



**HAL**  
open science

## Dealing with internal inconsistency in double-bounded dichotomous choice: an application to community-based health insurance

Hermann Pythagore Pierre Donfouet, P. Wilner Jeanty, Pierre-Alexandre Mahieu

### ► To cite this version:

Hermann Pythagore Pierre Donfouet, P. Wilner Jeanty, Pierre-Alexandre Mahieu. Dealing with internal inconsistency in double-bounded dichotomous choice: an application to community-based health insurance. *Empirical Economics*, 2014, 46 (1), pp.317-328. 10.1007/s00181-012-0665-2. halshs-00761049

**HAL Id: halshs-00761049**

**<https://shs.hal.science/halshs-00761049>**

Submitted on 18 Dec 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Dealing with internal inconsistency in double-bounded dichotomous choice: an application to community-based health insurance

Hermann Pythagore Pierre Donfouet ·  
P. Wilner Jeanty · P.-A. Mahieu

**Abstract** Contingent valuation method is commonly used in the field of health economics in an attempt to help policy makers in their policy-making decision process. The use of the double-bounded dichotomous choice format results in a substantial gain in statistical efficiency over the single-bounded dichotomous choice format. Yet, there is internal inconsistency with a downward mean shifting in the second responses. Using data from a community-based health insurance survey, this paper aims at testing whether double certainty calibration reduces internal inconsistency in a double-bounded dichotomous choice contingent valuation survey. Results suggest that double calibration significantly reduces internal inconsistency while maintaining the efficiency gain arising from the double-bounded format.

**Keywords** Contingent valuation · Internal inconsistency ·  
Double certainty calibration · Community-based health insurance.

**JEL Classification** C15 · D6 · I38

---

H. P. P. Donfouet (✉)  
CREM, UMR 6211, University of Rennes I, 7 Place Hoche, 35 065 Rennes Cedex, France  
e-mail: donfouetz@yahoo.fr

P. W. Jeanty  
Kinder Institute for Urban Research & Hobby Center for the Study of Texas, Rice University, MS 208  
6100 Main Street, Houston, TX 77005, USA

P.-A. Mahieu  
LEMNA, University of Nantes, Chemin de la Censive du Tertre, 44 322 Nantes Cedex, France

## 1 Introduction

Contingent valuation (CV) method typically remains the most widely used method for assessing consumer preferences for non-market goods and services. Also notable is the use of CV to value marketed goods and services for which no established market exists (see e.g., [Abdullah and Jeanty 2011](#)). The value obtained from CV survey is of interest for policy-making decisions that are based on cost-benefit analysis. Thus, researchers must estimate without any bias the value that respondents attach to a particular good or policy. In an attempt to reach this goal, the single-bounded dichotomous choice (SBDC) and the double-bounded dichotomous choice (DBDC) formats have been prominently used over the past years.

The use of the open-ended format in the form “how much are you willing to pay for X (or for policy A)?” has been rescinded in favor of the SBDC format in the form “are you willing to pay Y dollars for X (or for policy A)?” because the latter is incentive compatible and mimics behavior in regular markets where people usually purchase or decline purchase a good at the posted price ([Arrow et al. 1993](#); [Bishop and Heberlein 1979](#)). However, the SBDC format provides less information about each respondent’s willingness-to-pay (WTP), resulting in less efficiently estimated WTP measures. In order to obtain more information about respondents’ WTP, [Carson et al. \(1986\)](#) developed the DBDC approach, which consists of asking another yes/no question, where a higher or a lower bid amount is offered depending on the first response. A few years later, [Hanemann et al. \(1991\)](#) demonstrated that DBDC is more efficient than the SBDC format. Since then, the DBDC format has been the mainstay in CV surveys. Yet, some skepticism remains largely because empirical applications have shown that WTP amounts from the first and the second responses were not driven by the same underlying preferences, with the former being significantly higher than the latter ([McFadden 1994](#); [Cameron and Quiggin 1994](#); [Kanninen 1995](#); [Herriges and Shogren 1996](#); [Bateman et al. 2001, 2008](#); [Burton et al. 2003](#)). This is known as internal inconsistency emerging primarily from the second responses. In fact, [McFadden \(1994, pp. 705–706\)](#) stated that the double referendum elicitation format is internally inconsistent, causing some practitioners to abandon such elicitation format. Using data obtained from a community-based health insurance (CBHI) survey, this paper aims at testing whether the double certainty calibration reduces the internal inconsistency.

