

Anchoring bias in face-to-face Time-Trade-Off valuations of health states

Sesgo de anclaje en valoraciones de estados de salud por Time Trade-Off

Mario García-Molina and Liliana A. Chicaiza-Becerra

Received november 7th 2016 / Sent for modification january 13th 2017 / Accepted june 8th 2017

ABSTRACT

MG: Econ. M. Sc. M. Phil, Ph. D. Universidad Nacional de Colombia. Bogotá, Colombia. mgarciamo@unal.edu.co
LC: BA in Business Administration. Ph. D. Universidad Nacional de Colombia. Bogotá, Colombia. lachicaizab@unal.edu.co

Objective To test whether anchoring (a cognitive bias) occurs during face-to-face interviews to value health states by Time-Trade-Off.

Methods 147 Colombian subjects (111 males, 36 females) valued five EQ-5D health states better than death during a face-to-face interview. Subjects were randomly assigned to two different starting points.

Results Shapiro-Wilk test discarded normality, while non-parametric tests, including Kolmogorov-Smirnov and Wilcoxon-Mann-Whitney, showed that anchoring was significant in four out of five health states. A higher starting point increased the elicited value by 15 %-188 %. The size of the anchoring effect was not uniform among health states.

Conclusion Anchoring effects may bias face-to-face Time Trade-Off valuations. The size of the anchoring effect is relevant enough for health policy.

Key Words: Heuristics; cost-benefit analysis; outcome assessment (health care); observer variation; bias; methods (*source: MeSH, NLM*).

RESUMEN

Objetivos Comprobar si se presenta anclaje (un sesgo cognitivo) en las valoraciones de estados de salud por Time-Trade-Off realizadas en entrevistas en persona.

Métodos 147 sujetos colombianos (111 hombres, 36 mujeres) valoraron five estados de salud EQ-5D mejores que la muerte en una entrevista personal. Los sujetos se asignaron aleatoriamente a dos puntos de partida distintos.

Resultados El test de Shapiro-Wilk descartó normalidad de las distribuciones. Las pruebas no paramétricas de Kolmogorov-Smirnov y Wilcoxon-Mann-Whitney mostraron que el anclaje era significativo en cuatro de los cinco estados de salud. Un punto de partida más alto aumentaba el valor obtenido en un 15 %-188 %. El tamaño del anclaje no era uniforme en los distintos estados de salud.

Conclusiones El anclaje puede sesgar las valoraciones por Time Trade-Off realizadas en entrevistas. El efecto es lo suficientemente alto como para hacerlo relevante en términos de política.

Palabras Clave: Heurística; análisis costo-beneficio; evaluación de resultado; variaciones dependientes del observador; sesgo; métodos (*fuentes: DeCS, BIREME*).

Time-Trade-Off (TTO) is one of the main methodologies used for eliciting health-state utilities to calculate quality-adjusted life-years (QALYs) in health technology assessments. However, there are major differences in the results of TTO valuations (1), as well as in the implementation methods (2). Differences between TTO valuations in different populations might be attributed to different preferences, but might also be the result of small methodological changes.

Anchoring is a cognitive bias that arises when numerical estimates are affected by irrelevant information at hand (3). Even obviously random data unrelated to a particu-

lar question may lead people to focus on the information that is consistent with the anchor (4). Bias does not seem to disappear when the subject is aware of its existence (5). Anchoring has been found in a wide set of laboratory and real life situations (6-8), including health valuation (9,10). It may also be related to other cognitive biases (11,12).

A recent study found anchoring in TTO procedures as the starting point of subsequent valuations in a web survey (13). However, there are some issues to be tackled with respect to this result. First, the gold standard for TTO is face-to-face interviews. In addition, working with a heterogeneous population makes it difficult to isolate the anchoring effect. For instance, people of different ages may value differently. Also, the ten-year horizon of the standard TTO protocol has a different meaning for a young adult in their 20's than for an elderly person in their 60's. Since data are not normally distributed, isolating the anchoring effect by means of econometric regressions leaves room for discussion. Considering that the size of the anchoring effect may be affected by the lack of personal involvement and that heterogeneity in the general population makes it difficult to isolate the effect, in this paper we set out to find whether the starting point in a face-to-face TTO iteration procedure with a homogeneous highly educated population also induces anchoring in final health-state utilities.

