



ScienceDirect

Contents lists available at [sciencedirect.com](http://sciencedirect.com)  
Journal homepage: [www.elsevier.com/locate/jval](http://www.elsevier.com/locate/jval)

## Health Policy Analysis

# Economic Evaluation of New Models of Care: Does the Decision Change Between Cost-Utility Analysis and Multi-Criteria Decision Analysis?



Esther H.A. van den Bogaart, MSc, Mariëlle E.A.L. Kroese, PhD, Marieke D. Spreeuwenberg, PhD, Dirk Ruwaard, PhD,\*  
Apostolos Tsiachristas, PhD\*

## ABSTRACT

**Objectives:** To experiment with new approaches of collaboration in healthcare delivery, local authorities implement new models of care. Regarding the local decision context of these models, multi-criteria decision analysis (MCDA) may be of added value to cost-utility analysis (CUA), because it covers a wider range of outcomes. This study compares the 2 methods using a side-by-side application.

**Methods:** A new Dutch model of care, Primary Care Plus (PC+), was used as a case study to compare the results of CUA and MCDA. Data of patients referred to PC+ or care-as-usual were retrieved by questionnaires and administrative databases with a 3-month follow-up. Propensity score matching together with generalized linear regression models was used to reduce confounding. Univariate and probabilistic sensitivity analyses were performed to explore uncertainty in the results.

**Results:** Although both methods indicated PC+ as the dominant alternative, complementary differences were observed. MCDA provided additional evidence that PC+ improved access to care (standardized performance score of 0.742 vs 0.670) and that improvement in health-related quality of life was driven by the psychological well-being component (standardized performance score of 0.710 vs 0.704). Furthermore, MCDA estimated the budget required for PC+ to be affordable in addition to preferable (€521.42 per patient). Additionally, MCDA was less sensitive to the utility measures used.

**Conclusions:** MCDA may facilitate an auditable and transparent evaluation of new models of care by providing additional information on a wider range of outcomes and incorporating affordability. However, more effort is needed to increase the usability of MCDA among local decision makers.

**Keywords:** cost-utility analysis, health services research, local decision-making, multi-criteria decision analysis, new models of care.

VALUE HEALTH. 2021; 24(6):795–803

## Introduction

New models of care are implemented worldwide to increase accessibility, equity, and affordability of care by developing new approaches of collaboration and delivering health and social care.<sup>1</sup> Health authorities are experimenting with new models of care at the local level (so-called pilot sites, pioneer sites, or vanguards) that act as blueprints and inspiration to the rest of the healthcare system.<sup>2</sup> These models may take different forms because they are tailored to the local needs and context and are driven by local leaders from multiple organizations who collaborate and are responsible to improve access and quality of care for their local populations.<sup>3–6</sup> However, learning from and scaling up these local initiatives at a national level is challenging given the lack of a framework that assesses their cost-effectiveness.<sup>7,8</sup>

The appropriateness of the widely accepted cost-utility analysis (CUA) in local decision making is debatable. This is because

the decision context is different than in health interventions and technologies that are subject to reimbursement decisions or national clinical guidelines.<sup>9,10</sup> Furthermore, CUA includes quality-adjusted life-year (QALY) as a single measure of outcome and fails to incorporate outcomes of interest to local decision makers such as access to care, equity, patient satisfaction, and non-health-related quality of life.<sup>11–13</sup> It is widely argued that economic evaluations of new models of care should be carried out at a local level,<sup>13,14</sup> should be flexible to accommodate the selection of all relevant outcomes and costs at different levels (eg, individual, organizational, local),<sup>15–18</sup> and should incorporate the perspective of all relevant stakeholders.<sup>19</sup> Local decision makers are often using decision support tools, such as balanced scorecards and key performance indicators, to monitor various performances. However, these tools, similar to a cost-consequence analysis (CCA), lack clear decision rules to indicate cost-effectiveness.<sup>20,21</sup>

\*Dirk Ruwaard and Apostolos Tsiachristas share joint senior authorship.

Multi-criteria decision analysis (MCDA) is proposed as a suitable method for evaluating new models of care based on the performance on all relevant outcomes and perspectives of all relevant stakeholders.<sup>22-24</sup> In previous studies, MCDA was used to evaluate integrated care initiatives.<sup>25,26</sup> However, it is unclear if the adoption of MCDA in the evaluation of new models of care would alter funding decisions at local level. Therefore, this article aims to investigate whether the adoption of MCDA instead of traditional CUA alters the decision of investing in new models of care by using a new Dutch model of care as a case study. By conducting a side-by-side application of the CUA and MCDA, the applicability and suitability of both methods to support local decision makers about the broader value for money of new model of care is tested.

## Methods

### Setting and Decision Context

Primary Care Plus (PC+) was a new model of care implemented in the Dutch pioneer site Blue Care in the southern Netherlands in 2014 by primary and secondary care providers, local health authorities, the largest health insurer in the region, and a local patient organization.<sup>27,28</sup> Its triple aim was to improve population health and patient satisfaction by improving quality of and access to care while avoiding unnecessary outpatient hospital visits.<sup>29</sup> After its piloting, the decision of the multiple collaborating and responsible organizations (ie, stakeholders) was to disinvest or scale up its implementation in other areas in The Netherlands.

### Primary Care Plus versus Hospital-Based Outpatient Care

PC+ was compared to care-as-usual, which was hospital-based outpatient care (HBOC). GPs within the pioneer site Blue Care were able to refer non-acute and low-complex patients either to a medical specialist in 1 of the 2 PC+ centers in the city of Maastricht or to HBOC. Despite the fact that PC+ was available for all GPs within the region, only one-third of the GPs referred patients regularly.

In PC+, the medical specialist examined and/or treated the patient during a maximum of 2 consultations. Following PC+, the medical specialist referred the patient back to the GP with treatment advice or, if necessary, referred the patient to HBOC for further diagnosis and/or treatment. PC+ consultations were provided by medical specialists who worked in the HBOC in the Maastricht UMC+ who visited PC+ centers on a regular basis (weekly or biweekly).

### Study Design and Data

In this longitudinal prospective observational study with a 3-month follow-up, we used data of 2116 adult patients who visited initially PC+ (intervention group) or HBOC (care-as-usual group) between December 2014 and April 2018. Data on patients' health and wellbeing and experience of care were collected prospectively using a survey. Healthcare consumption and related costs in PC+ and hospital care were retrieved retrospectively from patient medical records. The different data sources were linked and merged into 1 data set. Diagnosis treatment combinations at baseline were used to select patients referred to HBOC for low-complex hospital care (1 or 2 consultations with a medical specialist) to define a control group comparable to the PC+ group.