Calibration has been one of the prominent techniques used in CV surveys to mitigate hypothetical bias, a tendency of respondents to state WTP amounts different from what they would declare in real settings (e.g., [Champ et al. 1997](#)). To the best of our knowledge, it has never been used to grapple with the issue of internal inconsistency in DBDC. Our results strikingly indicate that the double certainty calibration is effective at reducing the discrepancy between the welfare estimates from the single- and double-bounded models estimated from the same dataset while the efficiency gain arising from the follow-up question is maintained.

The rest of the paper is organized as follows. Section 2 provides some background on internal inconsistency in the DBDC context. Section 3 presents the methodology used, while Sect. 4 describes the survey design and data. Section 5 reports the empirical results of the study. Finally, Sect. 6 concludes.

## 2 Internal inconsistency and double certainty calibration

Several conjectures have been put forward in an attempt to explain the reasons underpinning internal inconsistency in the DBDC approach. Since the results of the CV surveys are not consequential, people may invest little effort in the valuation task and bear a range of values in mind rather than a single point (Guzman and Kolstad 2007). Hanley et al. (2009) argued that respondents would prefer to state a range of values instead of a point estimate because they are unsure about the value they place on the proposed good or policy. This uncertainty could lead to an overestimation of the mean WTP and a behavioral inconsistency (Murphy et al. 2005; Mahieu et al. 2012). Flachaire and Hollard (2007) maintained that the existence of a range may be the culprit for the internal inconsistency. Premised on the coherent arbitrariness principle, their model suggests that respondents uncertain about their true WTP are prone to anchoring effects.

The government wastage model was proposed by Carson et al. (1994) to explain the downward mean shifting in the second responses. In this model, the respondents saying “yes” to the initial bids for the provision of a public good might conceive of the higher follow-up bids as an attempt by the government to collect more funds than needed to cover the provision of the good and will say “no” to the follow-up bids since perceiving it as a waste. By the same token, respondents saying “no” to the initial bids might view the lower follow-up bids as an indication that the good being valued is of lower quality, and thus they will answer “no” to the follow-up bids. Another possible explanation is the strategic behavior model (Mitchell and Carson 1989) where the respondents answer the first questions truthfully but answer the second ones strategically. Individuals would tend to lower the bids by rejecting any additional bids proposed by the researcher. In order to avoid this strategic behavior while gaining efficiency, Cooper et al. (2002) proposed the one and one-half bound approach. Bateman et al. (2008) showed that the respondents were unfamiliar with the institutional procedures of the DBDC and they were surprised by the follow-up questions. Watson and Ryan (2007) found that, unlike respondents who felt uncertain about their responses to the first valuation question, respondents who felt certain are more likely to respond the second question in a more consistent way. A scale ranging from 1 to 5 was implemented to assess the degree of certainty, where respondents providing a score of 5 were considered certain about their WTP, while the other participants were considered uncertain.

In the current study, we use the follow-up certainty questions (FCQ) to calibrate the respondent’s WTP for both the first and the follow-up bids. As a result, we term our approach double certainty calibration. “Yes” respondents are asked how sure they are about their answers on a scale ranging from 1 to 10, where 1 means “very uncertain” and 10 “very certain.” Following Ethier et al. (2000) and Poe et al. (2002), we set a threshold of 7 out of 10 to decide whether to recode a “yes” to a “no” response. Thus, for the FCQ, all “yes” answers not receiving a score of 7 or higher are recoded as “no.” This rule is also applied to the follow-up bid of the second question. The rationale underlying the double certainty calibration is that respondents might be able to better assess the good being valued when they are given the opportunity to revise their answers. It may also be the case that a reduction in WTP as a result of calibration may lead to a reduction in the internal inconsistency in response patterns.

### 3 Econometric methods

Following Hanemann (1984), let  $v(p, q, y, s, \varepsilon)$  be the indirect utility function of the individual, where  $p$  represents the prices of the market goods,  $q$  the non-market good,  $y$  the respondent's income,  $s$  the socio-demographic characteristics such as age, income, and gender, and  $\varepsilon$  the stochastic component of preferences. Via the questionnaire, the respondent is confronted with the possibility of a change from an initial state (the status quo) to the proposed alternative (that is from  $q^0$  to  $q^1 > q^0$ ). In the status quo, the utility function of the respondent is given by  $v(p, q^0, y, s, \varepsilon)$ . When a change from the status quo  $q^0$  to the proposed alternative occurs, the utility function in the final state,  $q^1$ , is equal to  $v(p, q^1, y - A, s, \varepsilon)$ , with  $A$  being the proposed payment offered to the respondent. The respondent will answer "yes" to the offered bid if  $v(p, q^1, y - A, s, \varepsilon) \geq v(p, q^0, y, s, \varepsilon)$  and "no" otherwise. The model heavily relies on probabilities because the stochastic component of the utility function is not observable. Hence,

$$\Pr \{\text{response is "yes"}\} = \Pr \left\{ v(p, q^1, y - A, s, \varepsilon) \geq v(p, q^0, y, s, \varepsilon) \right\} \quad (1)$$

By using the compensating variation measure, the quantity  $C$  satisfies:

$$v(p, q^1, y - C, s, \varepsilon) = v(p, q^0, y, s, \varepsilon),$$

Thus, solving for  $C = C(p, q^0, q^1, y, s, \varepsilon)$  yields the respondent's maximum WTP for the change from  $q^0$  to  $q^1$ .

