Does don't know mean no? Analysis of 'don't know' responses in dichotomous choice contingent valuation questions

Peter A. Groothuis and John C. Whitehead

ABSTRACT

The 'don't know' response option in contingent valuation dichotomous choice questions is analysed using data from both willingness to pay and willingness to accept studies. An empirical analysis is conducted to determine whether respondents are stating a response similar to yes or no responses or a middle response. It is found that don't know responses are similar to no responses in the willingness to pay study. In the willingness to accept study, it is found that the `don't know' responses are similar to a middle response. It is further suggested that researchers consider calculating ambivalence bounds when a don't know response is a middle response.

I. INTRODUCTION

Dichotomous choice contingent valuation (CV) questions have been increasingly used since their introduction by Bishop and Heberlein (1979). Mitchell and Carson (1989) and the NOAA Panel on CV (1993) have recommended the referendum format for value elicitation questions. Dichotomous choice questions typically request a yes or no response to a bid value. For example, a stylized willingness to pay (accept) question asks: `Would you be willing to pay (accept) $A for the change in resource allocations (described elsewhere)? [YES or NO]' Bid values ($A) are varied across respondents and discrete choice econometric techniques are used to estimate willingness to pay and willingness to accept values. More controversially, the NOAA panel recommended including a middle response in dichotomous choice CV questions. A middle response might take the form of `don't know', `undecided', or `uncertain' options in addition to the standard `yes' and `no' answer categories.

The benefits of offering a middle response option are that uncertain survey respondents are not forced to construct a willingness to pay (WTP) or willingness to accept (WTA) value when answering a dichotomous choice question. Only those respondents who have previously formed values are left in the sample increasing the validity of `yes' and `no' answers. The cost of offering middle responses is a reduction in sample size and econometric efficiency. But is dropping the middle response the best procedure? Further, if middle responses are not dropped from the empirical analysis, what information do they provide and how should they be handled empirically? The analysis examines the effects of including a middle response in both WTP and WTA dichotomous choice studies.

II. BACKGROUND

Survey researchers disagree on whether a middle response should be offered. Payne (1951) says middle responses should be offered if you wish to elicit only the stronger opinions on the issue. However, if you want to know which way people are leaning then the middle response should not be offered. Presser and Schuman (1980) provide evidence supporting this suggestion by performing several middle response experiments to determine if there is any effect on nonresponse, effects on the proportion of yes/no responses, or any effects on the relationships among variables. Presser and Schuman (1980) find that omitting a middle response will increase item nonresponse since those who have true don't know responses or are uncertain will skip the question, not affecting the proportion of yes/ no responses.

In related telephone survey research, Bishop (1988) finds that the proportion of yes and no responses changes for some questions in attitude surveys when middle responses are offered. He suggests that middle responses reflect ambivalence about the choice. In a mail survey Ayidiya and McClendon (1990) find results similar to Bishop. Poe et al. (1988) also find that results can change if don't know responses are offered in mailed questionnaires. They further argue that survey instruments without the don't know option are superior because a higher percentage of substantive responses are received, relative to surveys which offer a don't know option.

In CV research, Milon (1989) offers a middle response in a mail survey and uses follow up questions to determine if the middle response is a `no' or a protest response. Milon suggests that

respondents are more likely to be uncertain, and give the middle response, as the `stakes-at-risk' related to the choice increases. Johannesson et al. (1993) assume that a don't know response implies indifference and discard these answers from their empirical analysis. Opaluch and Segerson (1989) argue that a CV question creates ambivalence due to the unfamiliarity of the topic. This ambivalence leads the respondent to answer with the `don't know' response. In an application of this theory Ready et al. (1995) find that the proportion of yes/no responses changes if multiple middle responses are included.

More recent research suggests two competing hypotheses about the information contained in don't know responses. Wang (1997) suggests that individuals have a valuation distribution or a range of values as opposed to a single value. He argues that if the offer falls in the middle range of the valuation distribution respondents will answer with the middle response. Wang (1997) develops an empirical model for don't know responses that is conceptually similar to ordered logistic regression. Carson et al. (1998) argue that don't know responses are similar to no responses and empirically test the hypothesis with the multinomial logit model. When don't know and no responses contain the same essential information, Carson et al. argue that researchers should follow a conservative approach and code don't know as no responses when calculating WTP. These competing hypotheses are considered and how researchers should use the middle response option in CV analysis is addressed.

III. THE DATA AND MODELS

Two data sets are employed. The first is a WTA study of the compensation required to site a hazardous waste landfill in a rural Pennsylvania County (Groothuis and Miller, 1994; Groothuis et al. 1998). The second is a WTP study for a sea turtle protection programme in North Carolina (Whitehead, 1992; 1993).

