VALUE IN HEALTH ▮ (2017) ▮▮▮–▮▮▮

Available online at www.sciencedirect.com

ScienceDirect

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

Comparing Analytic Hierarchy Process and Discrete Choice Experiment to Elicit Patient Preferences for Treatment Characteristics in Age-Related Macular Degeneration

Marion Danner, Dipl.Vw. MPH[1,*], *Vera Vennedey, MSc*[1], *Mickaël Hiligsmann, PhD*[2], *Sascha Fauser, DrMed*[3], *Christian Gross, MSc*[1], *Stephanie Stock, DrMed*[1]

[1]*Institute for Health Economics and Clinical Epidemiology, University Hospital of Cologne (AöR), Cologne, Germany;* [2]*Department of Health Services Research, CAPHRI School for Primary Care and Public Health, Maastricht University, Maastricht, The Netherlands;* [3]*Center for Ophthalmology, University Hospital of Cologne (AöR), Cologne, Germany*

ABSTRACT

Background: In this study, we conducted an analytic hierarchy process (AHP) and a discrete choice experiment (DCE) to elicit the preferences of patients with age-related macular degeneration using identical attributes and levels. **Objectives:** To compare preference-based weights for age-related macular degeneration treatment attributes and levels generated by two elicitation methods. The properties of both methods were assessed, including ease of instrument use. **Methods:** A DCE and an AHP experiment were designed on the basis of five attributes. Preference-based weights were generated using the matrix multiplication method for attributes and levels in AHP and a mixed multinomial logit model for levels in the DCE. Attribute importance was further compared using coefficient (DCE) and weight (AHP) level ranges. The questionnaire difficulty was rated on a qualitative scale. Patients were asked to think aloud while providing their judgments. **Results:** AHP and DCE generated similar results regarding levels, stressing a preference for visual improvement, frequent monitoring, on-demand and less frequent injection schemes, approved drugs, and mild side effects. Attribute weights derived on the basis of level ranges led to a ranking that was opposite to the AHP directly calculated attribute weights. For example, visual function ranked first in the AHP and last on the basis of level ranges. **Conclusions:** The results across the methods were similar, with one exception: the directly measured AHP attribute weights were different from the level-based interpretation of attribute importance in both DCE and AHP. The dependence/independence of attribute importance on level ranges in DCE and AHP, respectively, should be taken into account when choosing a method to support decision making.

Keywords: age-related macular degeneration, analytic hierarchy process, convergent validity, discrete choice experiment, patient preference(s).

Introduction

Discrete choice experiments (DCEs) and the analytic hierarchy process (AHP) are two of a range of multicriteria approaches that can be used to elicit decision makers' preferences and assist in decision making [1–4]. The German Institute for Quality and Efficiency in Health Care (IQWiG), the US Food and Drug Administration, the European Medicines Agency, and other health technology assessment agencies have tested these methods to support health care decision making, to supplement health economic analyses, or to get structured information, particularly on patient preferences [5–10]. It, however, remains unclear how these methods perform in comparison with each other and whether they deliver comparable results.

DCE and AHP use very different approaches to measure the relative importance of treatment attributes and levels to decision makers. To facilitate the reading of this article, the terms "attributes" and "levels" are used throughout, despite often being referred to within the AHP methodology as decision "criteria" and "subcriteria," respectively. AHP structures attributes and levels in a decision hierarchy, and respondents compare these to one another pairwisely at each level and in each cluster of the hierarchy. DCE, however, uses combinations of attribute levels to develop descriptions of hypothetical treatment options and asks respondents to repeatedly choose between two or more of these options in order. Although the AHP asks respondents to value the relative importance of two attributes or levels directly on a nine-point ratio scale, the DCE requires a repeated discrete

Conflicts of interest: The authors have indicated that they have no conflicts of interest with regard to the content of this article.

* *Address correspondence to:* Marion Danner, Institute for Health Economics and Clinical Epidemiology, University Hospital of Cologne (AöR), Gleueler Straße 176-178, Cologne 50935, Germany.

E-mail: marion.danner@uk-koeln.de.

1098-3015$36.00 – see front matter

http://dx.doi.org/10.1016/j.jval.2017.04.022

choice between options. The expression of a preference for or against an option in DCE requires individuals to simultaneously weigh attribute levels against each other. AHP uses a direct mathematical approach to calculate importance weights for attributes, levels, and, if necessary, treatment alternatives, whereas DCE, being rooted in utility theory, estimates coefficients of a utility function using regression analysis. Of note is the difference between the meaning of "preference" and "importance" in the two analytic approaches: coefficients in a DCE indicate a positive or negative preference for or against attribute levels in relation to the other levels of that attribute, whereas AHP importance weights are positive weight values for levels of an attribute and attributes, respectively, adding up to 1. These weights do not per se convey information about the direction of a preference (positive or negative). The AHP permits calculation of a "consistency ratio" (CR), which measures how plausible one pairwise comparison is with respect to other pairwise comparisons. Such consistency measurement is not part of the DCE, which relies on the assumptions of random utility theory. Plausibility checks testing patients' attention and task comprehension can nevertheless be performed in a DCE by including repeated choice sets or dominant options in the experiment. A detailed description of the AHP method is given in Saaty [11], Dolan et al. [12], and Hummel et al. [13] and of the DCE methodology in Bridges et al. [4], Hauber et al. [14], Hensher et al. [15] , and Muhlbacher et al. [16].

