Health Utility Scores in Alzheimer's Disease: Differences Based on Calculation With American and Canadian Preference Weights

Mark Oremus, PhD*, Jean-Eric Tarride, PhD, Natasha Clayton, Canadian Willingness-to-Pay Study Group, Parminder Raina, PhD

Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, ON, Canada

A B S T R A C T

Objective: Health utility scores quantify health-related quality-of-life (HRQOL) in Alzheimer's disease (AD). These scores are calculated by using preference weights derived from general population samples. We recruited persons with AD and their primary informal caregivers and examined differences in health utility scores calculated by using two sets of published preference weights. **Methods:** We recruited participants from nine clinics across Canada and administered the EuroQol five-dimensional (EQ-5D) questionnaire HRQOL instrument. We converted participants' EQ-5D questionnaire responses into two sets of health utility scores by using US and Canadian preference weights. We assessed agreement between sets by using the intraclass correlation coefficient. Bland-Altman plots depicted individual-level differences between sets. **Results:** For 216 persons with AD and their caregivers, mean health utility scores were higher when calculated with US instead of Canadian preference weights ($P < 0.0001$). The intraclass correlation coefficient (95% CI) was 0.79 (0.05–0.93) in the persons with AD group and 0.83 (0.30–0.94) in the caregiver group. Ninety-five percent of the individual differences in utility score fell between −0.16 and 0.03 for persons with AD and −0.15 and 0.05 for caregivers. Forty-three percent of these differences exceeded a minimum clinically important threshold of 0.074. **Conclusions:** In AD studies, researchers should calculate health utility scores by using preference weights obtained in the general population of their country of interest. Using weights from other countries' populations could bias the utilities and adversely affect the results of economic evaluations of AD treatments.
Keywords: EQ-5D, health economics, health-related quality-of-life, quality-adjusted life-years.

Copyright © 2014, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

Introduction

Alzheimer's disease (AD) can adversely affect cognition, function, behavior, and mood. Because of the complexity of its manifestations, medications developed to treat AD have been evaluated on all four domains [1]. These medications, that is, the three cholinesterase inhibitors (donepezil, rivastigmine, and galantamine) and one *N*-methyl-D-aspartate receptor (memantine) treat only the symptoms of disease in each domain. The medications do not delay, halt, or reverse disease progression, nor do they affect survival [1]. Therefore, the impact of AD medications becomes an issue of health-related quality of life (HRQOL). The question is whether alleviating the symptoms of disease enhances HRQOL [2–4].

HRQOL is measured by using generic (e.g., EuroQol five-dimensional [EQ-5D] questionnaire [5], Health Utilities Index Mark II or III [6]) or disease-specific (e.g., Dementia Quality-of-Life [2]) instruments. Responses to these instruments can be converted into health utility scores with algorithms derived primarily from samples of the general public. These algorithms are known as "preference weights." Health utility scores range from 0.0 (death)

to 1.0 (perfect health) and represent a person's overall health state. Some very poor health states (e.g., persistent vegetative state) may be represented by health utility scores below 0.0.

The generalizability of preference weights between populations is uncertain. Weights have been calculated for some populations yet not others. Researchers who wish to compute health utility scores for one population may be required to use preference weights from another population. No previous studies have examined the transferability of preference weights from one population to another in AD.

Transferability between populations is important because health utility scores are used to calculate quality-adjusted life-years (QALYs) in economic evaluations [7]. These evaluations are becoming an important evidentiary tool in health policy, as payers seek to allocate scarce resources to treatments that demonstrate value for money [8–10]. In AD, the importance of economic evaluations was highlighted when UK's National Institute for Health and Care Excellence (NICE) used economic data [11] to recommend that Britain's National Health Service delist coverage of cholinesterase inhibitors for persons with mild-stage disease.

Because existing AD medications do not modify the course of the disease, the life-years gained component of a QALY should not vary substantially across different treatments or placebo. Thus, the calculation of QALYs in AD becomes dependent on health utility scores. Using preference weights that are inappropriate for a particular population could produce biased QALYs for that population. A need exists to investigate the generalizability of preference weights between populations in AD.

Although no work to examine generalizability has been done in AD, earlier work in cardiac care found differences in health utility scores calculated by using US and UK preference weights [12]. Another study of 42 EQ-5D questionnaire health states also found differences between health utilities calculated by using US and UK weights [13].

The purpose of this study was to compare health utility scores obtained with the EQ-5D questionnaire and two different sets of preference weights. One set of weights was calculated in the United States [14] and the other in Canada [15]. Persons with AD and their caregivers formed the study sample. We included caregivers because they provide the preponderance of care for persons with AD. The stress and burden of the caregiving role are so intense that caregivers have been called the "hidden victims" of AD [16].