### Outcomes

Outcomes (or criteria in the MCDA context) were selected and grouped according to the Triple Aim<sup>30</sup> (ie, health and well-being, experience of care, and costs) following Stiefel and Nolan's overview of outcomes per aim<sup>31</sup> and were operationalized using indicators (ie, outcome measurements). A brief description of the outcomes related to the Triple Aim and the associated indicators is provided below and presented in Table 1. A more detailed description of the indicators is included in Appendix Table 1 in Supplemental Materials found at <https://doi.org/10.1016/j.jval.2021.01.014>.

Self-reported health and well-being was measured in patients using the EuroQol 5-dimensional questionnaire with 5 levels (EQ-5D-5L) and the Short Form Health Survey version 2 (SF-12v2) at baseline, 1-week, and 3-month follow-up. The EQ-5D-5L is a generic measure based on the construct of Health-Related Quality of Life and consists of questions on mobility, self-care, pain/discomfort, usual activities, and anxiety/depression.<sup>32</sup> The SF-12v2 focuses on the outcomes of "physical functioning" and "psychological well-being."<sup>33</sup> Both indicators can be applied to persons with all different types of diseases and complaints. The 2 criteria (ie, person-centeredness and access to care) related to the experience of care were measured using the Consumer Quality (CQ) index at 1-week follow-up. The CQ index is a scientifically based, standardized tool, which can be used throughout the care sector to measure self-reported experience of care from the patient perspective.<sup>34</sup> For person-centeredness, we used a multicomponent indicator based on 3 items of the CQ index, namely the degree that care matches an individual's needs, capabilities, and preferences, and jointly making informed decisions. Furthermore, we used the time between referral and start of treatment as the indicator of the criterion access to care. The indicator for the

**Table 1.** Criteria and related indicators relevant for Primary Care Plus.

Triple Aim	Criteria	Indicators
Health and well-being	Health-related quality of life	EuroQol 5-dimensional questionnaire with 5 levels (EQ-5D-5L)
	Physical functioning	Short-Form Health Survey version 2 (SF-12v2)
	Psychological well-being	Short-Form Health Survey version 2 (SF-12v2)
Experience of care	Person-centeredness	Consumer Quality (CQ) index
	Access to care	Consumer Quality (CQ) index
Costs of care	Costs of care	Healthcare consumption and related costs in Primary Care Plus and hospital care

*Note.* All data were collected through questionnaires except for costs that were retrieved from administrative databases.

criterion cost of care were the costs of all consultations with a medical specialist (including at PC+) and all hospital admissions from baseline to 3-month follow-up.

### Statistical Analysis and Propensity Score Matching

Descriptive statistics were produced in terms of means and standard deviations (SDs) for continuous variables, and frequencies for categorical variables. Moreover, propensity score matching (PSM) was performed to reduce observed confounding between the 2 groups by following a stepwise strategy.<sup>35-37</sup> First, we followed standard practice<sup>38</sup> and included in the propensity score model all possible confounding variables available in the data set: age, gender, residence at birth, educational level, baseline health status (SF-12v2 physical and mental component summary score), historical healthcare costs (12 months before baseline), and medical specialty referred to. Second, the 2 groups were matched using several PSM techniques, including exact matching, nearest neighbor greedy, caliper (0.25) and optimal matching, full matching, genetic matching, and inverse probability weighting. In addition to 1-to-1 matching, n-to-1 matching was used to keep a larger sample size since the HBOC group was smaller. The performance of the different PSM techniques on covariate balancing was assessed based on standardized mean differences (SMDs), Rubin's B (the absolute standardized difference of the means of the linear index of the propensity score in the intervention and (matched) control group), and Rubin's R (the ratio of intervention to (matched) control variances of the propensity score index) (see [Appendix Table 2](#) in Supplemental Materials found at <https://doi.org/10.1016/j.jval.2021.01.014>).<sup>39</sup> SMDs of <0.25, Rubin's B <25, and Rubin's R between 0.5 and 2 indicate sufficient balance between the 2 groups. The PSM technique with the lowest values on these performance indicators was chosen and compared to the covariate balance before PSM.

Furthermore, we produced a doubly robust estimation to further reduce confounding by fitting generalized linear regression models to the complete cases and including the potential confounders and weights from the PSM in the regression.<sup>40</sup> In the regressions with QALYs as dependent variable, we also included patients' baseline utility as suggested in the literature.<sup>41</sup> The distribution family and link function in each regression were selected based on goodness-of-fit using the Akaike's information criterion and the Bayesian information criterion.<sup>42-44</sup> All statistical analyses were performed in R, an open-source statistical programming environment (R Core Team 2016).<sup>45</sup>

### Cost-Utility Analysis

In the CUA, the cost of care related to PC+ and hospital care were used. QALYs were calculated using the area under the curve approach based on the EuroQol 5-dimensional questionnaire (EQ-5D-5L) and Dutch value set.<sup>41,46</sup> After performing the doubly robust estimation by using generalized linear regression models, the differences between the intervention and control group were expressed by calculating the incremental cost-effectiveness ratios (ICERs).<sup>47</sup>

### Multi-Criteria Decision Analysis

The MCDA included all criteria (ie, outcomes) presented in [Table 1](#) except of the EQ-5D-5L, since patients' health and well-being was already covered by the SF-12v2. Including the EQ-5D-5L next to the SF-12v2 would cause overlap of criteria and therefore double counting.<sup>48</sup> Cost calculation and doubly robust estimation was performed similar to the CUA. As the criteria were

measured on different scales, we standardized them using the relative standardization method (details on this method are presented in [Appendix 1](#) in Supplemental Materials found at <https://doi.org/10.1016/j.jval.2021.01.014>). Criteria weights (ie, the relative importance of each outcome) were retrieved from the SELFIE study, which estimated weights of Dutch stakeholders for new models of care.<sup>26,49</sup> The weights for all criteria were derived from the exact same criteria used in the SELFIE study except the criterion access of care for which we used the weight of the "continuity of care" criterion in the SELFIE study. Finally, the standardized performance scores were combined with the weights and converted into an overall value for the PC+ intervention and HBOC according to the value-based Multi-Attribute Utility Theory (MAUT) method, using a weighted sum approach.<sup>48,50,51</sup>

### Sensitivity Analyses

In both the CUA and MCDA, probabilistic sensitivity analysis using Monte Carlo simulation, with 5000 replications, was performed to address second-order uncertainty in the results.<sup>52,53</sup> Regarding the CUA, the confidence interval around the ICER was calculated using bootstrapped estimations of the mean cost and QALY differences.<sup>54</sup> This was graphically presented using a cost-effectiveness acceptability curve (CEAC).<sup>55</sup> The uncertainty in the MCDA was graphically presented using a conditional multi-attribute acceptability curve (CMAC).<sup>26</sup> In contrast to the CEAC, the probability of PC+ being cost-effective in the CMAC is based on the various included outcomes in the MCDA and shows the probability of PC+ to be accepted as the preferred alternative (ie, having the highest overall value), while the budget remains below a set threshold. This threshold refers to the budget available to be allocated to either PC+ or HBOC, for the treatment of a given population.