Hence, an equivalent condition to (1) is:

$$\Pr \{\text{response is "yes"}\} = \Pr \left\{ C(p, q^0, q^1, y, s, \varepsilon) \geq A \right\}, \quad (2)$$

In other words, respondents will say "yes" when their maximum WTP for the change from  $q^0$  to  $q^1$  exceeds the proposed bid  $A$ . Let  $G_c(\bullet)$  denote the cumulative distribution function of  $C$ , and  $g_c(\bullet)$  the corresponding density function. Then (2) becomes:

$$\Pr \{\text{response is "yes"}\} = 1 - G_c(A), \quad (3)$$

The form of the function  $G_c(A)$  and the distributional assumption of the stochastic component of the utility function determine the econometric model to be used. If the  $G_c(A)$  follows a probit standard distribution and the model to estimate is linear, then the expected mean WTP is:

$$\mu_{\text{SBDC}} = -\frac{\alpha}{\beta}, \quad (4)$$

where  $\alpha$  is the intercept and  $\beta$  the coefficient on the bid level representing the estimated marginal utility of income. Other socio-demographic variables are important

for benefit transfer and policy analysis but not needed for the tests implemented in the paper.

In order to obtain more precise welfare estimates, we consider the DBDC format. For the same level of the newly proposed CBHI, the DBDC protocol, as previously noted, entails posing two successive questions requesting a “yes” or “no” from the respondents. A positive response to the first bid question leads to a higher second bid (twice the first bid); the converse holds in the case of negative answer (half of the first bid). Thus, there are four possible responses: “yes–yes”; “yes–no”; “no–yes”; and “no–no.”

With the DBDC data at hand, we estimate the interval data probit model drawing upon [Hanemann et al. \(1991\)](#). In this model, the mean/median WTP estimates and the dispersion parameters are assumed to be the same across equations or questions and the coefficient of correlation between the random components in the two equations is assumed to be equal to one.

Let  $A^1$  denote the first bid and  $A^2$  the second bid. The bounds on the WTP are:

$$\begin{aligned} A^1 &\leq \text{WTP} < A^2 \text{ for the “yes–no” responses;} \\ A^1 &> \text{WTP} \geq A^2 \text{ for the “no–yes” responses;} \\ \text{WTP} &\geq A^2 \text{ for the “yes–yes” responses;} \\ \text{WTP} &< A^2 \text{ for the “no–no” responses;} \end{aligned}$$

The general form of the double-bounded model is:

$$\text{WTP}_{ij} = \mu_i + \varepsilon_{ij}, \quad (5)$$

where  $\text{WTP}_{ij}$  represents the amount the  $j$ th respondent is willing to pay, and  $i = 1, 2$  represents the first and the second answers, while  $\mu_1$  and  $\mu_2$  correspond to the means for the first and the second responses. Based on the assumption that  $\mu = \mu_1 = \mu_2$ , WTP for the  $j$ th individual can be written as:

$$\text{WTP}_j = \mu + \varepsilon_j \quad (6)$$

With the error term assumed to be normally distributed, the  $j$ th contribution to the likelihood function is:

$$\begin{aligned} L_j(\mu|A) &= \Pr\left(A^2 - \mu > \varepsilon_j > A^1 - \mu\right)^{YN} \cdot \Pr\left(\mu + \varepsilon_j > A^2\right)^{YY} \\ &\times \Pr\left(\mu + \varepsilon_j < A^2\right)^{NN} \cdot \Pr\left(A^1 - \mu > \varepsilon_j > A^2 - \mu\right)^{NY}, \end{aligned}$$

where

$$\begin{aligned} YY &= 1 \text{ for a “yes–yes” answer, } 0 \text{ otherwise} \\ NY &= 1 \text{ for a “no–yes” answer, } 0 \text{ otherwise;} \\ YN &= 1 \text{ for a “yes–no” answer, } 0 \text{ otherwise;} \\ NN &= 1 \text{ for a “no–no” answer, } 0 \text{ otherwise;} \end{aligned}$$

We derive inferential statistics for both  $\mu_{\text{SBDC}}$  and  $\mu_{\text{DBDC}}$  by the Delta method. However, since confidence intervals obtained from the Delta method are symmetric around the mean, hence not appropriate (Park et al. 1991), the 95% confidence intervals for mean WTP estimates are constructed by the Krinsky and Robb (1986) Krinsky and Robb (1986) Monte Carlo simulation, which we implement in Stata using the user-written command *wtpcikr* (Jeanty 2007).