We examined the Canadian weights because they are relatively new, having first been published in 2012 [15]. Also, several Canadian jurisdictions, including the provinces of Alberta, Ontario, and Québec, emphasize economic evaluations in decision making. The Canadian situation thereby provides a timely opportunity to investigate the generalizability of preference weights. We chose the US weights as a comparator because of the tendency to use preference weights from a "similar" population in cases in which weights do not exist for the population in question. The results of this study will be of interest to researchers and decision makers in other countries besides Canada. Because these individuals will have to select appropriate preference weights for their populations, they should be aware of potential pitfalls in the transferability of weights from similar populations.

Methods

Study Participants and Recruitment

We recruited persons with AD and their primary informal (unpaid) caregivers. Recruitment took place between November 2008 and August 2011 in nine memory or geriatric clinics across Canada. Persons were included if they had a diagnosis of AD based on *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision* criteria [17] or National Institute of Neurological and Communicative Disorders and Stroke – Alzheimer's Disease and Related Disorders Association criteria [18]. To avoid recruiting persons whose level of cognitive impairment would prevent them from answering the study questionnaire, participation was restricted to individuals with mild or moderate AD according to the Functional Assessment Staging in Alzheimer's disease scale [19]. Persons were included only if their primary informal caregiver also agreed to participate in the study. All participants had to speak English or French.

Recruiting physicians identified potentially eligible persons with AD before regular clinic visits. At the visits, research assistants approached these persons and their caregivers, explained the study, and obtained informed consent. Consenting participants were scheduled to complete a single one-on-one interview in the clinic or in the home.

Data Collection

The study questionnaire elicited sociodemographic information such as age and sex, followed by the EQ-5D questionnaire [5]. The EQ-5D questionnaire asks participants to rate their current health state on each of five dimensions, that is, mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Three response options are available: no problems, some/moderate problems, or extreme problems/inability to perform tasks. Persons with AD and caregivers rated their own individual HRQOL on the EQ-5D questionnaire; caregivers did not provide proxy HRQOL ratings for care recipients. A professional translator converted the questionnaire (except the EQ-5D questionnaire questions) into French for use with French-speaking participants from the province of Québec. Two French-speaking study investigators (M.O. and J.E.T.), one of whom has French as a first language, verified the translations. We used the established French Canada version of the EQ-5D questionnaire in the French-language questionnaire.

We converted EQ-5D questionnaire responses into two sets of health utility scores. One set was calculated with US preference weights and the other set was computed with Canadian preference weights. The US weights came from a sample of 4048 persons representing the civilian, noninstitutionalized, English- and Spanish-speaking US general population [14]. The Canadian weights came from a sample of 1145 English-speaking persons who were members of a market research panel, with quotas used to obtain an age and sex distribution that matched the Canadian general population [15]. In our study, participants who did not respond to one or more EQ-5D questionnaire dimensions were excluded from the computation of health utility scores.

Statistical Analysis

We reported sample characteristics as medians and interquartile ranges for continuous variables and frequencies for categorical variables. Differences between the two sets of health utility scores were tested by using the Wilcoxon signed rank sum test. The intraclass correlation coefficient (ICC) [20] was used to assess the level of agreement between the two sets of scores. Specifically, we chose the ICC (3,1) to examine the extent to which the US and Canadian weights gave the same utility scores. ICCs of more than 0.75 indicated excellent agreement [21].

We used Bland-Altman plots [22] to graphically depict individual-level differences between the two sets of health utility scores. To calculate these differences, we subtracted scores based on Canadian preference weights from scores based on US preference weights. The Bland-Altman plots show 95% limits of agreement, which encompass 95% of the individual-level differences in our sample. Narrow limits indicate smaller differences between the two sets of health utility scores.

To assess the clinical significance of these differences in score, we counted the number of participants for whom the difference exceeded a minimum clinically important difference (MCID) of 0.074. This MCID was estimated in a study using UK preference weights and participant data from eight published longitudinal studies [23].

To examine the impact of disease severity on health utility scores, we stratified persons with AD according to the severity of disease (mild or moderate) and recomputed the mean health utilities for each stratum. Caregiver strata were based on the disease severity of the care recipient.

We used SAS v9.2 (The SAS Institute, Cary, NC) to conduct all analyses except the computation of ICCs. ICCs were computed in SPSS v20 (IBM Corp, Armonk, NY) by using a two-way mixed-effects analysis of variance model and absolute agreement. The study received ethics approval from McMaster University