The sampling frame for the Pennsylvania study is Lawrence County residents. Lawrence County, a rural Western Pennsylvania area, is located 40 miles north of Pittsburgh and has New Castle as its largest city. By virtue of its rural nature and proximity to major industrial centres, it has been targeted several times as a potential site for a hazardous waste disposal facility. A random sample of households was drawn from the county telephone directory. The mail survey was conducted following Dillman (1978) during the spring of 1992. A 41% response rate was obtained.

Survey respondents were presented with information about hazardous waste and a policy proposal concerning the location of the hazardous waste landfill. The WTA question was framed as a referendum where respondents were offered a tax reduction of $A (either 100, 500, 1000, 1500, or 2000) to accept a hazardous waste facility in their county. The value elicitation question was:

```
   Suppose the State proposes to locate the hazardous waste landfill in your
   county. In return the State proposes to compensate people by reducing
State
   income tax by $A per family in your county per year. Would you be willing
   to accept this proposal?
```

Respondents were then given three answer choices: YES, NO or DON'T KNOW. To analyse the response, a linear approximation in the change in the indirect utility function is specified.

(1) $Dv = a_0 + a_1 A + a_2(p1 - q1) + a_3 Y$

where A is the amount of tax reduction, q1 is the probability of good health without a hazardous waste landfill, p1 is the probability of good health with a hazardous waste landfill and Y is income. The probabilities are the subjective answers to risk related questions in the survey. The expected signs are $a_1 > 0$, $a_2 > 0$, and $a_3 < 0$. Respondents are expected to be more likely to respond yes with a higher offer amount and with a higher change in the health risk probability. Respondents are expected to be less likely to respond yes as income increases (Groothuis et al. 1998).

The sampling frame for the North Carolina study is North Carolina households. A random sample was drawn from telephone directories of North Carolina cities and rural areas. The mail survey was conducted following Dillman (1978) during the winter of 1991. A response rate of 35% was achieved.

Survey respondents were informed about a programme designed to manage loggerhead sea turtle nesting habitat. Following the description of a sea turtle protection programme, respondents were presented with a value elicitation question and asked if they would be willing to donate a dollar amount $A (either 1, 5, 10, 25, 50 or 100) to preserve loggerhead sea turtles:

```
    Suppose that a $A contribution from each North Carolina household each
year
    would be needed to support and fund the loggerhead sea turtle program.
    Would you be willing to contribute $A each year to the `Loggerhead Sea
    Turtle Preservation Trust Fund' in order to support the loggerhead sea
    turtle program?
```

Respondents were then given three answer choices: YES, DON'T KNOW or NO. To analyse the WTP responses, we specify a linear approximation in the change in indirect utility function

(2) $Dv = b_0 + b_1 A + b_2[p2(q2 - r2)] + b_3 Y$

where p2 is the probability of demanding trips to observe sea turtle habitat, q2 is the probability of extinction without a management programme, r2 is the probability of extinction with a management programme, and Y is income. The probabilities are the subjective answers to risk related questions in the survey. The expected signs are $b_1 < 0$, $b_2 2> 0$, and $b_3 > 0$. Respondents are expected to be less likely to respond yes as the dollar amount requested increases. Respondents are more likely to respond yes as the extinction risk, weighted by the demand probability, increases and as income increases (Whitehead 1993).

Means of the variables are reported in Table 1. Fifty-five per cent of the Pennsylvania sample were not willing to accept the offer. Sixteen per cent were willing to accept the offer and 28% responded don't know. The most important reason (55%) for the don't know response was `I need more information' indicating that respondents have uncertainty in preferences. The second most

important response (40%) was `I am not sure how the tax cut or the hazardous waste facility affects my household' again indicating that respondents are uncertain. Five per cent of the respondents answered `I disagree with the question' indicating a protest against the compensation scheme. Thus 95% of the `don't know' respondents in the Pennsylvania sample answered because of uncertainty.

For the North Carolina sample the most common answer to the valuation question was don't know (36%). Thirty-one per cent (32%) of respondents answered yes (no). The most important reason (37%) given for the `don't know' response was `I need more information' suggesting that the dichotomous choice question created uncertainty. The second most important response (30%) was `I can't afford to contribute to the Trust Fund' suggesting that if forced to make a choice, the answer would be no. These respondents may be showing ambivalence: wanting the improved environmental quality but not wanting the loss in income. Fourteen per cent gave reasons that suggested they valued sea turtle protection. Thus for the North Carolina sample no clear reason for choosing don't know was found.