Moreover, we performed multiple imputation using multivariate imputation by chained equations, also known as fully conditional specification, to assess the impact of missing observations on the results of the CUA and MCDA.<sup>56</sup> Furthermore, in a univariate sensitivity analysis, we used the Short-Form Health Survey with 6 dimensions (SF-6D) utilities to calculate QALYs instead of the EQ-5D-5L utilities used in the main analysis.<sup>57,58</sup> This was done to make the results of the CUA more comparable to the results of the MCDA where quality of life was measured with the SF-12v2. Finally, local ranging standardization of performance scores was used, instead of relative standardization, to investigate the impact of the standardization method on the MCDA results (details on this method are presented in [Appendix 1](#) in Supplemental Materials found at <https://doi.org/10.1016/j.jval.2021.01.014>).

## Results

In total, 1783 patients from the intervention group and 272 from the control group with no missing observations in outcome and confounding variables were included. Before PSM, patients referred to PC+ had better physical health status, and the distribution of the referred medical departments was uneven between the 2 groups ([Table 2](#)).

### Selection of PSM Technique and Complete Cases

Nearest neighbor optimal matching with a 2:1 ratio was the best PSM technique, as it resulted in a high number of respondents with the least covariate imbalance (SMD >0.25) and with acceptable Rubin's B (0.8%) and Rubin's R (0.984) (see [Appendix Table 2](#) in Supplemental Materials found at <https://doi.org/10.1016/j.jval.2021.01.014>).

**Table 2.** Baseline characteristics before and after propensity score matching.

	Before PSM				After PSM (complete cases)			
	PC+	HBOC	P value	SMD	PC+	HBOC	P value	SMD
N	1783	272			340	190		
Age – mean (SD)	54.87 (15.96)	56.09 (16.00)	.239	0.077	58.03 (15.41)	59.80 (13.95)	.190	0.121
Gender (male) % (n)	38.6 (689)	34.9 (95)	.268	0.077	33.8 (115)	34.7 (66)	.907	0.019
Foreign-born % (n)	3.7 (66)	5.1 (14)	.327	0.070	5.9 (20)	5.3 (10)	.920	0.027
Low educational level % (n)	19.1 (340)	22.1 (60)	.281	0.074	23.2 (79)	23.7 (45)	.992	0.011
SF-12v2 PCS – mean (SD)	47.65 (9.26)	45.79 (9.94)	.002*	0.194	45.22 (9.53)	45.10 (9.94)	.891	0.012
SF-12v2 MCS – mean (SD)	50.99 (9.42)	50.13 (9.05)	.158	0.093	49.13 (10.16)	50.12 (9.15)	.266	0.102
Historical healthcare costs in € – mean (SD)	761.4 (2767.07)	811.7 (3069.00)	.783	0.017	772.22 (3239.36)	897.91 (3473.91)	.677	0.037
Medical specialty referred to			<.001**	0.731*			.813	0.191
Dermatology % (n)	31.8 (567)	12.1 (33)			7.9 (27)	11.1 (21)		
Gynecology % (n)	5.8 (104)	9.6 (26)			9.7 (33)	8.4 (16)		
Internal medicine % (n)	2.5 (44)	9.2 (25)			8.2 (28)	7.9 (15)		
Otorhinolaryngology % (n)	17.0 (303)	12.5 (34)			11.5 (39)	15.3 (29)		
Neurology % (n)	7.7 (138)	4.8 (13)			4.7 (16)	4.2 (8)		
Ophthalmology % (n)	7.9 (141)	5.9 (16)			5.9 (20)	6.8 (13)		
Orthopedics % (n)	19.3 (344)	39.7 (108)			44.7 (152)	40.5 (77)		
Rheumatology % (n)	6.6 (118)	3.7 (10)			3.8 (13)	3.7 (7)		
Urology % (n)	1.3 (24)	2.6 (7)			3.5 (12)	2.1 (4)		

HBOC indicates hospital-based outpatient care; MCS, mental component summary; PC+, Primary Care Plus; PCS, physical component summary; PSM, propensity score matching; SD, standard deviation; SF-12v2, Short-Form Health Survey version 2; SMD, standardized mean differences.

\* $P < .01$ ; \*\* $P < .001$ ; SMD  $> 0.25$ .

1016/j.jval.2021.01.014). Following PSM, 544 patients from the intervention group and 272 from the control group were matched. However, in this matched sample, complete data on all outcome measures (ie, criteria) during the 3-month follow-up were available from 530 patients (65.0%), with 340 patients (62.5%) in the PC+ group and 190 patients (69.9%) in the HBOC group (see Appendix Table 3 in Supplemental Materials found at <https://doi.org/10.1016/j.jval.2021.01.014>). As shown in Table 2, the differences in baseline characteristics between the 2 groups with complete cases were mitigated after PSM.

The illustrated graphics of the results of the PSM in terms of propensity score distributions and covariate balance are presented in Appendix Figure 1 and Appendix Figure 2 in Supplemental Materials found at <https://doi.org/10.1016/j.jval.2021.01.014>. Regarding the doubly robust estimation using generalized linear regression models, the model performance based on the Akaike's information criterion and Bayesian information criterion is presented in Appendix Table 4 and Appendix Table 5 in Supplemental Materials found at <https://doi.org/10.1016/j.jval.2021.01.014>.

### Results of the CUA

The results of the CUA are presented in Table 3 and show that PC+ led to a cost reduction of €259.04 (95% CI –447.03 to –71.05) and QALY gain of 0.002 (95% CI –0.004 to 0.007), indicating therefore that PC+ was the dominant strategy.

