# Novel methods in preference-based health outcome measurement

Zhang, Xin

*DOI:*
10.33612/diss.875541897

# CHAPTER 6

## General Discussion

# Introduction

With the advance of modern medicine, healthcare has evolved from physician-centered to patient-centered. The importance of patient involvement and the measurement of patient-reported outcomes have also gained substantial recognition and appreciation.[1] Regulatory agencies such as the US Food and Drug Administration, the National Institute for Health and Care Excellence in the UK, and the Dutch National Health Care Institute (Zorginstituut) actively encourage the recording of patient-reported outcomes (PROs) to supplement conventional clinical assessments.[2,3,4] Alongside clinical outcomes (e.g., survival, blood pressure), the measurement of patient-reported outcome (or subjective health outcomes, such as patient's experience, well-being, health status, health-related quality of life) offers valuable insights about health concerns from patients' perspective. This information is useful for tailoring care to individual patient needs, facilitating shared decision-making in clinical practice, and improving the overall quality of healthcare. This thesis centered around exploring the performance of a novel multi-attribute preference response (MAPR) measurement framework. This MAPR measurement framework integrated patient-centered, preference-based, electronic patient-reported outcome measures (ePROMs), a new and simpler preference-based method (Drop-Down, DD), and user-friendly electronic mobile applications. The ePROM used in this thesis is the generic health outcome measure CS-Base. Studies in this thesis explored the effects of various aspects of the CS-Base rooted in the MAPR framework, including the content (descriptive system), the preference-based method, the produced health-state values, and the overall performance of the integrated framework.

# PROMs founded on the MAPR framework

**Patient-centered content and outcomes**

As explained in the introduction, the relation between patient-centeredness and PROMs is intertwined. It is advocated that involving patients' input to determine the content and identify what matters to them should be a priority in the development of PROMs. The application of PROMs supports and enhances the implementation of patient-centered care, by actively involving patients to assess their health status and report their health outcomes themselves. To our knowledge, the CS-Base is the first preference-based generic PROM which is fully patient-centered in its development and construction. During its development, the 12 items that form

the CS-Base were selected by patients themselves from a pool of 47 candidate items. These candidate items were drawn from existing generic health status instruments by a systematic literature review, to ensure they cover all health domains.[5] The outcomes measured are all reported by patients. This is done by asking patients to describe their own health status, and assess their own preference for health.

The results of our studies showed that the CS-Base reflected patients' health conditions well, indicated by the fact that the number of observations of problems reported on the CS-Base items were consistent with the complaints that can be expected when looking at the distribution of health conditions reported. According to our findings from the study which compared the CS-Base with EQ-5D-5L, three of the 12 CS-Base items (mobility, hearing, vision) had the highest attribute (item) importance, the other 9 items had similar moderate importance. There were four similar items (mobility, usual/daily activities, anxiety, pain) between CS-Base and EQ-5D-5L, the other eight items are present in the CS-Base exclusively. The attribute (item) importance of the four similar items were comparable between the two measures. Among the eight CS-Base exclusive items, two (hearing, vision) showed high importance, the remaining six items resulted in similar or higher importance compared to the four similar items. Health problems were reported by no less than half of the patients on five of the eight exclusive CS-Base items. These observations indicate that almost all the items captured in the CS-Base were regarded as important by patients. This finding is reasonable, as these items were deliberately selected by patients according to their importance in the development of the CS-Base. One the other hand, studies of the EQ-5D-5L also reported that important health items were not appropriately captured by the EQ-5D-5L. For example, the items related to sleep, relationships, and tiredness,[6,7] as well as items related to mental health or diseases of the nervous system.[8] In another well-known generic health outcome measure, Health Utilities Index Mark 3, some CS-Base items such as hearing, vision and cognition, were also included as part of its eight items in total.[9]

Nevertheless, there might be limitations remaining. Although the CS-Base items were selected by patients, the collection of the candidate items was based on a selection of existing generic health outcome measures. The patients were not invited to suggest any additional items. So, some items may have been missed because they were not covered by existing measures.