IV. EMPIRICAL RESULTS

The probability of answering one of three response categories is estimated with the multinomial logit and ordered logit models (Greene, 1997). For the Pennsylvania sample, the initial assumption is that the `don't know' response is most closely related to a yes response. For the North Carolina sample, the initial assumption is that the don't know response is most closely related to a no response. These assumptions result in more conservative WTA and WTP welfare estimates. The multinomial logit model is estimated first with the coefficients for respondents who answered don't know and the more conservative response constrained to be equal. Then the coefficients are allowed to vary in a second multinomial logit model. Next, the response categories is arranged so that the don't know response is assumed to be a middle response and an ordered logistic regression model is run.

For the Pennsylvania sample (Table 2) the effects of the variables in each model are jointly significantly different from zero, according to the model [chi square] statistic. All of the coefficient estimates are of predicted sign and all but the income coefficients are significantly different from zero at conventional levels. The coefficient on income is only significant in the multinomial logit don't know coefficient vector.

The likelihood ratio (LR) specification test is used to determine if the constrained or unconstrained model is most appropriate. It is found that the unconstrained model yields a higher model [chi square] value and the LR test is significant at the 0.01 level. The model [chi square] statistic in the ordered logit model is also significantly different from zero and close to the unconstrained multinomial logit model. From these tests we conclude that respondents who state don't know are indicating a middle response reflecting uncertainty or ambivalence. The don't know response is not similar to the yes response. We speculate that if respondents were forced to choose between yes or no with the WTA question, they would likely choose to protest or not to answer.

In Table 3, we report the results of the WTP specification for the sea turtle preservation programme. The effects of all of the variables in each model are jointly significantly different from zero, according to the [chi square] statistic. Also, all of the coefficient estimates are of predicted sign and all but income in the unconstrained multinomial logit for don't know respondents are significantly different from zero. It is found that the unconstrained multinomial logit and the ordered logit model yield higher model [chi square] values than the constrained multinomial logit model. But the LR test indicates that this difference is insignificantly different from zero. From these tests it is concluded that respondents who state don't know are most closely indicating a no response. If respondents were forced to choose between yes or no with the WTP question, they would be more likely to choose the no response.

V. WTP AND WTA ESTIMATES

Of primary interest to CV researchers is obtaining estimates of WTA and WTP. Using the above specifications, we can calculate a point estimate of both the WTA and WTP. The point estimate for WTA and WTP are

(3) WTA = -[[a.sub.0] + [a.sub.2](p1 - q1) + [a.sub.3] Y]/[a.sub.1]

(4) WTP = -[[b.sub.0] + [b.sub.2][p2(q2 - r2)] + [b.sub.3] Y]/[b.sub.1]

where mean income and probability levels are used. Willingness to pay estimates and standard errors are calculated with LIMDEP (Greene, 1995). With the multinomial logit model we find a WTA estimate of $2663 when don't know and yes responses are constrained to be equal (Table 2). When don't know and yes responses are unconstrained in the multinomial logit model, WTA falls to $2390 for don't know respondents and $2903 for yes respondents. These estimates are not statistically different from each other, however.

A WTP estimate of $23 is calculated when the don't know and no responses are constrained to be equal in the multinomial logit model (Table 3). The WTP estimate is $29 when the respondents who answer `no' are compared to the yes respondents and $15 when the don't know respondents are compared to the yes respondents. The $15 estimate is not statistically different from zero.

If don't know responses are determined to be middle responses researchers could estimate ambivalence bounds using the two thresholds that ordered logit provides. The lower threshold indicates a shift from a no response to ambivalence (don't know) and the upper threshold is the break between ambivalence and a yes response. The lower bound for the WTA value is identical to Equation 3 with the ordered logit coefficients. The upper bound estimate uses the [mu] statistic in the calculation

(5) WTA = -[[a.sub.0] - [micro] + [a.sub.2](p1 - q1) + [a.sub.3] Y]/[a.sub.1]

where ([a.sub.0] - [micro]) is the upper threshold constant term. The ambivalence upper and lower bounds are $1370 and $4387.

To calculate the ambivalence bounds for the WTP study, Equation 4 is used with the ordered logit coefficients for the upper bound and

(6) WTP = -[[b.sub.0] - [micro] + [b.sub.2][p2(q2 - r2)] + [b.sub.3] Y]/[b.sub.1]

for the lower bound. The estimates of WTP are -$23 and $93. The negative WTP is not statistically different from zero.