The cost-effectiveness (CE) plane with the bootstrapped ICERs (see Appendix Figure 3 in Supplemental Materials found at <https://doi.org/10.1016/j.jval.2021.01.014>) shows that the majority of the bootstrapped ICERs (86.4%) are located in the southeast quadrant of the plane, indicating that PC+ is, besides less costly, also more effective compared to HBOC. However, a small proportion of points (13.6%) is located in the southwest quadrant, indicating that PC+ is less effective compared to HBOC. The CEAC

(Fig. 1) shows that the probability of PC+ being cost-effective decreases from 100% to 86.6% when the willingness to pay per QALY increases to €50 000, as a small proportion of bootstrapped ICERs was located in the southwest quadrant.

### Results of the MCDA

Aggregated performance for PC+ and HBOC is presented in Table 4. Psychological well-being was considered the most important criteria, with a stakeholders' mean weight of 0.308. Regarding the total scores, calculated by aggregating the average stakeholders' weights and standardized performance scores, the largest differences were found in the scores on accessibility (0.140 vs 0.126) and total costs (0.094 vs 0.061). In line with the result of the CUA, the overall total score shows that PC+ outperforms HBOC (0.730 vs 0.681). The unstandardized performance scores, the weights of the criteria from the viewpoint of the different stakeholders, and the value scores with standardized performance based on relative scaling from these different perspectives are presented in Appendix Tables 6, 7, and 8 in Supplemental Materials found at <https://doi.org/10.1016/j.jval.2021.01.014>.

The CMAC (Fig. 2) shows that the probability of PC+ being effective and affordable increases to 100% at a budget of €5 600 000 for the 3-month outpatient and inpatient hospital costs of 10 740 people eligible to be referred to PC+, or €521.42 per patient for the same period.

### Results of Univariate Sensitivity Analyses

The results of the CUA and MCDA on the imputed data set were similar to the results of the main analysis. However, in both analyses there was slightly more uncertainty regarding the effectiveness of PC+. This was illustrated by a higher number of bootstrapped ICERs (20.4%) in the southwest quadrant of the CE plane and a higher budget requested (€563.31 per patient) for

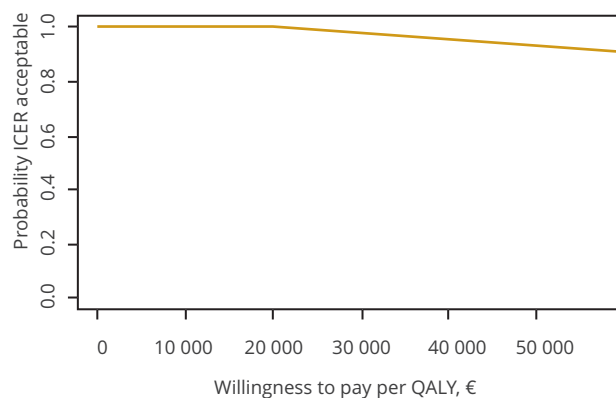


**Table 3.** Incremental cost-effectiveness ratio (ICER).

Alternative	Total cost (mean, SD)	Total QALY (mean, SD)	Incremental cost (95% CI)	Incremental QALY (95% CI)	ICER
PC+	490.34 (211.22)	0.191 (0.034)	−259.04 (−447.03 - −71.05)	0.002 (−0.004 - 0.007)	Dominant*
HBOC	763.54 (287.21)	0.188 (0.033)			

CI indicates confidence interval; HBOC, hospital-based outpatient care; ICER, incremental cost-effectiveness ratio; PC+, Primary Care Plus; QALY, quality-adjusted life-years; SD, standard deviation.

\*Dominant = less costs, better outcomes.

**Figure 1.** Cost-effectiveness acceptability curves showing the probability that PC+ is cost-effective when compared with care-as-usual over a range of willingness-to-pay for an additional QALY.

ICER indicates incremental cost-effectiveness ratio; PC+, Primary Care Plus; QALY, quality-adjusted life-year.

PC+ to be effective and affordable on the CMAC (see Appendix 2 in Supplemental Materials found at <https://doi.org/10.1016/j.jval.2021.01.014>). When SF-6D utility scores were used in the CUA instead of EQ-5D-5L utility scores, PC+ led to −0.000 (95% CI −0.003 to 0.002) less QALYs, illustrated with a majority of bootstrapped ICERs (79.1%) in the southwest quadrant of the CE plane (see Appendix 3 in Supplemental Materials found at <https://doi.org/10.1016/j.jval.2021.01.014>). Finally, although the overall value scores were lower when using local ranging standardization instead of relative standardization in the MCDA, PC+ was still the most preferred alternative (see Appendix 4 in Supplemental Materials found at <https://doi.org/10.1016/j.jval.2021.01.014>).

## Discussion

This study demonstrated a side-by-side performance of CUA and MCDA to evaluate a new model of care, PC+, for non-acute and low-complex patients and found comparable results. Both methods of economic evaluation showed that PC+ was very likely to be cost-effective compared to HBOC. Although the decision suggested by the CUA and MCDA was the same, this study highlights a number of interesting dissimilarities and potential synergies between the 2 methods.

First, in the CUA the opportunity costs of 1 unit of additional benefit, in this case 1 additional QALY, is estimated. In the MCDA, a new composite score of benefit is created using the scores on all included criteria.<sup>26</sup> Therefore, opportunity costs need to be calculated determining 1 unit of additional benefit of that composite score. By including costs as a criterion, it is argued that the opportunity costs are not addressed adequately in the MCDA;

therefore, costs should not be included.<sup>59–62</sup> However, an argument in favor of including costs in the MCDA, just like the other criteria, is that the relative contribution of costs to the decision-making process is made more explicit.<sup>51</sup> Additionally, when implementing new models of care, reducing costs is one of the aims on new models of care, alongside improving quality of and access to care. Furthermore, if costs of a new model of care reflect the value of the displaced interventions (eg, care-as-usual), then opportunity cost is incorporated in the decision. Therefore, incorporating costs as a criterion in MCDA of new models of care may be useful in local decision making.

Although CUA incorporates a multi-attribute measurement of outcome (ie, longevity and quality of life), MCDA can incorporate many more outcomes and has the ability to decompose them to support multifactorial decisions. Such decisions may be applied more frequently because of the importance at the local level where the decision is closer to the local needs and interests of several stakeholders. This is evident in the fact that local decision makers are using disaggregated outcomes (eg, key performance indicators) similar to a CCA.<sup>63</sup> In addition, MCDA can provide a systematic ranking of several alternative strategies that local decision makers may consider to optimize budget allocation.<sup>24</sup> However, multiple comparisons may be more complex in a CUA framework.