Specifically, the TTO method seeks to find how many years in perfect health are equivalent (indifferent) to a year in certain health state A. This is achieved by asking whether a person would prefer to spend the rest of their life (for example, 10 years) in health state A and then die, or to spend 10 years in perfect health and then die. If the person chooses 10 years in perfect health, then the question is asked again changing the number of years in perfect health until an equivalent number is obtained. By way of example, a person reports being indifferent to spending 10 years in health state A and then dying, compared to spending 2 years in perfect health and then dying. In that case, a year in the health state under study is equivalent to 0.2 years in perfect health. Theoretically speaking, the procedure could start at 10 years and go down from there, start in zero and go up from there or start at any other number and go up or down in any order depending on the response. The assumption of procedural invariance means that the result should be the same regardless of the starting point (14); however, if there is anchoring, the starting point would affect the result, which is the object of study of this article.

METHODS

Population. Participants included 147 final-year economics students (111 males, 36 females) aged 18 to 25, ta-

king the research methodology course at Universidad Nacional de Colombia in 2015. Participation was voluntary and they could withdraw from the experiment at any time, no questions asked. Only one person declined to participate. The subjects were informed that the study was looking for determinants of preferences for health states and that it did not represent any hazard or breach of confidentiality for them. Informed consents were signed to participate. No compensation was paid for participating.

The participants valued five EQ-5D health states by TTO. All of them had a 40-year time horizon, so that the results were closer to their life expectancy. The subjects were randomly allocated to two groups. For the first group, the first question in the iteration procedure asked to compare 40 years in perfect health to 40 years in the valued health state, and then, the number decreased in 4-year periods. For the second group, the first question compared 20 years in perfect health to 40 years in the valued health state, and then it decreased or increased in 4-year periods depending on the answer.

Only states valued better than death were considered. When a subject valued a state as worse than death, it was excluded from the sample for that state, considering that valuation of worse than death states implied a process different from the process under study.

Health states were described using the EQ-5D-3L system, which uses five dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) at three levels (1=no problems, 2=moderate problems, and 3=severe problems). For a clearer exposition, each health state shall be referred to hereinafter by means of a letter: health states were 32211 (L), 22323 (Y), 21221 (X), 11121 (M), 33333 (S).

Interviewers (10 people) received previous training for the task and were not aware of the hypothesis being tested. They were randomly assigned to one of the anchors. For their part, participants were assigned randomly to the interviewers. Six of the interviewers performed the 40 anchor and four used the 20 anchor.

Anchoring should appear in the form of higher valuations for the group starting at 40 than for the group starting at 20. Distribution normality was tested through Shapiro-Wilk test, while results were tested by Kolmogorov-Smirnov (K_S), Wilcoxon-Mann-Whitney and Kruskal-Wallis tests. Ordered logistic regressions and box and whisker plots were also used to confirm the difference between the distributions.

In order to measure the size of the anchoring, the Jacowitz and Kahneman's (15) Anchoring Index was used (difference between medians divided by the difference between anchors).

RESULTS

Descriptive statistics of the variables are reported in Table 1. The number of observations was calculated after excluding worse-than-death states.

For each health state, the mean valuation of the anchor-40 group was higher than for the anchor-20 group in the same state. The Shapiro-Wilk test showed that no variable followed a normal distribution. The ks test was applied to verify equal distributions (Table 2).

The first column of the ks test shows that all states but M have lower values for the anchor-20 group than for the anchor-40 group and that this result is

significant at 1 %. The second line shows the probability of a value in the anchor-20 group being higher than in the other group, which does not happen in any case. The third line shows the combined tests and the p value; all states but M have a different distribution for both groups, with lower values for the anchor-20 group. The results with ks are supported by the Wilcoxon-Mann-Whitney test, as well as by the Kruskal-Wallis test (not reported), an ordered logistic regression (not reported), and box and whisker plots (Figure 1). The anchoring effect, measured by the Anchoring Index, is low compared to different estimations in the laboratory and in business (15,16).

Table 1. Descriptive statistics

Health state	Anchor	Mean	Median	Utility	Observations	% Difference between utilities	Anchoring Index
L	20	13.6	12.3	0.306	44		
L	40	19.2	19.5	0.488	77	59	0.36
Y	20	12.5	11.8	0.294	40		
Y	40	17.6	18.0	0.450	65	53	0.31
X	20	23.4	24.8	0.619	48		
X	40	27.8	28.5	0.713	86	15	0.19
M	20	31.5	35.6	0.881	50		
M	40	33.4	36.0	0.900	85	2	0.04
S	20	5.8	4.0	0.100	12		
S	40	11.3	11.5	0.288	34	188	0.38