VI. CONCLUSIONS

Should CV researchers include a don't know response option? The analysis does not provide a clear answer. For instance, if researchers are interested in use values then respondents may have already constructed a value so that the don't know responses are unnecessary. Conversely, if researchers are interested in passive use (nonuse) values then unfamiliar respondents may not have previously constructed a WTP or WTA value to compare to an offer. In this case offering a `don't know' option would avoid a large amount of protest responses or non-response.

If a `don't know' option is offered how should researchers use the answer? Two competing hypotheses exist in the literature. The first is that the don't response is similar to the most conservative discrete response. If forced to make a choice, respondents will resort to the conservative response (e.g., no (yes) in a WTP (WTA) study). The results suggest that this is the case with the WTP study. The other hypothesis is that a don't know response is a middle response. The results suggest this is the case with the WTA study. When don't know responses are found to be middle responses, it is suggested that researchers estimate ambivalence regions by using the thresholds in ordered logit analysis. These regions provide insights that a point estimate alone ignores.

How do researchers determine what the choice of a middle response reveals about preferences? It is suggested that researchers rely on follow up questions and multiple choice regression methods to guide their decision on how to interpret a don't know response. When the constrained model and unconstrained multinomial model yield similar results conclude that the don't know response reflects the conservative response. When ordered logit provides the best fit conclude that the don't know response reflects an ambivalent response. Further, it is not suggested that researchers should drop the don't know responses from the analysis. The lost information about WTP and WTA may be misleading. CV researchers with multiple response data should use all the information available to them. Researchers must use their judgement on how to interpret and use don't know responses in CV analysis.

Table 1. Data summary

Variable	Description	Mean	S.D.
Pennsylvania Sample			
YES	YES, willing to accept SA	0.16	
NO	NO, not willing to accept SA	0.55	
DK	Don't Know	0.28	
A	Dollar Amount	1001.25	658.37
p1-q1	Probability Change	-0.71	0.29
Y	Income in thousands	30.37	16.01
North Carolina Sample			
YES	YES, willing to pay SA	0.31	
NO	NO, not willing to pay SA	0.32	
DK	Don't Know	0.36	
A	Dollar Amount	32.31	34.43
p2(q2-r2)	Probability Change	0.25	0.22
Y	Income in thousands	29.07	15.37

Table 2. WTA Logit models

Multinomial Logit

Constrained

DK, YES

Variable	Coefficient	b/St. Er.
Constant	1.031	2.02
A	0.001	2.93
p1-q1	2.848	6.10
Y	-0.016	-2.01
mu		
WTP	$2663.20	4.46
Upper Bound WTP		
Lower Bound WTP		
Beginning LL	-313.14	
Ending LL	-290.73	
Chi-squared	44.81	
% Correct Predictions	58.31	
LR Test		

Multinomial Logit

Unconstrained

| Variable | DK | | YES | |
	Coefficient	b/St. Er.	Coefficient	b/St. Er.
Constant	1.106	1.90	0.796	1.19
A	0.004	2.18	0.001	2.88
p1-q1	2.105	4.09	4.320	6.60
Y	-0.022	-2.49	-0.004	-0.32

mu				
WTP	$2389.50	3.42	$2903.20	4.47
Upper Bound WTP				
Lower Bound WTP				
Beginning LL		-313.14		
Ending LL		-272.22		
Chi-squared		75.83		
% Correct Predictions		62.07		
LR Test		31.03	[4 d.f.]	

Ordered Logit

No = 2, DK = 1,
Yes = 0

Variable	Coefficient	b/St.Er.
Constant	1.616	3.50
A	0.001	3.17
p1-q1	3.001	7.22
Y	-0.008	-1.11
mu	1.70	10.06
WTP		
Upper Bound WTP	$4387.00	4.12
Lower Bound WTP	$1369.70	5.72
Beginning LL	-313.14	
Ending LL	-279.00	
Chi-squared	68.27	
% Correct Predictions	60.82	
LR Test	23.46 [4 d.f.]	