Although the theoretical and methodological frameworks in AHP and DCE are different, the German IQWiG focused on testing these two methods to generate preference-based weights for the prioritization of the outcome-specific results of benefit and/or cost-effectiveness assessment [17]. The IQWiG stated that further research in both methods was needed to explore their methodological properties and aspects such as reliability. Recent guidance issued by the US Food and Drug Administration Center for Devices and Radiological Health similarly stresses that patient preference assessment is an active and evolving area of research and encourages further refinement of the methods [18]. One initial step in this refinement process could be to explore the convergent validity of the most commonly used methods, a step that was recently taken by several research projects (e.g., [19–21]). Few studies till now have compared DCE and AHP for preference elicitation, and we identified only three such studies in a health care setting [22–24]. Nevertheless, these studies were either not initially designed to compare methods or suffered from limited comparability because of differing attribute selection, labeling, and/or framing [22,23]. One of the studies did not elicit patients' preferences but rather those of physical therapists [24]. Although a range of studies compared other methods against each other in a single population with identical attributes and levels using tools such as a rating scale or best-worst scaling as compared with DCE or AHP [20,21,25–27], to date no study has compared AHP and DCE in a single patient population—an oversight that this study attempts to rectify. The study was performed on patients suffering from age-related macular degeneration (AMD).

The primary objective of our study was to compare preference-based weights for AMD treatment attributes and levels generated by two different elicitation methods—DCE and AHP. Our secondary objectives were to explore the ease of use of the instruments for patients and to obtain insights into the advantages and disadvantages of the techniques.

Methods

We used DCE and AHP to elicit and compare preferences of patients with AMD, a chronic progressive eye disease in elderly individuals. Methods and results have previously been published for both AHP and DCE and are only briefly summarized here [28,29]. Methods for comparing attribute and level importance weights and comparative ease of questionnaire use are described in more detail.

Treatment Alternatives

Three vascular endothelial growth factor inhibitors administered as intravitreal injections are used for the treatment of AMD in Germany. Although aflibercept and ranibizumab are approved for AMD, bevacizumab is not. The latter is nevertheless used off-label, costs considerably less, and is paid for under special contracts with sickness funds. Drugs also differ in the frequency of administration and monitoring according to the label.

Sample Size, Study Population, and Setting

On the basis of an empirically derived formula for DCE sample size calculation, a minimum sample size of 84 patients was required [30]. Consecutive patients receiving intravitreal injections with a vascular endothelial growth factor inhibitor were recruited at Cologne University Hospital. Patients with a diagnosis of neovascular AMD, a minimum visual function of 5%, and treatment experience of at least one intravitreal injection were included in the study. Data were collected by means of interviewer-assisted, paper-and-pencil questionnaires. The study was approved by the ethics committee of the Cologne University Hospital (approval no. 14-011).

Attributes and Levels

Details on attribute and level identification, selection, and the framing process, combining a literature review with expert and focus group interviews, have been previously published [31]. The key reasons for selecting specific attribute and level definitions on the basis of the preceding research are provided in Appendix 1 in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2017.04.022. The attributes and levels included in the study are shown in Figure 1.

Questionnaire Development and Data Collection

An English version of the questionnaire is provided in Appendix 2 in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2017.04.022. (The original questionnaire was in German.) The questionnaire consisted of four parts. It included an introduction with an explanation of the attributes and levels, questions regarding sociodemographic and disease status, and the AHP and DCE survey parts. AHP and DCE were allocated to patients first or second on a random basis to explore questionnaire order bias.

The AHP task consisted of 24 pairwise comparisons based on the AHP decision hierarchy (Fig. 1). Pairwise comparisons were performed at each level and in each cluster of the hierarchy. The DCE part of the questionnaire included 13 choice tasks. One was a duplicated task to check for patient attentiveness. In each task, participants had to choose between two options described by the levels of the five attributes. Examples of AHP pairwise comparisons of attributes/levels and a DCE choice scenario are shown in Figure 2. At the end of each survey, the patients were asked to rank the degree of difficulty they experienced with each method on a scale from 1 to 5, where 1 meant difficult, 2 meant rather difficult, 3 meant neutral, 4 meant rather easy, and 5 meant easy. Patients were encouraged to think aloud throughout the survey. Interviewers took notes to follow the patients' decision making and trace problems in providing judgments for either questionnaire part.