When comparing mean WTP from SBDC ( $\mu_{\text{SBDC}}$ ) and mean WTP from DBDC ( $\mu_{\text{DBDC}}$ ), it is expected that the former be larger than the latter due to internal inconsistency. Our primary concern is to investigate whether calibration reduces the gap between mean WTP from the SBDC and mean WTP from DBDC. We address this issue by carrying out a battery of statistical tests. First, we examine whether the use of calibration technique significantly reduces the mean WTP from both SBDC and DBDC. Accordingly, the following two tests are implemented:

$$\begin{aligned} H_0 : \Delta &= \mu_{\text{SBDC}_{\text{NC}}} - \mu_{\text{SBDC}_{\text{C}}} \leq 0 \\ H_1 : \Delta &= \mu_{\text{SBDC}_{\text{NC}}} - \mu_{\text{SBDC}_{\text{C}}} > 0 \end{aligned} \quad (7)$$

where  $\mu_{\text{SBDC}_{\text{NC}}}$  and  $\mu_{\text{SBDC}_{\text{C}}}$  are the mean WTP for the SBDC without calibration and with calibration, respectively.

$$\begin{aligned} H_0 : \Delta &= \mu_{\text{DBDC}_{\text{NC}}} - \mu_{\text{DBDC}_{\text{C}}} \leq 0 \\ H_1 : \Delta &= \mu_{\text{DBDC}_{\text{NC}}} - \mu_{\text{DBDC}_{\text{C}}} > 0 \end{aligned} \quad (8)$$

where  $\mu_{\text{DBDC}_{\text{NC}}}$  and  $\mu_{\text{DBDC}_{\text{C}}}$  are the mean WTP for the DBDC without calibration and with calibration, respectively.

Furthermore, we carry out a one-tailed difference in difference test by testing the following null hypothesis:

$$\begin{aligned} H_0 : \Delta &= (\mu_{\text{SBDC}_{\text{NC}}} - \mu_{\text{DBDC}_{\text{NC}}}) - (\mu_{\text{SBDC}_{\text{C}}} - \mu_{\text{DBDC}_{\text{C}}}) \leq 0 \\ H_1 : \Delta &= (\mu_{\text{SBDC}_{\text{NC}}} - \mu_{\text{DBDC}_{\text{NC}}}) - (\mu_{\text{SBDC}_{\text{C}}} - \mu_{\text{DBDC}_{\text{C}}}) > 0 \end{aligned} \quad (9)$$

Testing the null hypothesis in Eq. (9) entails testing whether the discrepancy between mean WTP from SBDC and mean WTP from DBDC is lower with calibration than without calibration. Performing these tests, particularly the test in Eq. (9), is relatively complex given that the correlation between the first and the second answers renders non-independent the values obtained for the two elicitation questions. Hence, the covariance between WTP measures from the initial and the follow-up questions is different from zero, precluding the use of conventional paired *t* tests.

Bootstrap technique remains an effective way to side-step this issue and to undertake the test (Efron 1993). For both SBDC and DBDC, calibrated and uncalibrated, we simulate mean WTP using 50,000 replications and the results for each simulation are saved to a dataset.<sup>1</sup> We then load and merge the datasets and calculate the difference

---

<sup>1</sup> We use the user-written-wtpcikr-command to perform the Monte Carlo simulations and save the results to datasets (Jeanty 2007).

in mean WTP as in Eqs. (7) to (9). Once the difference in mean WTP is calculated, the achieved significance level (ASL) is also calculated to test the corresponding null hypothesis. The lower the ASL relative to conventional significance levels, the more likely the internally inconsistent patterns in responses are reduced by the use of the double certainty calibration.