Table 3. WTP Logit models

Multinomial Logit

Constrained

DK, YES

Variable	Coefficient	b/St. Er.
Constant	0.812	1.83
A	0.019	3.27
p2-(q2-r2)	-2.789	-3.73
Y	-0.019	-1.74
mu		
WTP	$22.61	2.46
Upper Bound WTP		
Lower Bound WTP		
Beginning LL	-221.56	
Ending LL	-206.34	
Chi-squared	30.44	
% Correct Predictions	47.03	
LR Test		

```
                    Multinomial Logit

                    Unconstrained

                    DK                          YES

Variable            Coefficient   b/St. Er.   Coefficient   b/St. Er.

Constant                0.798        1.64         0.854        1.68
A                       0.017        2.61         0.023        3.52
p2-(q2-r2)             -2.450       -2.97        -3.276       -3.47
Y                      -0.015       -1.27        -0.024       -1.89
mu
WTP                    $15.42        1.33        $29.37        3.34
Upper Bound WTP
Lower Bound WTP
Beginning LL                       -221.56
Ending LL                          -204.69
Chi-squared                          33.74
% Correct Predictions                46.04
LR Test                       330         [4 d.f.]

                    Ordered Logit

                    No = 2, DK = 1,
                    Yes = 0

Variable            Coefficient   b/St. Er.

Constant                1.483        3.94
A                       0.015        3.73
p2-(q2-r2)             -2.600       -4.35
Y                      -0.017       -1.93
mu                      1.72         9.63
WTP
Upper Bound WTP        $93.12        4.96
Lower Bound WTP       -$22.64       -1.31
Beginning LL                -221.56
Ending LL                   -205.32
Chi-squared                   32.49
% Correct Predictions         48.51
LR Test                 2.05    [4 d.f.]
```

REFERENCES

Ayidiya, S. A. and McLendon McKee, J. (1990) Response effects in mail surveys, Public Opinion Quarterly, 54, 229-47.

Bishop, G. F. (1988) Experiments with middle response alternative in survey questions, Public Opinion Quarterly, 51, 220-32.

Bishop, R. C. and Heberlein, T. A. (1979) Measuring values of extramarket goods: Are indirect measures biased? American Journal of Agricultural Economics, 61(5), 926-30.

Carson, R. T., Hanemann, W. M., Kopp, R. J., Krosnick, J. A., Mitchell, R. C., Presser, S., Ruud, P. A. and Kerry Smith, V. with Conaway, M. and Martin, K. (1998) Referendum design and contingent valuation: The NOAA panel's No-Vote recommendation, Review of Economies and Statistics, 80, 335-8.

Dillman, D. A. (1978) Mail and telephone surveys: The total design method, John Wiley, New York, NY.

Green, W. (1995) LIMDEP: User's Manual, Econometric Software, Bellport, NY.

Greene, W. H. (1997) Econometric Analysis, Prentice Hall, 3rd Edition.

Groothuis, P. A., Van Houtven, G. and Whitehead, J. C. (1998) Using contingent valuation to measure the compensation required to gain community acceptance of a LULU: The case of a hazardous waste disposal facility, Public Finance Review, 26, 231-49.

Groothuis, P. A. and Miller, G. (1994) Locating hazardous waste facilities: The influence of NIMBY beliefs, American Journal of Economics and Sociology, 53, 335-46.

Johannesson, M., Johansson, P-O., Kristrom, B. and Gredthan, U. G. (1993) Willingness to pay for antihypertensive therapy--further results, Journal of Health Economies, 12, 95-108.

Milon, J. W. (1989) Contingent valuation experiments for strategic behavior, Journal of Environmental Economics and Management, 17, 292-308.

Mitchell, R. and Carson, R. (1989) Using Surveys to Value Public Goods: The Contingent Valuation Method, Resources for the Future: Washington, DC.

National Oceanic and Atmospheric Administration (1989) Appendix I-Report on the NOAA Panel on Contingent Valuation, Federal Register, 58, 4602-14.

Opaluch, J. J. and Segerson, K. (1989) Rational roots of `Irrational' behavior: New theories of economic decision making, Northeastern Journal of Agricultural and Resource Economics, 18, 81-95.

Payne, S. L. (1951) The Art of Asking Questions, Princeton: Princeton University Press.

Poe, G. S., Seeman, I., McLaughlin, J., Mehl, E. and Dietz, M. (1988) `Don't know' boxes in factual questions in mail questionnaire, Public Opinion Quarterly, 52, 212-22.

Presser, S. and Schuman, H. (1980) The measurement of a middle position in attitude surveys, Public Opinion Quarterly, 44, 108-23.

Ready, R., Whitehead, J. C. and Blomquist, G. C. (1995) Contingent valuation when respondents are ambivalent, Journal of Environment Economies and Management, 29, 181-96.

Wang, H. (1997) Treatment of `Don't know' responses in contingent valuation surveys: A random valuation model, Journal of Environmental Economics and Management, 32, 219-32.

Whitehead, J. C. (1992) Ex ante willingness to pay with supply and demand uncertainty: Implication for valuing a sea turtle protection program, Applied Economics, 24, 981-8.

Whitehead, J. C. (1993) Total economic values for coastal and marine wildlife: Specification, validity and valuation issues, Marine Resource Economics, 8, 119-32.