Validity, Feasibility and Acceptability of Time Trade-Off and Standard Gamble Assessments in Health Valuation Studies: A Study in a Multiethnic Asian Population in Singapore

Hwee-Lin Wee, PhD,[1] Shu-Chuen Li, PhD,[2,3] Feng Xie, PhD,[4] Xu-Hao Zhang, MSc,[2] Nan Luo, PhD,[4] David Feeny, PhD,[5,6] Yin-Bun Cheung, PhD,[7] David Machin, PhD,[8,9,10] Kok-Yong Fong, FRCP (Edin),[1,11] Julian Thumboo, FRCP (Edin)[1,11]

[1]Department of Rheumatology & Immunology, Singapore General Hospital, Singapore; [2]Department of Pharmacy, National University of Singapore, Singapore; [3]Discipline of Pharmacy & Experimental Pharmacology, School of Biomedical Sciences, University of Newcastle, Australia; [4]Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, ON, Canada; [5]Kaiser Permanente North-west Center for Health Research, Portland, OR, USA; [6]Health Utilities Incorporated, Dundas, ON, Canada; [7]MRC Tropical Epidemiology Group, London School of Hygiene and Tropical Medicine, London, UK; [8]National Cancer Center Singapore, Singapore; [9]Clinical Trials & Epidemiology Research Unit, Singapore; [10]School of Health and Related Research, University of Sheffield, Sheffield, UK; [11]Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore

ABSTRACT

Objectives: To assess the validity, feasibility and acceptability of standard gamble (SG) and time trade-off (TTO) assessments in a multiethnic Asian population.

Methods: Through in-depth interviews performed among Chinese, Malay, and Indian Singaporeans (education ≥ 6 years), we assessed validity of SG/TTO methods for eliciting health preferences by hypothesizing that 1) SG/TTO scores for three hypothetical health states (HS) would exhibit ranked order (decreasing scores with worse HS); and 2) more subjects would rate the most severe HS as worse than dead. Subjects also evaluated feasibility and acceptability of SG/TTO using a 10-point visual analog scale (VAS) and open-ended questions. Ratings were compared using Kruskal–Wallis, Wilcoxon signed-rank tests or tests of proportions.

Results: Validity: In 62 subjects (90% response rate), as hypothesized, SG and TTO scores exhibited ranked order with increasing HS severity (SG: 0.85, 0.08, −19.00; TTO: 0.85, 0.00, −0.18). More subjects rated the most severe HS as worse than dead (SG: 8%, 39%, 59%; TTO: 8%, 45% and 62%).

Feasibility: Subjects felt SG and TTO were easy to understand (median VAS scores: 8.0 vs. 8.0, $P = 0.87$) and to complete (8.0 vs. 8.0, $P = 0.84$). Acceptability: SG and TTO were well accepted, with TTO less so than SG (median [interquartile range] offensiveness: 2.0 [0, 4.0] vs. 2.0 [0, 3.0], $P = 0.045$). Overall, subjects did not have a clear preference for SG/TTO (50% vs. 45%, $P = 0.70$).

Conclusions: This study suggests the validity, feasibility and acceptability of SG and TTO for population-based HS valuation studies in a multiethnic Asian population.

Keywords: Asia, attitude to death, comparative study, quality of life, social values, value of life.

Introduction

Preference-based health-related quality of life (HRQoL) instruments, including the EQ-5D [1], Health Utilities Index [2] and the SF-6D (derived from the SF-36) [3] summarize HRQoL in a single index score and are used in cost-utility analyses with the aim of informing clinical policy and resource allocation in health care [4]. Each preference-based HRQoL instrument typically comprises a health classification system for describing the respondent in terms of a health state (HS), and a utility function that maps each HS to a utility score. The utility function is typically derived from a population-based HS valuation study in which respondents express their preferences for individual HS. These values are then aggregated using statistical modeling techniques to derive the utility function [5,6].

A variety of valuation methods have been employed in eliciting HS preferences, with choice-based valuation methods clearly preferred [7]. Standard gamble (SG) and time trade-off (TTO) are two choice-based methods with demonstrated acceptability, reliability, and validity in Western sociocultural contexts [7,8]. Both SG and TTO have established theoretical underpinnings [9,10], with the former being frequently referred to as the "gold standard" because it is directly based on the axioms of expected utility theory [9] although the latter requires the additional assumption

Address correspondence to: Julian Thumboo, Department of Rheumatology and Immunology, Singapore General Hospital, Outram Road, Singapore 169608. E-mail: julian.thumboo@sgh.com.sg

10.1111/j.1524-4733.2008.00361.x

 1098-3015/08/S3 S3–S10

that utility in additional healthy time is linear with respect to time [11]. Nevertheless, TTO is preferred by some (though not by others [12]) for its ease of implementation [9,10]. Importantly, studies have shown that health utilities elicited by both methods are usually different, with SG typically generating higher scores [13,14]. Hence, in planning studies to use choice-based preference measures such as SG and TTO in settings in which there has been relatively little experience with these measures, it is important to obtain empiric evidence with regard to validity, feasibility, and acceptability of each of these methods.

