VALUE IN HEALTH ▮ (2016) ▮▮▮–▮▮▮

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

# How Many Alternatives Can Be Ranked? A Comparison of the Paired Comparison and Ranking Methods

*Minsu Ock, MD, MPA*[1], *Nari Yi, MPH*[2], *Jeonghoon Ahn, PhD*[3], *Min-Woo Jo, MD, PhD*[1,*]

[1]*Department of Preventive Medicine, University of Ulsan College of Medicine, Seoul, South Korea;* [2]*Vital Statistics Division, Statistics Korea, Daejeon, South Korea;* [3]*Department of Health Convergence, Ewha Womans University, Seoul, South Korea*

A B S T R A C T

**Objectives:** To determine the feasibility of converting ranking data into paired comparison (PC) data and suggest the number of alternatives that can be ranked by comparing a PC and a ranking method. **Methods:** Using a total of 222 health states, a household survey was conducted in a sample of 300 individuals from the general population. Each respondent performed a PC 15 times and a ranking method 6 times (two attempts of ranking three, four, and five health states, respectively). The health states of the PC and the ranking method were constructed to overlap each other. We converted the ranked data into PC data and examined the consistency of the response rate. Applying probit regression, we obtained the predicted probability of each method. Pearson correlation coefficients were determined between the predicted probabilities of those methods. The mean absolute error was also assessed between the observed and the predicted values. **Results:** The overall consistency of the response rate was 82.8%. The Pearson correlation coefficients were 0.789, 0.852, and 0.893 for ranking three, four, and five health states, respectively. The lowest mean absolute error was 0.082 (95% confidence interval [CI] 0.074–0.090) in ranking five health states, followed by 0.123 (95% CI 0.111–0.135) in ranking four health states and 0.126 (95% CI 0.113–0.138) in ranking three health states. **Conclusions:** After empirically examining the consistency of the response rate between a PC and a ranking method, we suggest that using five alternatives in the ranking method may be superior to using three or four alternatives.

***Keywords:*** consistency, discrete choice experiments, paired comparison, ranking.



## Introduction

Two types of methods can be used in health economics to elicit stated preferences: cardinal methods and ordinal methods [1]. Typical examples of cardinal methods are the standard gamble (SG) and the time trade-off (TTO), and examples of ordinal methods include discrete choice experiments (DCE) and ranking methods [2,3]. First of all, the SG has a rigorous theoretical background because this task is conducted through comparisons of two alternatives under uncertainty about possible events or episodes [1]. There have been, however, several weak points from the perspective of feasibility, because people could not easily understand the concept of event probability. Furthermore, the values from the SG can be affected by loss aversion and risk attitude of respondents [1]. The TTO was then suggested by Torrance [4] as an alternative to the SG, and the person trade-off (PTO) was provided by Nord [5] to estimate disability weights that can be used to calculate disability-adjusted life-years. Although a TTO is considered to be easier for the public to understand than an SG, time preference can have an effect on the values of a TTO [6]. Furthermore, the lack of theoretical basis and ethical concerns have been constantly raised as criticism of the PTO [7–9].

Ordinal methods, such as the DCE, have been recently used to overcome the limitations of cardinal methods. For example, the EuroQol Group has tried to adopt the DCE for the five-level version of the EuroQol five-dimensional questionnaire valuation study on a trial basis [10]. In addition, in the Global Burden of Disease (GBD) Study 2010, the primary method for eliciting respondent preferences was a paired comparison (PC, a type of DCE), in which respondents were asked to select the better health state between two options [11]. Although a DCE could be easily conducted in the general population, there might be some difficulties in a study design when there are many alternatives to be compared. In the case of the GBD Study 2010, 220 unique health states were compared with each other and approximately 30,000 respondents participated in household surveys or a Web survey [11]. Obtaining a large sample size or increasing the number of questions for each respondent is needed to compare a large number of alternatives. If there are many alternatives to

* *Address correspondence to*: Min-Woo Jo, Department of Preventive Medicine, University of Ulsan College of Medicine, 86 Asanbyeongwon-gil, Songpa-Gu, Seoul 138-736, South Korea.

E-mail: mdjominwoo@gmail.com.

1098-3015$36.00 – see front matter Copyright © 2016, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

http://dx.doi.org/10.1016/j.jval.2016.03.1836

be compared, ranking methods can be viable alternatives to a DCE because the data from a ranking method, in particular a complete ranking method, can have more statistical information than those from a DCE [12,13].

Nevertheless, ranking methods also cannot infinitely increase the number of alternatives in each question because of cognitive burden. The cognitive burden increases substantially with an increase in the number of alternatives to be ranked, although there seems to be no consensus on the number of alternatives to be ranked [14]. Furthermore, in comparison with a DCE, the lack of consideration of the analytical method for a ranking method might be another problem [15]. Existing analytical methods for ranked data, such as the rank-ordered logit model, are difficult to apply if the alternatives to be ranked have no attributes or dimensions. Consequently, there have been attempts to convert the ranking method data into PC data [15,16]. Such analytical methods, however, may not reflect the actual preferences of respondents if the data from the ranking method do not agree with the converted PC data, that is, if there are logically inconsistent responses between PC data and ranking data [12]. These weaknesses of the analytical methods for ranking data could act as barriers to promoting extended application of ranking methods.

