GENERAL DISCUSSION
GENERAL DISCUSSION

Importance of the studies and overview of the issues

Over the past decades, as advances in medical treatments have improved survival and reduced key morbidities, perceived health status or health-related quality of life (HRQoL) assessments are becoming more and more relevant. HRQoL is a measure of perceived health status consisting of physical, mental, and social domains [1]. Regulatory bodies such as the Food and Drug Administration (FDA) [2] and the National Institute for Health and Care Excellence (NICE) [3] actively encourage measuring patient-reported HRQoL in addition to traditional clinical assessments. The public and patients are asked to make value judgments about different aspects of health to obtain a subjective yet numerical expression of the quality of a certain health state.

HRQoL instruments can be developed on the basis of various measurement frameworks. However, when comparing HRQoL across different populations, conducting disease modeling studies, and performing economic evaluations of various healthcare interventions, it is more reasonable to use preference-based instruments. These differ from other measures in that, by expressing the evaluation in a single metric score, they explicitly incorporate weights that reflect the importance attached to specific health aspects. A single metric score represents the quality of a health state holistically and is referred to as a health-state value.

The aim of this thesis was to investigate the specific problems associated with preference-based measures of health states and with the methodology used to derive health-state values. Health-state values can be at a cardinal or an interval measurement level [4]. If these values are transformed or normalized on a scale ranging from death to full health, they become utilities. Such utilities are often used in computing quality-adjusted life years, a key notion in cost-effectiveness analysis. Throughout this thesis we applied discrete choice modeling, which produces values, which cannot be formally referred to as utilities, but for our methodological investigations the use of values is sufficient.

Since the research underlying the present thesis required a large number of respondents, we needed to find a simple instrument enabling a self-completion format. Therefore, we selected the EQ-5D, a relatively simple and widely used generic preference-based instrument. For example, the guidelines for pharmacoeconomic research in the UK and in the Netherlands recommend the EQ-5D instrument for health-state evaluations [5, 6]. In the EQ-5D the description of any health-state can be presented by means of five health domains (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression), where each domain has a limited number of levels. Different health domains (also called attributes) are weighted on the basis of comparisons the respondents have to make. For example, they may be asked to choose between health-state descriptions. Their choices allow the researchers to determine the relative importance of these attributes and of the levels for each attribute [7, 8].
The valuation framework of discrete choice modeling, which was selected for this thesis, is consistent with the random utility model in economic theory [9]. Now it is used to augment economic evaluations in the realm of healthcare with information about the value of non-health outcomes such as waiting time, location of treatment, and type of care [10]. Moreover, this technique has been used to elicit personal and societal preferences in health-valuation studies [10]. Discrete choice models started to attract attention as an alternative to the valuation frameworks of time trade-off and standard gamble; these latter techniques were complex and prone to bias (time preference, states worse than dead, risk averse, scale compatibility) [11-15]. A promising feature of discrete choice models is that the derived values only relate to the attractiveness of a health state. They are not expressed in trade-offs between improved health and something else, as in time trade-off and standard gamble. Discrete choice is considered a relatively easy task for the respondents since it mimics individual everyday choices: ‘Which of the available options is more preferable?’ In comparison, conventional valuation techniques, such as time trade-off and standard gamble, are based on an iterative procedure, whereby a choice is made between each of the presented options till the point of indifference is reached. This task is challenging for the respondents and has led to biases. By contrast, respondents can tackle up to 20 discrete choice tasks without feeling fatigued. This enables the researchers to get enough information to specify more complex health values. Discrete choice tasks can be conducted online, thereby cutting out the interviewer (who would be needed for the time trade-off or standard gamble) and eliminating possible interviewer effects on the responses [16, 17]. However, a known limitation of discrete choice models is their inability to produce values in the form of absolute numbers. All values in discrete choice models are relative and are located on a latent scale from ‘best’ to ‘worst’ states. That limitation can be attributed to the location of ‘dead’, which is unknown since that option was not included as a choice option in the response tasks. Therefore, one of the main problems with discrete choice models is how to normalize the scale to dead–full health (0.0 – 1.0). To solve this problem and enable utility calculations, the design should include a task extension or additional tasks. In this thesis, we did not normalize the values on dead and full health to generate utilities. Instead, we focused on other aspects of health-state measurement. Specifically, we devoted our attention to four issues: the content and description of health states; problems with preference-based estimation (interactions); whose responses should be considered (those of the general population or of patients); and the respondents’ attention to the discrete choice tasks.