**Simple and statistically robust preference-based methods**

Preference-based methods apply special techniques to elicit patients' or respondents' preference for health, so to derive weights (coefficients) assigned to attribute (item) levels. These weights are subsequently computed into a single index which reflects the overall quality of the health state. Most of the existing preference-based methods originate from the fields of social sciences (e.g., psychology, economics, marketing) or health sciences (e.g., health economics, clinimetrics), and are based on one of the three fundamental measurement models for subjective phenomena: item response theory (IRT),[10] discrete choice experiments (DCEs),[11] and valuation techniques.[12] Commonly used preference-based methods such as standard gamble (SG),[13] time trade-off (TTO),[14] are valuation techniques stemming from the field of health economics. These valuation techniques have been criticized to be susceptible to both theoretical and empirical limitations including time preference, loss aversion, adaptation, being cognitively demanding, interviewer effects and others. The DCEs emerged later as an alternative to these conventional valuation techniques. DCE method entails comparisons of two or more hypothetical health states. It has been rapidly adopted and widely used because it's easier to perform than TTO or SG, and free from some of the limitations of conventional methods.[15]

The Drop-Down (DD) method introduced in this thesis is a simpler preference-based method compared to DCE. As in the DCE, patients are asked to compare and make a trade-off between two hypothetical health states. In the DD method, patients own health states as assessed by themselves are presented and they are asked, several times, to select the item that hinders them the most. In the DCE, hypothetical states are presented. It could be challenging for patients to imagine those states since they have no prior experience with them. The health states presented to patients through the drop-downs are states that they can likely imagine or may have even experienced, as these states are based on patients' own health states, with each varying on only one level of one item. Thus, the DD method could be simpler to perform than the DCE method, as well as TTO or SG, which all present hypothetical health states.

Another strength of the DD method is that it is powerful regarding statistical performance. As evidenced by our studies, the DD method produced statistically robust coefficients. All the coefficients followed a logical order and were all statistically significant. The DD coefficients were far more precise with small standard errors and small confidence intervals, when compared to the Better-Worse (BW) method which entails paired comparison of patients' own health state with hypothetic health states. The DD coefficients demonstrated good face validity when compared to those of the EQ-5D-5L. Differences between both instruments were clearly

discernible *across the levels of all items.* Additionally, the coefficients of each of the item levels were rather *comparable across all the 12 items* to a high degree. However, as for the EQ-5D-5L (coefficients derived by TTO method), the differences of coefficients between levels of all items were very subtle or even overlapped. The preference-based method used in the EQ-5D-5L (US version compared in our study) is TTO, which is susceptible to various limitations as described above. This might lead to the unsatisfying face validity of EQ-5D-5L coefficients.

We assume the reasons behind the outperformance of the DD method over the BW method and the EQ-5D are that it eliminates two factors that could introduce bias or distortion. Firstly, the DD method effectively excludes the distorting factor of "status quo" (or risk aversion, loss aversion). Such a distorting phenomenon is observed in the BW method,[16] where respondents disproportionately chose their own health states as better than those hypothetic health states. This is probably because patients are unwilling to make balanced trade-offs between their own health states and hypothetic states, due to the risk aversion. Individuals can have a strong tendency to remain in their current health state, as they may weigh losses heavier than gains in decision-making.[17,18] Loss aversion doesn't play a role in the DD method, therefore it is capable of gathering richer information that carries more data variation than the BW method and contributes to more powerful statistical results. In addition, the DD method is less susceptible to adaptation, which is another prominent distorting factor affecting valuations of health states. Once patients get adapted to their health status over time, they tend to report relatively better health status using PROMs. Studies showed that health-state values are typically higher (better health status) among patients with chronic illness or disability, compared to the values among non-patients who only imagine themselves in hypothetical health conditions.[19] This drawback can now be averted in the DD method, since each DD produced a health state better than in a previous DD, and the tasks are based on forced-selection. In each DD task, patients are asked to select one item that disturbed them most. Therefore, the concern of adaption can be mitigated in the DD method.

Despite many advantages of the DD method rooted in the MAPR framework, there remains a limitation: it produces values instead of utilities. It is less straightforward in generating utilities due to the absence of an anchor, such as "dead" or time duration in the tasks. The "utilities" are health-state values, anchored on a unidimensional scale ranging from 0.0-1.0, where 0.0 = dead, 1.0 = full health. Utilities enable for the computation of quality-adjusted life years and can be directly used for health economic evaluation. A separate study is needed to generate data that can be used to rescale the DD-based values (where the location of "dead" is unknown) to

utilities.[20] Moreover, although those typical distorting factors as described above do not seem present in the DD method according to our findings, we cannot rule out the possibility that other distorting factors or biases may play a role in the task of the DD method. Because the DD task is so simple and so connected to the natural experience of patients ("what is disturbing me most in my disease state?"), however, we consider it as highly unlikely that a significant factor or bias mechanism would play a role in the background.