In our present study, we compared the consistency of the response rates between a PC and a ranking method to explore the feasibility of applying analytical methods for DCE to ranking data. We also assessed how many alternatives can be ranked, considering Pearson correlation coefficients between the PC and the ranking method. Thus, specifically, we compared the stated preference of two health states from a PC and that of three, four, and five health states from a ranking method.

## Methods

### Study Design and Health States

A household survey was conducted in a representative sample of 300 members of the general population around the capital area in South Korea. Sampling was performed using the multistage stratified quota method. A sample quota was assigned to each of the regions of the capital area (Seoul city, Inchon city, and Gyeonggi province) according to sex, age, and education level as defined by June 2013 resident registration data available through the Ministry of Administration and Security, South Korea. The survey was performed between March 25, 2014, and April 4, 2014. Data were collected using a survey program involving computer-assisted face-to-face interviews. The survey program provided the time recordings of the respondents. Thus, we could identify how much time each respondent spent performing each valuation method. We used a total of 222 health states, which reflected a diversity of health outcomes as a consequence of various disease causes. Of the 222 health states, 220 health states were from the GBD 2010 disability weight study and 2 health states were “full health” and “being dead.” Each health state (except “full health” and “being dead”) was made up of brief lay descriptions that explained the meaning of the health states in various aspects of health [11]. M. Ock initially translated the 220 health states from the GBD 2010 disability weight study into Korean, and M.-W. Jo modified them. A reverse translation process was also performed by a bilingual person, and M. Ock and M.-W. Jo reconfirmed the translation.

### Valuation Method and Survey Procedure

A computer program was developed for this study using the design of a previous study [11]. Participants in this program were initially asked for details about their sex, age, and educational level. Next, they performed three valuation methods (a PC, a PTO, and a ranking method). First, the respondents were asked to select the healthier option between two health states in each PC, considering physical or mental problems. To elicit a preference for health states, we asked the respondents to imagine experiencing the health problems for the rest of their lives. Second, the respondents compared the health benefit of two life-saving programs for eliciting trade-offs between “being dead” and less fatal health states in the PTO [11]. The purpose of the PTO, however, was to erase the memory of the PC in this study, and so we did not analyze the results of the PTO. Third, the respondents were asked to determine the ranking of health states in order of good health in the ranking method, considering physical or mental problems. In the same way as for the PC, we also asked them to imagine experiencing the health problem for the rest of their life time. Each respondent conducted a PC 15 times, a PTO 3 times, and a ranking method 6 times. In each of the two questions involving a ranking method, the respondents ranked three, four, and five health states. After completing the valuation methods to assess health states, the participants were also asked about their clinical information, including ambulatory care visits in the past 2 weeks, hospitalization in the past 12 months, and morbidity.

### Composition of the Questions

To compare the results between the PC and the ranking methods, the health states of the PC and the ranking methods were constructed to overlap each other. Table 1 shows the overall composition of the questions for the valuation methods. “Being dead” should be included in 1-A and 3-E, whereas “full health” should be included in 1-H and 3-F. The fifth and sixth questions of the ranking method (3-E and 3-F) will be reference points to other questions in terms of selecting the health states compared or ranked. For example, if H1, H2, H3, and H4 are selected among 220 health states in 3-E, the 5 health states to be ranked are H1, H2, H3, H4, and “being dead.” Then, the 4 health states of 3-C are randomly determined from the 5 health states in 3-E and the 3 health states of 3-A are randomly chosen from the 4 health states in 3-C. The 2 health states of 1-A to be compared are “being dead” and a randomly selected health state among 4 health states from 3-E (excluding “being dead”). The 6 questions from 1-B and 1-G are similar to playing a full league with 4 soccer teams. That is, all the possible combinations of comparison in the 4 health states from 3-E (excluding “being dead”) are the 6 questions from 1-B and 1-G (${}_4C_2 = 6$). If the 5 health states from 3-F are selected (including “full health”), the health states of 1-H to 1-O, 3-B, and 3-C are randomly determined from 3-F in the same way. As a result, each respondent evaluated 10 health states (including “being dead” and “full health”) using the PC and the ranking methods.

### Analysis

Initially, descriptive analyses for sociodemographic factors were performed. Before comparing the results between the PC and the ranking methods, we converted the ranked data into PC data. For example, if the orders of health states were “full health” > H5 > H6 in ranking the 3 health states (e.g., 3-B), they were converted as follows: “full health > H5,” “full health > H6,” and “H5 > H6.” This conversion was applied to the other ranking methods (i.e., ranking four health states and ranking five health states).

After conversion, we examined the consistency of the response rate. We defined the consistency of response rate as follows: (the number of coincident responses between the PC and the converted ranking methods)/(the number of converted responses in the ranking method) × 100%. For example, as mentioned earlier, assume that the orders of health states in