Existing studies comparing SG and TTO have largely taken a quantitative approach [7,13,15], with no studies (to the best of our knowledge) having addressed qualitative aspects to better understand individual subject's preferences and behavior when completing these exercises. We therefore conducted such a study in a multiethnic Asian population, in which a diversity of views might be expected. In this study, which is likely to be the first head-to-head comparison of SG and TTO methods in an Asian population, we aimed to assess the validity, feasibility and acceptability of SG and TTO and to evaluate if systematic differences in SG and TTO scores observed in other studies [13,14] were also observed in this Asian population. We defined 1) acceptability as the degree to which subjects are satisfied with SG and TTO and have no objections to these methods; and 2) feasibility as the extent to which SG or TTO exercise may be done practically and successfully. We then aimed to move beyond descriptive statistics (e.g., completion rate and missing data) to gain an insight into factors influencing individuals' preferences for SG or TTO.

Methods

Subjects and Study Design

In this Institutional Review Board approved study, in-depth interviews were conducted among consenting Chinese, Malay, and Indian Singaporeans with at least 6 years of education in either English or their mother-tongue (i.e., Chinese, Malay or Tamil) by interviewers of the same ethnic group. The various mother-tongue versions of the questionnaire were translated from the source English version using a standardized method of forward and back translations by independent native speakers of the target languages. To achieve adequate representation, two male subjects (one speaking English, the other his respective mother tongue) and two female subjects (one speaking English, the other her respective mother tongue) from each age band (20–29, 30–39, 40–49, 50–59, >60) were recruited from the Singaporean general population, giving a minimum of 20 subjects per ethnic group.

Subjects were required to perform several tasks. First, subjects expressed their preferences for three HS using SG and TTO (detailed in the Appendix). The HS were selected to represent varying degrees of impaired health as defined in the EQ-5D MVH protocol [5]: mildly impaired (11122), moderately impaired (23321) and severely impaired (32313), and were administered in randomized order (sequence generated using STATA) [16]. Additionally, the selected HS needed to be plausible. For example, we considered it difficult for subjects to conceive HS 31111, where apart from being confined to bed, there was no impairment on the other dimensions of health. Furthermore, each selected HS needed to have at least a level two impairment for each health dimension to increase data variability. Each HS in the EQ-5D is described by a five-digit code where each digit represents a single-item health dimension (i.e., mobility, self-care, usual activities, pain/discomfort, and anxiety/depression), with value ranging from 1 (no problem) to 3 (severe problem). To ensure comparability of SG and TTO scores, we chose to rate each state on a dead-to-perfect health scale (score range 0–1) in both SG and TTO methods. On such a scale, HS regarded as worse than dead would be assigned negative scores.

Second, subjects evaluated the following aspects of feasibility of SG and TTO on a 10-cm 0 to 10 horizontal visual analog scale (VAS): 1) ease with which they understood the instructions; 2) ease with which they completed the exercise; and 3) amount of concentration needed. Acceptability of SG and TTO were similarly evaluated by asking subjects to rate the degree of offensiveness of each method using another horizontal VAS. In addition, for the TTO, their ease with discussing trading off life years and for SG, ease with thinking in terms of chance, willingness to take health risks and willingness to take financial risks. Higher scores indicate greater ease, less concentration needed, less offensiveness (i.e., greater acceptability) and greater willingness. Subjects were considered risk averse if their risk attitudes for either health or finances were less than 5 points on the VAS and risk loving if risk attitudes were more than 5 points on the VAS. Subjects were also asked to give their views on the following questions, with their answers recorded verbatim: "How can we improve the clarity of SG and TTO instructions?" and "Do you find it offensive to talk about giving up certain number of years of your life? Why? Is it a taboo?" Finally, subjects were asked to specify if they preferred SG or TTO, and to explain their choice. The purpose of asking subjects to evaluate each component of the SG and TTO exercises before asking them to state their preference for SG or TTO was to allow us to gain an insight into the factors that influenced their overall preference for either method.