Furthermore, the CMAC seems to be a more suitable tool than CEAC to support local decision makers. This may be because it uses a wider range of outcomes, beyond the QALY, relevant for decision making, such as equity, patient satisfaction, and access to care considerations, without requiring an additional axis and without increasing the complexity of the interpretability of the results. Additionally, the CMAC uses the budget available to be allocated across different models of care for a specific population instead of a range of willingness-to-pay thresholds for a QALY, which is less relevant to health and social care budget holders at local level.

Our study showed that the results of the CUA may be sensitive to utility measures that may even alter the decision. Using SF-6D utilities instead of EQ-5D-5L utilities changed the decision in our case study from dominant to cost-saving at lower QALYs. This change may be caused by the fact that the SF-6D is more efficient in detecting small changes in health-related quality of life.<sup>64</sup> The sensitivity of CUA results to the selection of the utility measure is evident in the literature<sup>65</sup> and may make CUA less attractive to local decision makers as they often cannot determine the most appropriate utility measure to be routinely collected for the whole population in their catchment area. Additionally, the MCDA results showed little difference between PC+ and HBOC in the score on physical functioning and psychological well-being. The MCDA outcome of PC+ outperforming HBOC was driven mainly by accessibility of care and the total costs.

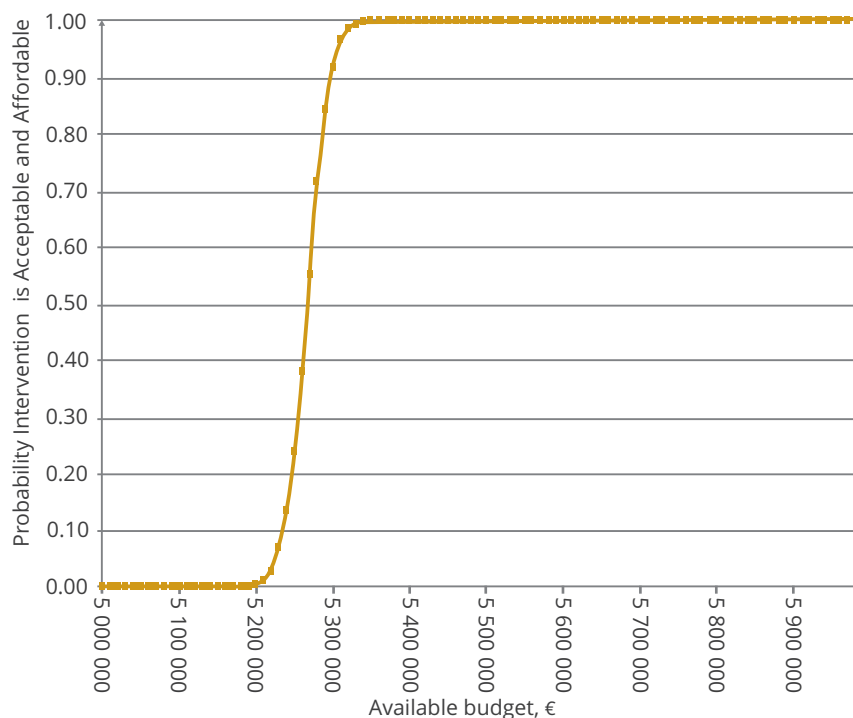
Although MCDA is an exhaustive, flexible, and inclusive approach with a growing popularity to evaluate interventions in healthcare, it also has some drawbacks.<sup>66,67</sup> First, the subjectivity of the weighting approach in the MCDA, which is generally based

**Table 4.** Aggregated weights and standardized performance.

Criteria	Standardized performance* score (SE)		Stakeholders' mean weight (SE)	Total score	
	PC+	HBOC		PC+	HBOC
Physical functioning	0.706(0.006)	0.708 (0.008)	0.248 (0.013)	0.175	0.175
Psychological well-being	0.710 (0.006)	0.704 (0.007)	0.308 (0.011)	0.219	0.217
Person-centeredness	0.707 (0.001)	0.708 (0.002)	0.144 (0.007)	0.102	0.102
Accessibility of care*	0.742 (0.004)	0.670 (0.006)	0.188 (0.013)	0.140	0.126
Total costs*	0.841 (0.020)	0.540 (0.015)	0.112 (0.014)	0.094	0.061
Overall total score				0.730	0.681

HBOC indicates hospital-based outpatient care; PC+, Primary Care Plus; SE, standard error.

\*Standardized performance based on relative standardization with accessibility of care and total costs being reverse coded with a lower standardized score referring to a higher nonstandardized score.

**Figure 2.** Conditional multi-attribute acceptability curve showing the probability of PC+ to be accepted as the preferred alternative over a range of available budgets.

PC+ indicates Primary Care Plus.

on human judgment, can be subject to bias.<sup>67</sup> Together with the ad hoc aggregation of outcomes, the results of the MCDA are context-dependent and therefore less generalizable to other situations. However, it should be mentioned that the QALY is also a preference-based health state classification system, using preferences from a general population sample.<sup>68</sup> Furthermore, it is questionable to what extent generalizability of results is desirable in decision making in a local context, although it would be beneficial when local decision makers set a core set of criteria and elicit the weights, which can be reused.

Additionally, compared with CUA, MCDA is often described as a complex or burdensome process.<sup>49,69,70</sup> This is related to some practical issues that might arise when using MCDA, such as the need to learn relevant techniques to perform MCDA or to have a

facilitator to help using these techniques, the need to expand data collection to assess the full range of outcomes, and the need for relevant software or programs to analyze data and translate the model output in recommendations.<sup>15,71</sup> However, local decision makers are increasingly using decision support tools (eg, balanced score cards and analysis of key performance indicators) with similar data and analytical needs with MCDA, which facilitate its feasibility and applicability at the local level.

