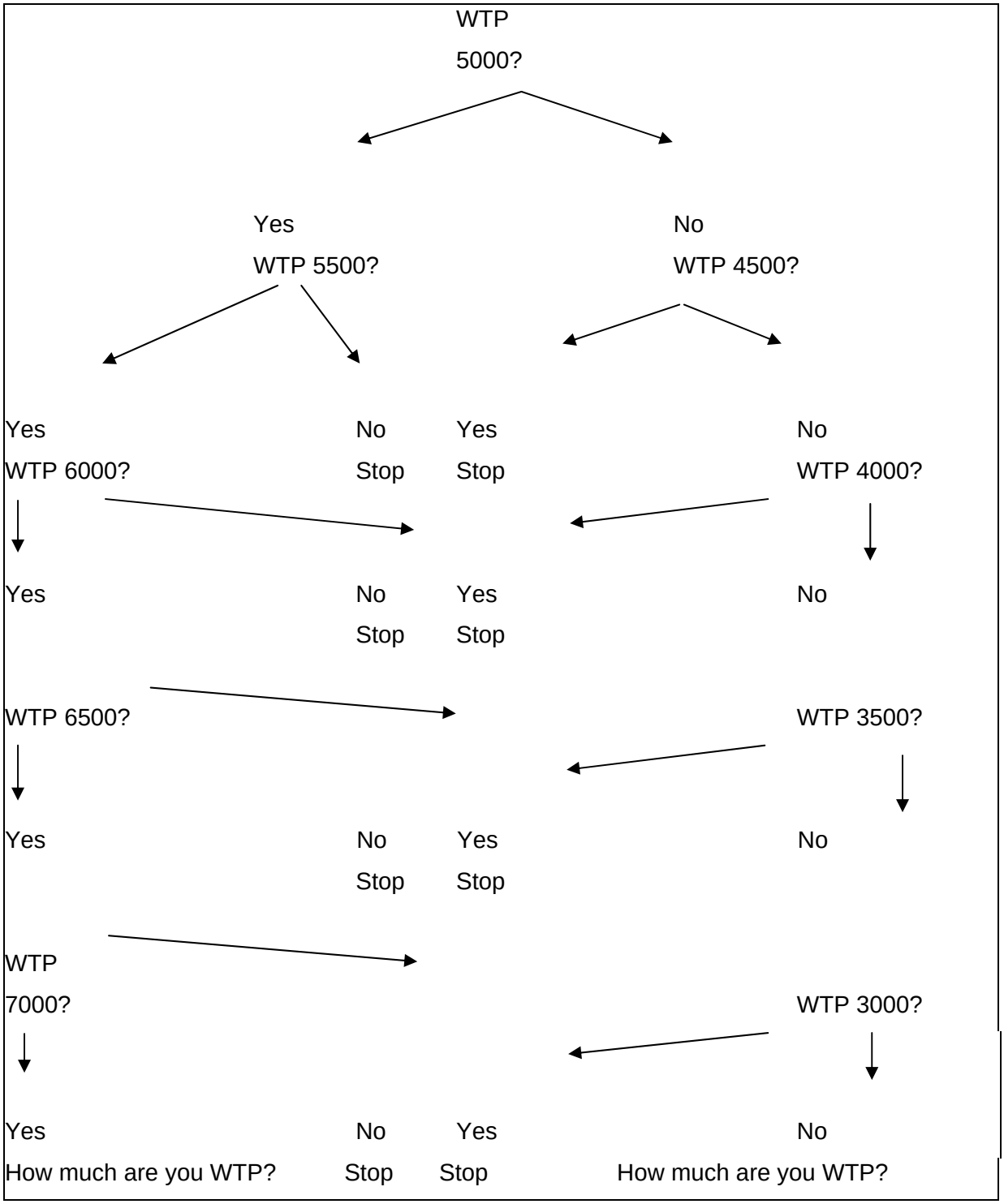
*** $p < 0.001$

** $p < 0.01$

* $p < 0.05$

Appendix 1

Box 1 Bidding algorithm for 5000CFA initial bid



Box 2 Description of CV scenario for 25% reduction in maternal deaths

I would like to ask you some questions concerning your feelings about complications and deaths amongst women who are pregnant. I'm going to do this by asking you to think about a hypothetical situation where the number of deaths and complications is reduced.

To give you a little background, health services aim to prevent illness and disease. However, in many cases, they can be re-organised in order to prevent more illness and disease from occurring. For example, different types of training could be provided for midwives.

I'm going to describe the benefits of a re-organisation of health services. This re-organisation is expected to save the lives of some mothers through the prevention of serious complications. In addition to this, the re-organisation is expected to reduce the number of other complications.

I'd like to know how important this improvement is to you.

At the moment, it's estimated that there are around **60** deaths in Nouna District every year amongst pregnant women. These women die during pregnancy, childbirth, or at a time shortly after, because of complications.

Imagine now, that if a re-organisation was to occur this year, there would be fewer deaths. I'd like you to think about a situation where **45** women would now die next year.

So, what this means is that **15** women next year will be saved as a result of this re-organisation.

If enough individuals make a contribution, the Government and other agencies will also contribute. With enough funds from individuals, the re-organisation can take place, and the lives of **15** women will definitely be saved next year in your District.

What I want to do now is present you with different sums of money and ask which amount of money you will be prepared to pay so that this re-organisation takes place.

In answering these questions, please think about how much you can afford to pay.