**Issue 1: Content and description of health states**

Various generic instruments that have been developed to provide values or utilities of health states define and describe the same health states in different ways. Consider, for example, the relatively simple instrument developed by the EuroQol Group the EQ-5D [18, 19]. In the
standard version (EQ-5D-3L) each of the attributes can take on three levels [20]. However, it was presumed to have specific drawbacks - a restricted discriminatory power, lack of sensitivity to smaller changes in health states, a substantial proportion of respondents report themselves as being in the best health state (ceiling effect) - which prompted an update of the instrument [21, 22]. For the new version, the EQ-5D-5L, the number of levels used to classify health states was increased from three to five, and the phrasing of health-state descriptions was changed. According to several earlier studies, the differences between the levels in the EQ-5D-5L are subtle and may be hard to distinguish, which might have caused some of the language versions (notably the English) to show inconsistencies at the upper or lower levels of health attributes [23-25]. Such inconsistencies would affect the validity of the estimated values. The current Dutch study (Chapter 1) used discrete choice modeling, thereby deviating from earlier studies [23-25], and found no inconsistencies for either version of the EQ-5D. However, the overall weights for the attributes were different in the two versions. In the EQ-5D-5L the highest weight was attributed to anxiety/depression, while in the EQ-5D-3L the highest weight was attributed to mobility. A change made in the wording of the description of the mobility attribute from ‘confined to bed’ to ‘unable to walk’ is a possible explanation for the shift in the level of importance. In that light, researchers should be cautious about describing health states with existing instruments or revising the descriptions. Apparently even small differences in wording could affect individual responses and thereby the elicited values. Another concern is how comprehensively the health state needs to be described. Insufficient detail could lead the respondents to start guessing, thus distorting the actual description of the health state in question. On the other hand, excessively detailed descriptions would overload the respondents with information to be processed, thereby creating fatigue and leaving more room for random responses.

Issue 2: Interactions between attributes
Attention was also given to the inclusion of interactions between distinct health attributes. Not only did we find that such interactions exist, but we showed that the combined effect of two separate health attributes is stronger than the sum of their individual effects. Most models are main-effects models, which take into account the individual effects but not the interactions of distinct health attributes. However, we suggest that interactions of health attributes should be taken into account. In that regard, future studies might test whether specific interactions exist between health attributes of interest. For example, a researcher might seek to verify that the most salient interactions found in the present study would again appear with a different version of an instrument and a different setting. The results of the current study are applicable only to the original version of the generic preference-based EQ-5D instrument, where the health-state descriptions are based on health attributes with three levels each. It would be important to find out whether similar results can be achieved with other preference-based instruments, such as HUI or SF-6D, because the set of attributes used for the construction of instruments can influence
the strength of the interactions between these attributes. However, these instruments use larger sets of attributes and levels, which makes it harder to evaluate all possible interaction terms. In this case, we have to rely on theoretical knowledge and the literature to test the significance of specific interactions between the health attributes of interest.