Furthermore, where the use of a CUA is straightforward, the MCDA is flexible to incorporate multiple objectives and to allow for the different goals between interventions and places. Therefore, the MCDA facilitates dialog and forces decision makers to think about and clearly express relevant values.<sup>15,50</sup> However, this flexibility is also challenging since many MCDA methods are

available.<sup>66,72</sup> The differences of these methods are not only related to practical issues, but also to the underlying fundamental theories and beliefs.<sup>73</sup>

To overcome these practical challenges and to prevent the misuse of MCDA, specific attention for the use of the MCDA in the evaluation of new models of care in a local context is needed.<sup>73,74</sup> Therefore, more interaction between researchers and decision makers is needed to deliberate on best approaches. Furthermore, MCDA can be of added value to the widely used CUA. With this, informed allocation of limited resources to improve local population outcomes and promote sustainability of local healthcare can be ensured.

### Limitations

Regarding the case study used in this research, some limitations exist related to the design of the study, such as the absence of costs related to primary care, the relatively short follow-up period of 3 months, and the small study sample (especially the control group). Therefore, comparing CUA and MCDA using other case studies without these limitations could be beneficial. However, this case study incorporated all necessary methodological steps, and therefore it serves as a valuable example. Moreover, in local decision making, routinely collected data and non-randomization is common; therefore, other case studies will almost certainly face similar limitations.

Furthermore, we performed PSM to reduce observed confounding. However, because of the limited time points of measurement available, it was not possible to adjust for unobserved confounding (eg, by performing a difference-in-difference analysis). We also acknowledge that potential unobserved bias may have been transferred to the imputed data, although we expect unobserved confounding to be minimal as this was a relatively healthy population and the results were adjusted for many observed confounders.

In addition, we used the criteria weights of the SELFIE study, which included people with more complex healthcare needs (ie, people with multimorbidity) than the population included in our case study. Although this was not ideal, we expect that it had little impact on the total performance scores in the MCDA as (1) the criteria in the SELFIE study were defined in terms of general outcome concepts grouped by the Triple Aim, which were very similar to our criteria and allowed us to use slightly different indicators to measure same outcomes as in SELFIE, and (2) the SELFIE criteria weights are likely to be generalizable to the national level as they were derived from approximately 750 Dutch stakeholders including patients, partners of patients, providers, payers, and policy makers from across The Netherlands.<sup>75</sup>

Finally, as mentioned before, different MCDA methods that are based on different theories can be used in the evaluation of new models of care. In this case study, the value-based MAUT method was used.<sup>51</sup> Besides different value-based methods, other categories of MCDA methods are the outranking, reference level, and goal programming methods. Using different MCDA methods may lead to different decision outcomes. However, regardless of the method used, the main goal of the MCDA is to structure the decision and to support the assessment process.

### Conclusions

Regardless of the decision outcome, the MCDA is of added value to the CUA in the evaluation of new models of care as it allows the incorporation of relevant outcomes and perspectives of involved stakeholders in local decision making. This may result in a more informative, auditable, and transparent decision-making process at the local level. However, because of the practical

challenges related to the use of the MCDA approach, the question is to what extent it will be embedded in the existing local decision-making process. Therefore, further formalization and validation of the MCDA approach in the evaluation of new models of care in a local context is required.

### Supplemental Material

Supplementary data associated with this article can be found in the online version at <https://doi.org/10.1016/j.jval.2021.01.014>.

### Article and Author Information

**Accepted for Publication:** January 14, 2021

**Published Online:** April 20, 2021

doi: <https://doi.org/10.1016/j.jval.2021.01.014>

**Author Affiliations:** Department of Health Services Research, Care and Public Health Research Institute (CAPHRI), Faculty of Health Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands (van den Bogaart, Kroese, Spreeuwenberg, Ruwaard); Research Center for Technology in Care, Zuyd University of Applied Sciences, Heerlen, The Netherlands (Spreeuwenberg); Health Economics Research Centre, Nuffield Department of Population Health, University of Oxford, Oxford, UK (Tsiachristas).

**Correspondence:** Ms E. H. A. van den Bogaart, MSc, Department of Health Services Research, Care and Public Health Research Institute (CAPHRI), Faculty of Health Medicine and Life Sciences, Maastricht University, Duboisdomein 30, Maastricht 6229 GT, The Netherlands. Email: [e.vandenbogaart@maastrichtuniversity.nl](mailto:e.vandenbogaart@maastrichtuniversity.nl)

**Author Contributions:** *Concept and design:* van den Bogaart, Kroese, Spreeuwenberg, Ruwaard, Tsiachristas

*Acquisition of data:* van den Bogaart, Kroese, Ruwaard

*Analysis and interpretation of data:* van den Bogaart, Kroese, Spreeuwenberg, Ruwaard, Tsiachristas

*Drafting of the manuscript:* van den Bogaart, Tsiachristas

*Critical revision of the paper for important intellectual content:* van den Bogaart, Kroese, Spreeuwenberg, Ruwaard, Tsiachristas

*Statistical analysis:* van den Bogaart, Spreeuwenberg

*Provision of study materials or patients:* van den Bogaart, Kroese, Ruwaard

*Obtaining funding:* Kroese, Spreeuwenberg, Ruwaard

*Administrative, technical, or logistic support:* Kroese, Spreeuwenberg, Ruwaard

*Supervision:* Kroese, Spreeuwenberg, Ruwaard, Tsiachristas

**Conflict of Interest Disclosures:** Drs van den Bogaart, Kroese, Spreeuwenberg, and Ruwaard reported receiving grants from nonprofit private health insurers CZ and VGZ via the cooperative "Reinforcing Primary Care South of The Netherlands" (in Dutch: Versterking Eerstelijns Zuid-Nederland) and the Province of Limburg, The Netherlands during the conduct of the study. Dr Tsiachristas acknowledges financial support from the National Institute for Health Research Oxford and Thames Valley Applied Research Collaboration and the National Institute for Health Research Oxford Biomedical Research Centre.

**Funding/Support:** For this study, funding was received from the nonprofit private health insurers CZ and VGZ via the cooperative "Reinforcing Primary Care South of The Netherlands" (in Dutch: Versterking Eerstelijns Zuid-Nederland) and the Province of Limburg, The Netherlands.

**Role of the Funder/Sponsor:** The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

**Acknowledgment:** We acknowledge employees from the Maastricht UMC+ for their contributions in providing us with data. We would also like to acknowledge the employees of Transmural Interactive Patient Platform and Care in Development for their contributions in facilitating to send and receive the patient questionnaires. Finally, Ms van den Bogaart would like to acknowledge the colleagues of the Health Economics Research Centre of

the University of Oxford for supporting her visit and providing useful comments on the paper during a seminar.