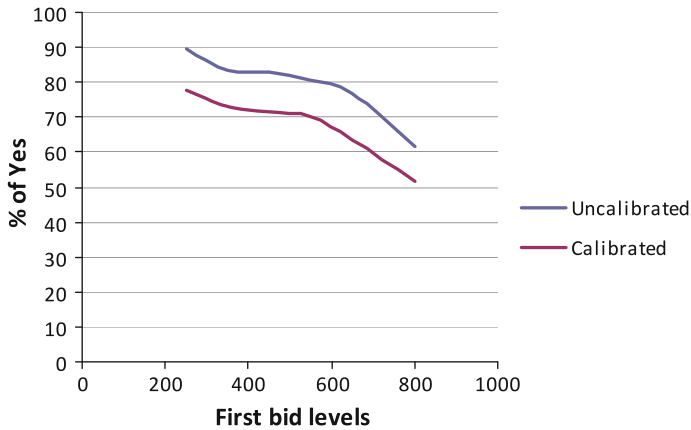
#### 4 Survey design and data

The good being valued in the study is the provision of CBHI to the rural households in Bandjoun, a province located in West of Cameroon. Given that most rural households are excluded from formal insurance, CBHI has emerged as a concept and a strategy to reach the poor living in rural areas with adequate health care services. Organized and managed in a participatory manner, CBHI are small scale, voluntary health insurance programs, (Tabor 2005). CBHI is now adopted in many developing countries (see for instance Dong et al. 2003; Dror et al. 2007; Ataguba et al. 2008; Asenso-Okyere et al. 1997). Recently, policymakers in Cameroon have adopted a health strategic plan to promote CBHI. The plan aims at: (a) putting in place CBHI in each health district by 2015 and (b) covering at least 40% of the population by the CBHI by 2015. Under the micro-insurance innovation facility housed at the International Labor Organization's Social Finance Program and supervised by the European Development Research Network, face-to-face interviews were conducted in 2009 in six villages on a sample of 369 rural household heads selected by a two-stage cluster sampling technique. In an attempt to conduct a state-of-the art CV, guidance provided in Arrow et al. (1993), Carson (2000), and Whittington (2002) were followed.

In order to minimize interviewer bias, we ensured that the numerators were well trained and were randomly assigned to the respondents as done in Abdullah and Jeanty (2011). Another source of concern was hypothetical bias arising from a respondent's stated WTP being lower or higher than her true WTP. In order to avoid hypothetical bias, we took pain in presenting the scenario to the respondents as real as possible by reminding them of their budget constraint and of the possibility of exclusion due to a failure to honor their payment on time. The scenario explained to the respondents the concept of CBHI, the operation of CBHI, the benefits associated to CBHI, and the premium that they would have to pay to receive such benefits.

Focus groups and pre-tests conducted helped determine the following initial bids: 250, 350, 450, 550, 650, and 800 CFA francs. As a standard practice in the DBDC question format, the follow-up bids were twice (respectively half) the initial bids if the respondent answered "yes" (respectively "no") to the first valuation question. Furthermore, the FCQ were included after both the initial and the follow-up bids. The FCQ asked the "yes" respondents to rate on a 10-point numerical likert scale ranging from 1 "very uncertain" to 10 "very certain" how sure they felt that they would actually pay for the CBHI. This self-reported certainty level was used to re-code responses to the WTP question and to provide an estimate of mean WTP similar to the actual WTP. Parallel to Ethier et al. (2000), Poe et al. (2002), Ready et al. (2010), and Morrison and





**Fig. 1** The aggregate demand curve for CBHI

Brown (2009) a threshold 7 out of 10 was set. Then, all “yes” answers were recoded as “no” if the score was strictly inferior to 7.<sup>2</sup>

Figure 1 conspicuously portrays that both the uncalibrated and the calibrated first yes responses give rise to a downward sloping Hicksian demand function, indicating that insurance is a normal good: as the premiums increase, the households are less willing to pay for CBHI.

## 5 Results

We begin with the uncalibrated results. The first panels in Tables 1 and 2 provide uncalibrated mean WTP estimates from SBDC and DBDC. As can be seen, there is downward mean shifting in the second responses ( $\mu_{SBDC} > \mu_{DBDC}$ ). Indeed, there is clear pattern of internal inconsistency in the uncalibrated responses from the respondents. This finding is consistent with previous researches (Hanemann et al. 1991; Cameron and Quiggin 1994; Herriges and Shogren 1996; DeShazo 2002; Bateman et al. 2008).

Following Loomis and Ekstrand (1998), we compare the efficiency gain of the DBDC over the SBDC. The ratio of the confidence interval to the mean WTP is used as a relative measure of efficiency of WTP estimates ( $CI/mean = (Upper\ bound - lower\ bound)/mean\ WTP$ ). The lower the ratio the higher the efficiency. A close look at the mean WTP estimates in Tables 1 and 2 confirms that the ratio of the confidence interval to the mean WTP of DBDC is lower than that of the SBDC ( $0.16 < 0.68$ ),<sup>3</sup> indicating that DBDC yields more efficient welfare estimates than does SBDC.