Table 2. Non-parametric tests for normality and equal distributions

Health state	Shapiro-Wilk P value (normality)	Kolmogorov Smirnov P value (anchor-20 group < anchor-40 group)	Kolmogorov Smirnov P value (anchor-20 group < anchor-40 group)	Kolmogorov Smirnov Exact P value (equal distributions)	Wilcoxon-Mann-Whitney (equal distributions)
L	0.02916 **	0.002 *	0.995	0.003 *	0.0009 *
Y	0.00190 *	0.000 *	1.000	0.001 *	0.0052 *
X	0.00045 *	0.001 *	0.899	0.001 *	0.0082 *
M	0.00000 *	0.137	0.494	0.236	0.6184
S	0.00505 *	0.001 *	0.940	0.001 *	0.0030 *

* Significant at 1 %; ** Significant at 5 %

DISCUSSION

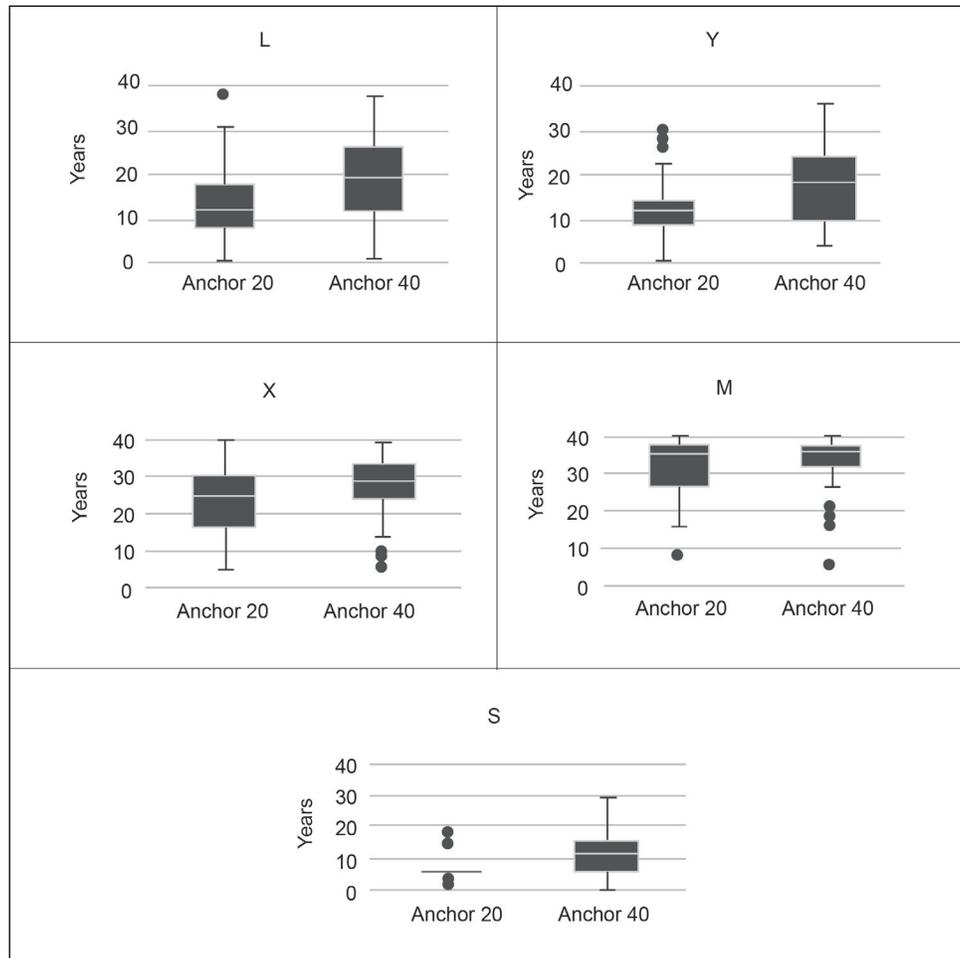
The results show that the starting point may act as an anchor in TTO health state valuations even during a face-to-face interview with educated people, so results are not affected by misunderstanding of the procedure by the subjects. The result for the health state M shows that anchoring might not be present in some health states, perhaps some near perfect health. The Anchoring Index is low compared to other situations (e.g. business) but the effect is large enough to be policy relevant, as the estimated utility using the 40-year anchor may be 15 % to 188 % higher than that estimated for the 20-year anchor. In addition, since not all states are equally affected, the effect should not be discarded in health technology assessments

because it may have a different impact on each branch of a decision tree. These results coincide with the only previous study on anchoring in TTO (13).