**Integrated MAPR measurement framework**

A strategic, coordinated, integrated approach to PROM assessment is recommended to meet the needs of multiple stakeholders.[21] An ideal integrated approach should strive for a harmonized and streamlined process encompassing the selection, collection, analysis, reporting, and application of PROMs.[22] Specifically, it should address key considerations such as capturing what matters to patients, establishing an non-burdensome pathway for patients to provide meaningful patient-centered data, utilizing robust analysis methods, and running in a compatible information technology (IT) system.

The MAPR measurement framework can serve as a pioneer to an integrated approach to PROMs assessment. Constructed within the MAPR framework, several PROMs have been developed including the generic CS-Base PROM studied in this thesis, as well as disease-specific PROMs like the Infant Health-Related Quality of Life Instrument (IQI),[23] the Transplant PROM (TXP),[24] and the cardiovascular disease-specific PROM (CVD-specific PROM).[25] As discussed above, all these PROMs founded on the MAPR framework are completely patient-centered in their development and construction, which ensures that patient-centered, patient-reported data is collected. All these MAPR PROMs consist of a brief but comprehensive descriptive system, comprising only several items (no more than 18), each with 4 levels. The assessment tasks are easy and not time consuming. From our studies, the whole assessment process of CS-Base took less than two minutes. A simple and statistical powerful preference-based method (the DD method) is applied to these MAPR PROMs for collecting patients' preference of health, which is the central part in the construction of preference-based PROMs.

Smart mobile applications have been specially designed for developing and running the MAPR electronic PROMs (ePROMs), containing a HealthFan and a HealthSnApp. Applying the HealthFan, health items extracted through a comprehensive literature review are systematically organized into distinct main domains of physical, mental, and social health, along with additional subdomains. Patients are then asked to select several items they deem most important.

Organized in a HealthFan, the presentation of the candidate items is visually appealing and informative, which facilitates the patients to conduct the selection of items.

The HealthSnApp is devised for a seamless transition between the two tasks (Task 1: descriptive task of health state, Task 2: preference-based task). HealthSnApp ensures an interactive communication and is convenient to perform in a self-completed format. This application is useful in several ways. Firstly, it facilitates data collection from patients by allowing them to perform the tasks remotely. Secondly, together with an integrated database server, the HealthSnApp allows for rapid storage and retrieval of data as well as a continuous update of the weights of the levels for the items. Such an integrated method of measurement and data collection has not been implemented in the field of preference-based health outcome measurement before. It is different from a widely applied health outcome measure such as the EQ-5D, where three key areas of divergence can be identified in its construction and analysis. In distinctive studies conducted in different countries using the EQ-5D, the following areas were found to be dissimilar: different measurement (preference-based) methods were used; different designs were used in conducting the measurement; and different analyses of the response data were applied.[26,27]

Overall, it seems that the CS-Base, along with other MAPR ePROMs, offers the advantage of running on a compatible IT system, which enhances user-friendliness, facilitates efficient data storage, seamless data exchange and utilization, and provides a standardized and integrated platform for the application of PROMs. Therefore, the MAPR framework can be regarded as a successful endeavor in exploring an integrated approach. It establishes a coordinated framework that can probably effectively address the key considerations which are essential for a high-quality implementation of PROMs. Although we see potential benefits of applying MAPR PROMs in clinical trials and real-world studies, their application is in a very early stage. There is much to explore regarding their performance in the real-world application and their contributions to the healthcare system, as well as refining the detailed steps and improving the whole framework.

Despite the advantages elaborated above, there is a potential challenge of using ePROMs in clinical practice. While not exclusive to the CS-Base or other MAPR ePROMs, it should be noted that not all patients possess the capacity to accurately assess their own health conditions, nor can all individuals successfully complete the required tasks using a smartphone or computer interface. Some specific groups of patients are unable to do this, for example, because of cognitive impairments (e.g., mentally ill, young children, elderly), communication deficits, or

because of severe distress caused by their illness. In that case the assessment done by proxies (e.g., caregivers, spouse) may be a second-best solution.

## Patients' perspective versus societal perspective

A dispute remains on the use of patients' preferences or the preferences of the general public when generating utilities for health economic evaluation. Regulatory agencies commonly advocate for using utilities based on the preferences of the general population, as they represent the societal perspective.[3,28] This is justified by the fact that healthcare is publicly funded, and the general public, as taxpayers are potential users of the healthcare system and should be involved in resource allocation decisions. On the other hand, patients are more likely the primary users of specific health policies, disease management approaches, or medical interventions. As we mentioned earlier, there are differences between the preferences of the general population and those of patients when it comes to health-related matters. These preference disparities significantly impact the estimated utilities and can consequently affect the cost-effectiveness outcomes.[29] As a result, utilities derived from the general population may not optimally cater to the specific needs of the intended patient population.