## REFERENCES

- Turner A, Mulla A, Booth A, et al. *The international knowledge base for new care models relevant to primary care-led integrated models: a realist synthesis*. 2018.
- NHS. *New Care Models: Vanguard—developing a blueprint for the future of NHS and care services*. London: NHS England; 2016.
- Collins B. *New care models: emerging innovations in governance and organisational form*. King's Fund; 2016.
- Ham C. *Making sense of integrated care systems, integrated care partnerships and accountable care organisations in the NHS in England*. London: The King's Fund; 2018.
- Craig P, Petticrew M. Developing and evaluating complex interventions: reflections on the 2008 MRC guidance. *Int J Nurs Stud*. 2013;50(5):585–587.
- Starling A, Health F. Implementing new models of care: lessons from the new care models programme in England. *Int J Care Coord*. 2018;21(1-2):50–54.
- Robert G, Fulop N. *Perspectives on context: a selection of essays considering the role of context in successful quality improvement*. Health Foundation London; 2014.
- Lewin S, Hendry M, Chandler J, et al. Assessing the complexity of interventions within systematic reviews: development, content and use of a new tool (iCAT\_SR). *BMC Med Res Methodol*. 2017;17(1):76.
- Salas-Vega S, Bertling A, Mossialos E. A comparative study of drug listing recommendations and the decision-making process in Australia, The Netherlands, Sweden, and the UK. *Health Policy*. 2016;120(10):1104–1114.
- Tsiachristas A, Rutten-van Molken MP. Evaluating complex interventions. In: *Handbook Integrated Care*. Springer; 2017:297–313.
- Rabarison KM, Bish CL, Massoudi MS, Giles WH. Economic evaluation enhances public health decision making. *Front Public Health*. 2015;3:164.
- Garrison Jr LP, Kamal-Bahl S, Towse A. Toward a broader concept of value: identifying and defining elements for an expanded cost-effectiveness analysis. *Value Health*. 2017;20(2):213–216.
- Wildman J, Wildman JM. Combining health and outcomes beyond health in complex evaluations of complex interventions: suggestions for economic evaluation. *Value Health*. 2019;22(5):511–517.
- Sefton T. Getting less for more: economic evaluation in the social welfare field. LSE STICERD Research Paper No CASE044. 2000.
- Campbell M, Fitzpatrick R, Haines A, et al. Framework for design and evaluation of complex interventions to improve health. *BMJ*. 2000;321(7262):694–696.
- Craig P, Dieppe P, Macintyre S, et al. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ*. 2008;337:a1655.
- Coast J. Is economic evaluation in touch with society's health values? *BMJ*. 2004;329(7476):1233–1236.
- Palfrey C, Thomas P. *Evaluation for the Real World: The Impact of Evidence in Policy Making*. Policy Press; 2012.
- Garrison Jr LP, Pauly MV, Willke RJ, Neumann PJ. An overview of value, perspective, and decision context—a health economics approach: an ISPOR Special Task Force report [2]. *Value Health*. 2018;21(2):124–130.
- Weir E, d'Entremont N, Stalker S, Kurji K, Robinson V. Applying the balanced scorecard to local public health performance measurement: deliberations and decisions. *BMC Public Health*. 2009;9:127.
- McDonald B. *A review of the use of the Balanced Scorecard in healthcare*. Newcastle: BMCD Consulting; 2012.
- Goetghebeur MM, Wagner M, Khoury H, Levitt RJ, Erickson LJ, Rindress D. Bridging health technology assessment (HTA) and efficient health care decision making with multicriteria decision analysis (MCDA): applying the EVIDEM framework to medicines appraisal. *Med Decis Making*. 2012;32(2):376–388.
- Bots PW, Hulshof JA. Designing multi-criteria decision analysis processes for priority setting in health policy. *JMCDA*. 2000;9(1-3):56–75.
- Baltussen R, Niessen L. Priority setting of health interventions: the need for multi-criteria decision analysis. *Cost Eff Resour Alloc*. 2006;4(1):14.
- Tsiachristas A, Cramm JM, Nieboer A, Rutten-van Molken M. Broader economic evaluation of disease management programs using multi-criteria decision analysis. *Int J Technol Assess Health Care*. 2013;29(3):301–308.
- Rutten-van Molken M, Leijten F, Hoedemakers M, et al. Strengthening the evidence-base of integrated care for people with multi-morbidity in Europe using multi-criteria decision analysis (MCDA). *BMC Health Serv Res*. 2018;18(1):576.
- Schippers E. *Proeftuinen en pilots' betere zorg met minder kosten*. Den Haag: Ministry of Health WaS; 2013.
- van den Bogaart EHA, Spreeuwenberg MD, Kroese M, van den Boogaart MW, Boymans T, Ruwaard D. Referral decisions and its predictors related to orthopaedic care: a retrospective study in a novel primary care setting. *PLoS One*. 2020;15(1):e0227863.
- Munton T, Martin A, Marrero I, Llewellyn A, Gibson K, Gomersall A. *Evidence: Getting Out of Hospital?* London: The Health Foundation; 2011.
- Berwick DM, Nolan TW, Whittington J. The triple aim: care, health, and cost. *Health Aff (Millwood)*. 2008;27(3):759–769.
- Stiefel M, Nolan K. *A guide to measuring the triple aim: population health, experience of care, and per capita cost*. IHI innovation series white paper. Cambridge, MA: Institute for Healthcare Improvement; 2012.
- Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20(10):1727–1736.
- Ware Jr J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care*. 1996;34(3):220–233.
- Delnoij D, Hendriks M, De CQ-index. het meten van klantervaringen in de zorg. *TSG*. 2008;86(8):440–446.
- D'Agostino Jr RB. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med*. 1998;17(19):2265–2281.
- Stuart EA. Matching methods for causal inference: a review and a look forward. *Stat Sci*. 2010;25(1):1–21.
- Baser O. Too much ado about propensity score models? comparing methods of propensity score matching. *Value Health*. 2006;9(6):377–385.
- Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res*. 2011;46(3):399–424.
- Rubin DB. Using propensity scores to help design observational studies: application to the tobacco litigation. *Health Serv Outcomes Res Methodol*. 2001;2(3/4):169–188.
- Funk MJ, Westreich D, Wiesen C, Sturmer T, Brookhart MA, Davidian M. Doubly robust estimation of causal effects. *Am J Epidemiol*. 2011;173(7):761–767.
- Manca A, Hawkins N, Sculpher MJ. Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. *Health Econ*. 2005;14(5):487–496.
- Akaike H. Information theory as an extension of the maximum likelihood principle. In: *Second International Symposium on Information Theory*, BN Petrov, F Csaki (Eds), BNPBF Csaki Budapest: Akademiai Kiado; 1973.
- Schwarz G. Estimating the dimension of a model. *Annals Stat*. 1978;6(2):461–464.
- Claeskens G, Hjort NL. *Model Selection and Model Averaging*. Cambridge Books; 2008.
- R: A language and environment for statistical computing. [computer program]. Vienna, Austria; 2016.
- Lamers LM, McDonnell J, Stalmeier PF, Krabbe PF, Busschbach JJ. The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. *Health Econ*. 2006;15(10):1121–1132.
- O'Brien BJ, Drummond MF, Labelle RJ, Willan A. In search of power and significance: issues in the design and analysis of stochastic cost-effectiveness studies in health care. *Med Care*. 1994;32(2):150–163.
- Marsh K, IJzerman M, Thokala P, et al. Multiple criteria decision analysis for health care decision making-emerging good practices: report 2 of the ISPOR MCDA Emerging Good Practices Task Force. *Value Health*. 2016;19(2):125–137.
- Leijten F, Tsiachristas A, Boland M, et al. *SELFIE 2020 Work Package 4: Development of an analytical framework to perform comprehensive evaluation of integrated care programmes for multi-morbidity using multi-criteria decision analysis*. Rotterdam, The Netherlands: Institute of Health Policy and Management and University of Bergen (UiB), Norway; 2017.
- Thokala P, Duenas A. Multiple criteria decision analysis for health technology assessment. *Value Health*. 2012;15(8):1172–1181.
- Marsh K, Goetghebeur M, Thokala P, Baltussen R. *Multi-Criteria Decision Analysis to Support Healthcare Decisions*. Springer; 2017.
- Broekhuizen H, Groothuis-Oudshoorn CG, Hauber AB, Jansen JP, MJ JJ. Estimating the value of medical treatments to patients using probabilistic multi criteria decision analysis. *BMC Med Inform Decis Mak*. 2015;15:102.
- Broekhuizen H, Groothuis-Oudshoorn CGM, van Til JA, Hummel JM, IJzerman MJ. A review and classification of approaches for dealing with uncertainty in multi-criteria decision analysis for healthcare decisions. *Pharmacoeconomics*. 2015;33(5):445–455.
- Fenwick E, Marshall DA, Levy AR, Nichol G. Using and interpreting cost-effectiveness acceptability curves: an example using data from a trial of management strategies for atrial fibrillation. *BMC Health Serv Res*. 2006;6:52.
- Al MJ. Cost-effectiveness acceptability curves revisited. *Pharmacoeconomics*. 2013;31(2):93–100.
- Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? *Int J Methods Psychiatr Res*. 2011;20(1):40–49.
- Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. *Med Care*. 2004;42(9):851–859.