The first panel in Table 3 shows a difference between  $\mu_{SBDC}$  and  $\mu_{DBDC}$  ( $\Delta = 90.97$ ). Nevertheless, it is possible to use the FCQ to calibrate respondents’

<sup>2</sup> Using the threshold level 8 and above would result in very few yes responses. This same pattern of data was found in the study of Whitehead and Cherry (2007) where the certainty level of 7 was used.

<sup>3</sup> The DBDC yields four times efficiency gains as compared to the SBDC.

**Table 1** Mean willingness to pay for SBDC

Statistic	Value without calibration	Value with calibration
$\mu$ SBDC	1064.95 (140.74)	975.49 (179.41)
Krinsky–Robb <sup>a</sup>	[875.43 1598.91]	[755.43 2056.21]
CI/Mean	0.68	1.33

<sup>a</sup> Confidence interval of the mean WTP is obtained by Monte Carlo simulations on 50,000 draws. The standard errors are in brackets computed by the Delta method

**Table 2** Mean willingness to pay for DBDC

Statistic	Value without calibration	Value with calibration
$\mu$ DBDC	973.98 (38.13)	935.72 (94.07)
Krinsky–Robb	[ 901.70 1052.87]	[ 753.49 1127.12]
CI/Mean	0.16	0.40

WTP and also reduce the discrepancy between the mean WTP calculated from the SBDC and DBDC and maintain the efficiency gain as well.

In fact, as it is conspicuous in Table 3, the use of double certainty calibration reduces the discrepancy between the mean WTP from the SBDC and DBDC by about 60% ( $\Delta = 39.77$ ). Furthermore, for both SBDC and DBDC, we test whether calibration makes a difference. As shown in Table 4, the null hypothesis is rejected at the one percent significant level ( $p$  value is 0.003 and 0.005 for SDBDC and DBDC, respectively). In other words, mean WTP from SBDC and mean WTP from DBDC are both reduced when calibration is applied. In fact, before the double calibration, the mean WTP of SBDC and DBDC were 1064.95 CFA francs and 973.98 CFA, respectively; while these means become 975.49 CFA francs and 935.72 CFA francs, respectively, when the calibration is implemented.

Finally, Table 3 clearly shows that the discrepancy between mean WTP estimate from SBDC and mean WTP estimate from DBDC is smaller with calibration (39.77) than without (90.97). We test whether the difference is statistically lower with calibration. As shown in Table 4, the test resoundingly rejects the null hypothesis that the discrepancy is higher with calibration ( $p$  value=0.005). This result substantiates our claim that the double certainty calibration is effective at reducing the internal inconsistency in the DBDC approach.

Our results also suggest that even with calibration, the DBDC approach yields more efficient WTP estimates. For instance, in Tables 1 and 2, the ratio of the confidence interval to the mean WTP of DBDC is lower than that of SBDC ( $0.40 < 1.33$ ). In other words, confidence intervals around the mean WTP estimates of DBDC are still tighter than the one around the mean WTP estimates of SBDC. In theory, calibrating the responses of respondents must not affect the efficiency gain of the DBDC over SBDC though the central tendency could be affected. As Alberini et al. (2003) argued, there is no reason to believe that allowing uncertain responses will affect the efficiency of welfare estimates. Our finding echoes this contention.

**Table 3** Difference between mean WTP for SBDC and DBDC

Statistic	Value without calibration	Value with calibration
$\mu_{SBDC}$	1064.95 (140.74)	975.49 (179.41)
$\mu_{DBDC}$	973.98 (38.13)	935.72 (94.07)
$\Delta = \mu_{SBDC} - \mu_{DBDC}$	90.97	39.77

**Table 4** The achieved significance level of the test of internal inconsistency

Test	<i>P</i> value
$H_0 : \Delta = \mu_{SBDC_{NC}} - \mu_{SBDC_C} \leq 0$	0.003
$H_1 : \Delta = \mu_{SBDC_{NC}} - \mu_{SBDC_C} > 0$	
$H_0 : \Delta = \mu_{DBDC_{NC}} - \mu_{DBDC_C} \leq 0$	0.005
$H_1 : \Delta = \mu_{DBDC_{NC}} - \mu_{DBDC_C} > 0$	
$H_0 : \Delta = \mu_{SBDC_{NC}} - \mu_{DBDC_{NC}} - \mu_{SBDC_C} - \mu_{DBDC_C} \leq 0$	0.005
$H_1 : \Delta = \mu_{SBDC_{NC}} - \mu_{DBDC_{NC}} - \mu_{SBDC_C} - \mu_{DBDC_C} > 0$	

## 6 Conclusions

The use of CV method in the health sector is gaining popularity as policymakers may capitalize on results from the CV survey to improve the well-being of their populations. Over the past decades, there has been a shift from SBDC to DBDC because of the statistical efficiency gains from the DBDC. Nevertheless, the DBDC format has been criticized on the ground that responses from the first question are inconsistent with those from the second question, causing a downward mean shifting in WTP estimates derived from the latter. Empirical evidence from previous studies has confirmed this internal inconsistency in DBDC.