The experiment was applied to a student population, which is younger than the general population. Therefore, a 40-year time horizon was applied instead of the usual 10-year time horizon. Nevertheless, the point is that anchoring was observed in the young adult population and that it should be considered in surveys applied in the general population or patients. The fact that this is a highly educated population should not affect the results because anchoring susceptibility does not seem to be related to demographic and cognitive measures (7).

For this study, the Ping-Pong scheme used in British TTO EQ-5D studies (changing the year of comparison up

Figure 1. Box and whisker plots



and down in the questions) was not considered, since it is taken as equivalent to the consider-the-opposite method (17) to avoid anchoring. However, anchoring was found in the results obtained by Augestad using the Ping-Pong scheme (13). This study only addressed one specific scheme as the objective was isolating the relevant variables. The effect of different schemes should be the subject of future studies.

The distributions are not normal, hence the importance of using non-parametric tests, as regression-based results would have been less conclusive.

The conclusion of this exercise is that anchoring is present in TTO even when face-to-face interaction increases the subject's attention and understanding of the task •

Conflict of interest: None.

Acknowledgment: The work was supported by Universidad Nacional de Colombia.

REFERENCES

1. Arnesen T, Trommald M. Roughly right or precisely wrong? Systematic review of quality-of-life weights elicited with the time trade-off method. *J Health Serv Res Policy*. 2004; 9: 43–50.
2. Attema A, Edelaar-Peeters Y, Versteegh MM, Stolk EA. Time trade-off: one methodology, different methods. *Eur J Health Econ*. 2013; 14 (Suppl 1): S53-S64.
3. Tversky A, Kahneman D. Judgment under uncertainty: heuristics and biases. *Science*. (80) 1974; 185 (4157): 1124-1131.
4. Mussweiler T, Strack F. Hypothesis-consistent testing and semantic priming in anchoring paradigm: a selective accessibility model. *J Exp Soc Psychol*. 1999; 35(2): 136-164.
5. Welsh MB, Begg SH, Bratvold R. Efficacy of bias awareness in debiasing oil and gas judgments. In: McNamara DS, Trafletton JG, eds. *Proceedings of the 29th meeting of the Cognitive Science Society*. Austin, Texas: Cognitive Science Society; 2007. p. 1647-1652.
6. Furnham A, Boo HC. A literature review of the anchoring effect. *The J Socio Econ*. 2011; 40(1): 35-42.
7. Welsh MB, Delfabbro PH, Burns NR, Begg SH. Individual differences in anchoring: traits and experience. *Learning and Individual Differences*. 2014; 29: 131-140.
8. Welsh MB, Begg SH. What have we learnt? Insights from a decade of bias research. *Australian Petroleum Production and Exploration Association Journal*. 2016; 56(1): 435-450.

9. McNamee P, Ternent L, Gbangou A, Newlands D. A game of two halves? Incentive compatibility, starting point bias and the bidding game contingent valuation method. *Health Econ.* 2010; 19(1): 75-87.
10. Stalmeier PFM. Discrepancies between chained and classic utilities induced by anchoring with occasional adjustments. *Med Decis Making.* 2002; 22(1): 53-64.
11. Macbeth G, Razumiejczyk E. Modelado matemático del efecto del anclaje sobre el sesgo de subconfianza. *Revista de Psicología.* 2012; 30(2): 223-248.
12. Belaus A, Reyna C, Freidin E. Medición y manipulación de normas sociales en juegos experimentales de corrupción. *Cuad Econ.* 2016; 35(68): 353-377.
13. Augestad LA, Stavem K, Kristiansen IS, Samuelsen CH, Rand-Hendriksen K. Influenced from the start: Anchoring bias in time trade-off valuations. *Quality of Life Research.* 2016; 25: 2179-2191.
14. Lenert L, Treadwell J. Effect of failure to maintain procedural invariance on utility elicitation. *Med Decis Making.* 1999; 19(4):473-481.
15. Jacowitz KE, Kahneman D. Measures of anchoring in estimation tasks. *Personality and Social Psychology Bulletin.* 1995; 201: 1161-1166.
16. Luppe MR, Lopes LP. Anchoring heuristic and the estimation of accounting and financial indicators. *International Journal of Finance and Accounting.* 2012; 1(5): 120-130.
17. Mussweiler T, Strack F, Pfeiffer T. Overcoming the inevitable anchoring effect: considering the opposite compensates for selective accessibility. *Pers Soc Psychol Bull.* 2000; 26(9): 1142-1150.