**Issue 3: Whose responses?**

Conventionally, the health-state values used in economic evaluations are derived from a representative community sample, a convention based on several arguments [26, 27]. One is that, as taxpayers, these persons are deemed to represent the general population [28, 29]. Another is the ‘veil of ignorance’ argument [28], whereby the general population is presumed to have never experienced the impaired health states under evaluation and to be blind to its own self-interest. Accordingly, representatives of the general population would embody principles of justice and equity and, thereby, ensure a fair distribution of resources. However, a community sample consists mainly of healthy or relatively healthy persons, who may be inadequately informed or have insufficient imagination to make an appropriate judgment about the impact of (severe) health states. Instead, judgments made by patients are put into the spotlight, assuming that people who have direct experience with impaired health will provide more reliable and informative health-state valuations [30, 31]. A lack of consensus on whose values to use motivated the investigation of whether any differences between the public and patients’ judgments actually exist. In that vein, an important aspect of this thesis was to compare different population samples. A comparison was made between healthy respondents and patients regarding health-state valuations, and a comparison was made between the general public and patients regarding new medical treatments (Chapters 3 and 4). We did not find large differences between the health values elicited from the general population and patients, but we did find differences between those of patients and healthy respondents. The values elicited from the general population can be quite similar to the values of healthy respondents (in case the majority of the population is healthy) or to those of patients (in case the majority of the population has some experience of disease). These results suggest that the sample of a representative general population should be constructed with some caution. While sampling based on demographic characteristics is important, it is also important to ensure an equal representation of the proportion of respondents with and without specific diseases. The reason for such caution is to avoid that overall population values reflect only healthy individuals or only patients. However, it may happen that patients who have certain diagnosed diseases might not have experienced some of the health states they are asked to evaluate and therefore might not be able to give informed judgments based on actual experience. By the same logic, healthy respondents could have been patients in the past and, thus, have experienced the health states under evaluation, enabling them to give informed answers based on experience. This discrepancy raises the issue of experience-based values. In the future, researchers could consider asking respondents to value health states that lie nearby their own (current or past) health
status. In that way, the health states would not be completely hypothetical. Unfortunately, we did not ask the respondents in our studies to specify whether they had experienced the health states under evaluation, but it would be an interesting area for future study.

**Issue 4: Attention of respondents to the discrete choice tasks**

The last issue raised in this thesis was prompted by a basic assumption in the preference-based measurement framework. It is generally assumed that respondents pay equal attention to all components of information (e.g. left-side versus right-side alternatives, specific attributes used in the description) presented in the response task. These tasks require the respondent to make considered choices among two or more scenarios described by specific attributes with certain levels. Using an eye-tracking device, we investigated whether the respondents were paying attention to all of the information that was presented in the discrete choice tasks. Disregarding specific attributes would lead to an incomplete understanding of the health-state descriptions and, thus, to biased results with the statistical models applied in health evaluations. However, the eye-tracking study revealed that the respondents did pay attention to all of the information elements and were not fatigued after completion of the task. It should be mentioned that we used a specific layout (EQ-VT, a standardized valuation study protocol) for the study. The effect of applying eye-tracking to different versions or different layouts, such as a changed ordering of the attributes, different color schemes, different placement of information cues on the screen, could be an interesting topic for future research. The results would give insight into the optimal layout and appearance of the health valuation instrument, allowing researchers to simplify it and make it more attractive for the respondents. The effectiveness of discrete choice studies depends strongly on the survey design, and not only on the experimental design but also on the layout design. Careful selection of health-state descriptions and survey elements (phrasing, instructions, ease of navigation, information notes, and visual cues) are required to conduct a study. When these standards are met, the respondents are more likely to take account of the full range of information presented and make more informed judgments. Moreover, the problem of fatigue would be reduced by making the design simple and attractive.

To sum up, this thesis has provided evidence in support of using discrete choice modeling for health-state measurements. The present study has drawn attention to some important issues: the content and description of health states; interactions between health attributes; the construction of samples to derive the health values; and the importance of the survey design to enhance the respondents’ attention to the response tasks. We pointed out that preference-based health-state measurement is associated with several methodological drawbacks that might warrant attention in future research. First, a simple main-effect model for health-state measurement may not be sufficiently accurate to produce credible health-state values. Therefore, interaction terms would need to be tested and studied carefully. Second, researchers would need to be prudent when
developing or modifying a preference-based instrument. Even small differences in the phrasing or the valuation technique in combination with particular statistical models may affect the expected results and elicited values. Third, the judgments of healthy people may differ from those of patients who are actually experiencing health limitations. It would be advisable to use values based on assessments by patients instead, as patients are likely to be more adequately informed than healthy people or more adept at imagining certain health states. Accordingly, patients may be better motivated to make an informed judgment about the impact on perceived health of such states. However, if the patient community wants to have a central role in defining value, robust models would be needed to incorporate the patient voice in a value assessment that is free from adaptation and other biases. Fourth, it is important to invest in the design and layout of preference-based instruments in general and discrete choice tasks in particular, with the objective of making the task attractive to the respondents without causing fatigue. We believe that a logical, simple, and reliable measurement model is needed for the measurement of a subjective phenomenon such as health status. The discrete choice model and various related choice models are probably qualified candidates.
REFERENCES


5. Guidelines for pharmacoeconomic research, updated version. College voor zorgverzekeringen, Diemen March 2006