Using a CBHI CV survey, the purpose of this paper was to investigate whether applying certainty calibration to both initial and follow-up bids reduces this internal inconsistency. To the best of our knowledge, this study was the first to use certainty calibration to contend with the issue of internal inconsistency in the DBDC approach.

Consistent with previous CV studies, we find a substantial difference between uncalibrated SBDC and DBDC WTP estimates. Using parametric bootstrap technique, we also show that, for both SBDC and DBDC, calibrated and uncalibrated mean WTP estimates are statistically different with the former being lower than the latter. Finally, considering the most salient focus of this study, we provide statistical evidence supporting the contention that calibration is effective at reducing internal inconsistency between the single- and double-bounded welfare estimates in a DBDC contingent valuation study. In our application, the internal inconsistency is reduced by as much as 60%. Furthermore, the double certainty calibration does not affect the efficiency gain

arising from the second question. This fundamental result implies that researchers wanting to explore the double certainty calibration to mitigate the internal inconsistency in future CV studies will not need to worry about the efficiency gain inherent in applying the DBDC approach.

**Acknowledgments** We are thankful to the two anonymous referees for their useful comments and suggestions. A grant from the International Labor Organization (40052113/0) is also gratefully acknowledged. We will also like to thank the African Doctoral Dissertation Research Fellowship offered by the African Population and Health Research Center (APHRC) in partnership with the International Development Research Centre (IDRC) for financial assistance. Any errors or omissions remain our sole responsibility.

## References

- Abdullah S, Jeanty PW (2011) Willingness to pay for renewable energy: evidence from a contingent valuation survey in Kenya. *Renew Sustain Energy Rev* 15(6):2974–2983
- Alberini A, Boyle K, Welsh M (2003) Analysis of contingent valuation data with multiple bids and response options allowing respondents to express uncertainty. *J Environ Econ Manag* 45(1):40–62
- Arrow K, Solow PR, Leamer EE, Radner R, Shuman H (1993) Report of NOAA panel on contingent valuation method. *Fed Regist* 58(10):4601–4614
- Asenso-Okyere WK, Osei-Akoto I, Anum A, Appiah EN (1997) Willingness to pay for health insurance in a developing economy. A pilot study of the informal sector of Ghana using contingent valuation. *Health Policy* 42(3):223–237
- Ataguba J, Ichoku EH, Fonta W (2008) Estimating the willingness to pay for community healthcare insurance in rural Nigeria. *Poverty and Economic Policy*, Dakar
- Bateman I, Burgess GD, Hutchinson WG, Matthews DI (2008) Learning design contingent valuation (LDCV): NOAA guidelines, preference learning and coherent arbitrariness. *J Environ Econ Manag* 55(2):127–141
- Bateman JJ, Langford IH, Jones AP, Geoffrey NK (2001) Bound and path effects in double and triple bounded dichotomous choice contingent valuation. *Resour Energy Econ* 23(3):191–213
- Bishop RC, Heberlein T (1979) Measuring the values of extra-market goods: are indirect measures biased? *Am J Agric Econ* 61(5):926–930
- Burton AC, Carson KS, Chilton SM, Hutchinson WG (2003) An experimental investigation of explanation for inconsistencies in responses to second offers in double referenda. *J Environ Econ Manag* 46(3):472–489
- Cameron TA, Quiggin J (1994) Estimation using contingent valuation data from a “dichotomous choice with follow-up” questionnaire. *J Environ Manag* 27(3):218–234
- Carson RT (2000) Contingent valuation: a user’s guide. *Environ Sci Technol* 34(8):1413–1418
- Carson RT, Hanemann M, Mitchell RC (1986) Determining the demand for public goods by simulating referendums at different tax prices. Department of Economics, University of California, San Diego
- Carson RT, Wilks L, Imber D (1994) Valuing the preservation of Australia’s kakadu conservation zone. *Oxford Econ Pap* 46:727–749
- Champ PA, Bishop RC, Brown TC, McCollum DW (1997) Using donation mechanisms to value nonuse benefits from public goods. *J Environ Econ Manag* 33(2):151–162
- Cooper J, Hanemann MW, Signorelli G (2002) One and one-half bids for contingent valuation. *Rev Econ Stat* 84(4):742–750
- DeShazo JR (2002) Designing transactions without framing effects in iterative question formats. *J Environ Econ Manag* 43(3):360–385
- Dong H, Kouyate B, Cairns J, Mugisha F, Sauerborn R (2003) Willingness-to-pay for community-based insurance in Burkina Faso. *Health Econ* 12(10):849–862
- Dror M, Radermacher R, Koren R (2007) Willingness to pay for health insurance among rural and poor persons: field evidence from seven micro health insurance units in India. *Health Policy* 82(1):12–27
- Efron T (1993) An introduction to the bootstrap. Chapman and Hall, New York
- Ethier RG, Poe GL, Schulze WD, Clark JE (2000) A comparison of hypothetical phone and mail contingent valuation responses for green pricing electricity programs. *Land Econ* 76(1):54–67