No agreement has been reached on the dispute of patients' perspective or societal perspective. Each of both the perspectives has its pros and cons. While the general public's preferences reflect broader societal considerations, patients possess unique insights into the consequences and impact of their own health conditions. Ignoring patients' preferences could lead to a lack of alignment between the evaluated interventions and the outcomes that truly matter to those directly affected. To reconcile the two perspectives, a more balanced approach can be pursued. In our study, a two-step approach was used to estimate utilities of the CS-Base, accordingly, two different samples were used in each step, a patients-sample and a general-population sample. There might be a potential benefit of our approach, the CS-Base utilities are derived in by combining the patients' perspective with the societal perspective. The values of the CS-Base health states are derived from patients who experienced the health states themselves. The location of "dead" for anchoring values to the utility scale is derived from a representative sample of the general population. In this manner, the two-step approach we applied can be a balanced solution to engage both patients and the general public's preference.

# Future perspectives

Advancement in various aspects can be expected to facilitate further evolutions of MAPR ePROMs. From the measurement aspect, first, more studies should be conducted to evaluate the usefulness and easiness of the newly developed DD method. Second, the current preference-based methods developed within the MAPR framework have a close relationship with the assessment tasks as conducted within the item-response theory (IRT) measurement framework. Both the MAPR and the IRT are probabilistic measurement models. IRT is a more advanced and sophisticated framework for preference-based health outcome measurement compared with conventional DCE or other preference-based measurement frameworks. However, the analysis performed for the current preference-based method in this thesis is based on conventional statistical models as used for the DCE and other related models. A next step can be to fully construct the MAPR assessment tasks and its analysis as in the IRT measurement framework. Furthermore, the current approach for generating utilities for MAPR ePROMs is based on two distinct steps that needs two separate studies. This might reduce the efficiency of generating utilities. To address this issue, an integrated study approach can be considered. For instance, searching for possibilities to include "dead" as an anchor point in an additional task as part of the ePROMs, while also modifying the software to accommodate this change. Regarding the software, we anticipate the substitution of textual content with icons or animations to create more attractive and user-friendly ePROMs. This is likely to overcome language barriers and save time for researchers as legitimate translations are no longer needed. Such standardized content of the ePROM can also be helpful in pooling data from different countries while maintaining methodological consistency. Concerning the application of these new patient-centered and preference-based ePROMs, such as the CS-Base, which are founded on a compelling measurement framework, they can be used in a variety of settings. For example, they could be used as a standard component of routine health outcomes measurement[30] in clinical practice. They can also be integrated into the health information system of hospitals. Applying these ePROMs could facilitate healthcare providers to better track and address patient-specific needs, leading to improved treatment outcomes and ultimately enhancing the quality of healthcare.

# References


1 Porter ME, Larsson S, Lee TH. Standardizing patient outcomes measurement. N Engl J Med. 2016;374(6):504-506.

2 Food and Drug Administration. Guidance for industry on patient-reported outcome measures: use in medical product development to support labeling claims. 9 Dec 2009. Accessed 23 Jun 2023. Available from: https://www.regulations.gov/document/FDA-2006-D-0362-0005

3 National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2013 [Internet]. London: National Institute for Health and Care Excellence (NICE); 4 Apr 2013. Process and Methods Guides No.9.

4 Zorginstituut Nederland. Guideline for economic evaluations in healthcare. 16 Jun 2016.

5 Krabbe PFM, van Asselt ADI, Selivanova A, Jabrayilov R, Vermeulen KM. Patient-centered item selection for a new preference-based generic health status instrument: CS-Base. Value Health. 2019;22(4):467-473.

6 Finch AP, Brazier J, Mukuria C. Selecting bolt-on dimensions for the EQ-5D: Testing the impact of hearing, sleep, cognition, energy, and relationships on preferences using pairwise choices. Med Decis Making. 2021;41(1):89-99.

7 Yang Y, Rowen D, Brazier J, Tsuchiya A, Young T, Longworth L. An exploratory study to test the impact on three "bolt-on" items to the EQ-5D. Value Health. 2015;18(1):52-60.

8 Finch AP, Brazier J, Mukuria CW. Validity and responsiveness of preference-based measures: A critical overview of reviews. Value Health. 2015;18:A744.