58. Petrou S, Hockley C. An investigation into the empirical validity of the EQ-5D and SF-6D based on hypothetical preferences in a general population. *Health Econ*. 2005;14(11):1169–1189.
59. Marsh KD, Sculpher M, Caro JJ, Tervonen T. The use of MCDA in HTA: great potential, but more effort needed. *Value Health*. 2018;21(4):394–397.
60. Campillo-Artero C, Puig-Junoy J, Culyer AJ. *Does MCDA Trump CEA?* Springer; 2018.
61. Claxton K. Three questions to ask when examining MCDA. *Value Outcomes Spotlight*. 2015;1(1):18–20.
62. Claxton K, Sculpher M, Palmer S, Culyer AJ. Causes for concern: is NICE failing to uphold its responsibilities to all NHS patients? *Health Econ*. 2015;24(1):1–7.
63. Amelung V, Stein V, Goodwin N, Balicer R, Nolte E, Suter E. *Handbook Integrated Care*. Springer; 2017.
64. Chen J, Wong CK, McGhee SM, Pang PK, Yu WC. A comparison between the EQ-5D and the SF-6D in patients with chronic obstructive pulmonary disease (COPD). *PLoS One*. 2014;9(11):e112389.
65. Joore M, Brunenberg D, Nelemans P, et al. The impact of differences in EQ-5D and SF-6D utility scores on the acceptability of cost-utility ratios: results across five trial-based cost-utility studies. *Value Health*. 2010;13(2):222–229.
66. Marsh K, Lanitis T, Neasham D, Orfanos P, Caro J. Assessing the value of healthcare interventions using multi-criteria decision analysis: a review of the literature. *Pharmacoeconomics*. 2014;32(4):345–365.
67. Ghabri S, Josselin JM, Le Maux B. Could or should we use MCDA in the French HTA process? *Pharmacoeconomics*. 2019;37(12):1417–1419.
68. McDonough CM, Tosteson ANA. Measuring preferences for cost-utility analysis: how choice of method may influence decision-making. *Pharmacoeconomics*. 2007;25(2):93–106.
69. Towse A. Net clinical benefit: the art and science of jointly estimating benefits and risks of medical treatment. *Value Health*. 2010;13(Suppl 1):S30–S32.
70. Guo JJ, Pandey S, Doyle J, Bian BY, Lis Y, Raisch DW. A review of quantitative risk-benefit methodologies for assessing drug safety and efficacy-report of the ISPOR Risk-Benefit Management Working Group. *Value Health*. 2010;13(5):657–666.
71. Tony M, Wagner M, Khoury H, et al. Bridging health technology assessment (HTA) with multicriteria decision analyses (MCDA): field testing of the EVI-DEM framework for coverage decisions by a public payer in Canada. *BMC Health Serv Res*. 2011;11:329.
72. Jit M. MCDA from a health economics perspective: opportunities and pitfalls of extending economic evaluation to incorporate broader outcomes. *Cost Eff Resour Alloc*. 2018;16(Suppl 1):45.
73. Thokala P, Devlin N, Marsh K, et al. Multiple criteria decision analysis for health care decision making—an introduction: Report 1 of the ISPOR MCDA Emerging Good Practices Task Force. *Value Health*. 2016;19(1):1–13.
74. Mullen PM. Quantifying priorities in healthcare: transparency or illusion? *Health Serv Manage Res*. 2016;17(1):47–58.
75. Rutten-van Molken Maureen, Karimi Milad, Leijten Fenna, et al, on behalf of the SELFIE consortium. Comparing patients' and other stakeholders' preferences for outcomes of integrated care for multimorbidity: a discrete choice experiment in eight European countries. *BMJ Open*. 2020;10(10):e037547.