- Flachaire E, Hollard G (2007) Starting point bias and respondent uncertainty in dichotomous choice contingent valuation surveys. *Resour Energy Econ* 29(3):183–194
- Guzman RM, Kolstad CD (2007) Researching preferences, valuation and hypothetical bias. *Environ Resour Econ* 37(3):465–487
- Hanemann MW (1984) Welfare evaluations in contingent valuation experiments with discrete response. *Am J Agric Econ* 66(3):332–341
- Hanemann M, Loomis WJ, Kanninen B (1991) Statistical efficiency of double-bounded dichotomous choice contingent valuation. *Am J Agric Econ* 73(4):1255–1263
- Hanley N, Kristrom B, Shogren JF (2009) Coherent arbitrariness: on value uncertainty for environmental goods. *Land Econ* 85(1):41–50
- Herriges JA, Shogren JF (1996) Starting point bias in dichotomous choice valuation with follow-up questioning. *J Environ Econ Manag* 30(1):112–131
- Jeanty PW (2007) Constructing krinsky and robb confidence intervals for mean and median willingness to pay (wtp) using stata. Sixth North American Stata Users' Group Meeting, Boston, August 13–14
- Kanninen BJ (1995) Bias in discrete response contingent valuation. *J Environ Econ Manag* 28(1):144–125
- Krinsky I, Robb AL (1986) On approximating the statistical properties of elasticities. *Rev Econ Stat* 68(4):715–719
- Loomis J, Ekstrand E (1998) Alternative approaches for incorporating respondent uncertainty when estimating willingness to pay: the case of the mexican spotted owl. *Ecol Econ* 27(1):29–41
- Mahieu PA, Riera P, Giergiczy M (2012) The influence of cheap talk on willingness-to-pay ranges: some empirical evidence from a contingent valuation study. *J Environ Planing Manag* 55(6):753–763
- McFadden D (1994) Contingent valuation and social choice. *Am J Agric Econ* 76(4):689–708
- Mitchell RC, Carson RT (1989) Using surveys to value public goods: the contingent valuation method. Resource for the future, Washington
- Morrison M, Brown TC (2009) Testing the effectiveness of certainty scales, cheap talk, and dissonance-minimization in reducing hypothetical bias in contingent valuation studies. *Environ Resour Econ* 44(3):307–326
- Murphy JJ, Stevens TH, Weatherhead D (2005) Is cheap talk effective at eliminating hypothetical bias in a provision point mechanism? *Environ Resour Econ* 30(3):327–343
- Park T, Loomis JB, Creel M (1991) Confidence interval for evaluating benefits estimates from dichotomous choice contingent valuation studies. *Land Econ* 67(1):64–73
- Poe GL, Clark JE, Rondeau D, Schulze WD (2002) Provision point mechanisms and field validity tests of contingent valuation. *Environ Resour Econ* 23(1):105–131
- Ready RC, Champ PA, Lawton JL (2010) Using respondent uncertainty to mitigate hypothetical bias in a stated choice experiment. *Land Econ* 86(2):363–381
- Tabor SR (2005) Community-based health insurance and social protection policy. Social protection discussion paper series, World Bank, Washington
- Watson V, Ryan M (2007) Exploring preference anomalies in double bounded contingent valuation. *J Health Econ* 26(3):463–482
- Whitehead JC, Cherry TL (2007) Willingness to pay for a green energy program: a comparison of ex-ante and ex-post hypothetical bias mitigation approaches. *Resour Energy Econ* 29(4):247–261
- Whittington D (2002) Improving the performance of contingent valuation studies in developing countries. *Environ Resour Econ* 22(1):323–367