9 Feeny D, Furlong W, Torrance GW, Goldsmith CH, Zhu Z, DePauw S, Denton M, Boyle M. Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. Med Care. 2002;40(2):113-28.

10 Rasch G. An item analysis which takes individual differences into account. Br J Math Stat Psychol. 1966;19(1):49-57.

11 McFadden D. Conditional logit analysis of qualitative choice behavior. In: Zarembka P. Frontiers in Econometrics. Academic Press, San Diego. 1974.

12 Lugnér AK, Krabbe PFM. An overview of the time trade-off method: Concept, foundation, and the evaluation of distorting factors in putting a value on health. Expert Rev Pharmacoecon Outcomes Res. 2020;20 (4):331-342.

13 von Neumann J, Morgenstern O. Theory of Games and Economic Behavior (16th ed). Princeton University Press. Princeton. 2004.

14 Dolan P. Modeling valuations for EuroQol health states. Med Care. 1997;35(11):1095-1108.

15 McFadden D. Economic choices. Am Econ Rev. 2001;91(3):351e378. Nobel Lecture, Dec 2000.

16 Shahabeddin Parizi A, Vermeulen KM, Gomes-Neto AW, van der Bij W, Blokzijl H, Buskens E, Bakker SJ, Krabbe PF. Using a novel concept to measure outcomes in solid organ recipients provided promising results. J Clin Epidemiol. 2021;139:96-106.

17 Kahneman D, Tversky A. Prospect theory: An analysis of decision under risk. Econometrica. 1979;47(2):263-291.

18 Samuelson W, Zeckhauser RJ. Status quo bias in decision making. J Risk Uncertain. 1988;1:7-59.

19 Menzel P, Dolan P, Richardson J, Olsen JA. The role of adaptation to disability and disease in health state valuation: A preliminary normative analysis. Soc Sci Med. 2002;55(12):2149-2158.

20 Krabbe PFM, Jabrayilov R, Detzel P, Dainelli L, Vermeulen KM, van Asselt ADI. A two-step procedure to generate utilities for the Infant health-related Quality of life Instrument (IQI). PLoS One. 2020;15(4):e0230852.

21 van der Wees PJ, Nijhuis-Van Der Sanden MW, Ayanian JZ, Black N, Westert GP, Schneider EC. Integrating the use of patient-reported outcomes for both clinical practice and performance measurement: Views of experts from 3 countries. Milbank Q. 2014;92(4):754-775.

22 Calvert M, Kyte D, Price G, Valderas JM, Hjollund NH. Maximising the impact of patient reported outcome assessment for patients and society. BMJ. 2019;364:k5267.

23 Jabrayilov R, van Asselt ADI, Vermeulen KM, Volger S, Detzel P, Dainelli L, Krabbe PFM. A descriptive system for the Infant health-related Quality of life Instrument (IQI): Measuring health with a mobile app. PLoS One. 2018;13(8):e0203276.

24 Shahabeddin Parizi A, Krabbe PFM, Buskens E, van der Bij W, Blokzijl H, Hanewinkel V, Annema C, Bakker Stephan JL, Vermeulen KM. Health items with a novel patient-centered approach provided information for preference-based transplant outcome measure. J Clin Epidemiol. 2020;126:93-105.

25 Ties D, Singh TK, Zhang X, van Veghel D, Schalkers I, Groot HE, Krabbe PFM, van der Harst P. What really matters: A patient-centered instrument to evaluate health-related quality of life in cardiovascular disease. Eur Heart J Qual Care Clin Outcomes. 2022;8(7):722-729.

26 Norman R, Cronin P, Viney R, King M, Street D, Ratcliffe J. International comparison in valuing EQ-5D health states: A review and analysis. Value Health. 2009;12(8):1194-2000.

27 Janssen MF, Bonsel GJ, Luo N. Is EQ-5D-5L better than EQ-5D-3L? A head-to-head comparison of descriptive systems and value sets from seven countries. Pharmacoeconomics. 2018;36(6):675-697.

28 Canadian Agency for Drugs and Technologies in Health. Guidelines for the economic evaluation of health technologies: Canada (4th Edition). 2017.

29 Rowen D, Mulhern B, Banerjee S, Tait R, Watchurst C, Smith SC, Young TA, Knapp M, Brazier JE. Comparison of general population, patient, and carer utility values for dementia health states. Med Decis Making. 2015;35(1):68-80.

30 Wikipedia. Routine health outcomes measurement. 30 Aug 2023. Accessed 8 Sep 2023. Available from: https://en.wikipedia.org/wiki/Routine_health_outcomes_measurement.