ScienceDirect

Contents lists available at **sciencedirect.com**
Journal homepage: **www.elsevier.com/locate/jval**

Health Policy Analysis

Evaluating Complex Health and Social Care Program Using Multi-Criteria Decision Analysis: A Case Study of "Better Together in Amsterdam North"

Milad Karimi, PhD, Lennart van der Zwaan, MSc, Kamrul Islam, PhD, Joost van Genabeek, PhD, Maureen Rutten-van Mölken, PhD

ABSTRACT

Objectives: Multi-criteria decision analysis (MCDA) has been recommended to support policy making in healthcare. However, practical applications of MCDA are sparse. One potential use for MCDA is for the evaluation of programs for complex and vulnerable patients. These complex patients benefit from integrated care programs that span healthcare and social care and aim to improve more than just health outcomes. MCDA can evaluate programs that aim to improve broader outcomes because it allows the evaluation of multiple outcomes alongside each other. In this study, we evaluate an innovative integrated care program in the Netherlands using MCDA.

Methods: We used an innovative MCDA framework with broad outcomes of health, well-being, and cost to evaluate the Better Together in Amsterdam North (BSiN) program using preferences of patients, partners, providers, payers, and policy makers in the Netherlands. BSiN provides case management support for a period of 6 months. Seven outcomes that previous research has deemed important to complex patients were measured, including physical functioning and social relationships and participation.

Results: We find that the program improved the overall MCDA score marginally, and, thus, after 6 and after 12 months, BSiN was preferred to usual care by all stakeholders. BSiN was preferred to usual care, mostly owing to improvements in psychological well-being and social relationships and participation.

Conclusions: The integrated healthcare and social care program BSiN in the Netherlands was preferred to usual care according to an MCDA evaluation. MCDA seems a useful method to evaluate complex programs with benefits beyond health.

Keywords: complex care, discrete choice experiment, health technology assessment, multi-criteria decision analysis, well-being.

VALUE HEALTH. 2021; 24(7):966–975

Introduction

The care of vulnerable complex patients suffers from fragmentation and a lack of person centeredness.[1-3] Complex patients are those who not only experience multiple (physical and mental) chronic conditions but also functional and cognitive impairments and social and financial problems.[2] These vulnerable patients need complex care that is targeted to more than just improving health and have significant nonmedical needs.[2,4] Integrated care is proposed as a solution and is aimed to achieve high-quality and cost-effective care for complex patients.[5,6] Integrated care is defined as care that is coordinated, proactive, person centered, and provided by 2 or more care professionals. As a response to the needs of complex patients, innovative programs have been established.

One integrated care program for complex patients in the Netherlands is the program "Better Together in Amsterdam North" (Dutch acronym: BSiN). BSiN provides a period of intensive case management. During the case management process, the case manager and the BSiN patient develop a targeted "life action plan," which they aim to execute over 6 months. This case management approach ensures that all the various care providers are well aligned and that the individual is seeking the right type of care from the right services. The BSiN program is made up of an alliance of 12 healthcare and social care provider organizations, representing primary healthcare (eg, general practitioners), secondary healthcare (hospital), mental health services, welfare (debt services case managers and social workers), and social security (municipality return-to-work coordinators, home-care services). Thus, what is particularly unique about BSiN is the emphasis on problems in multiple domains such as health, financial, or housing problems.

The evidence of the effectiveness and cost-effectiveness on integrated care for vulnerable people with complex needs is sparse and mixed[7] making implementation, upscaling, or continuation of these programs difficult. Evaluating integrated care programs, such as BSiN, is challenging because these are complex multi-faceted interventions[8] that aim to improve the "Triple Aim":

(1) health and well-being, (2) experiences with care, and (3) costs.[8] Owing to this complexity, traditional cost-effectiveness evaluations that focus on health outcomes and costs are not sufficient for the evaluation of integrated care programs.[8] A broad evaluation method needs to be applied that investigates multiple outcomes.[9] In addition, these programs require a joint effort by various stakeholders, such as patients, informal care takers, professionals, payers, and policy makers, and, thus, all their views should be considered.[9]

Multi-criteria decision analysis (MCDA) can provide a decision-making technique for evaluating integrated care because it can include multiple outcomes[10] that cover the Triple Aim. In MCDA, the "criteria" (or outcomes) can be reviewed separately but also integrated into 1 score by applying a weighting to the outcomes.[10] In addition, MCDA can consider multiple perspectives by using different weights, reflecting the different preferences of the different stakeholders.[10] The MCDA process facilitates decision-making about integrated care programs by providing a structured and explicit process to evaluate BSiN. Although MCDA has been recommended as a means to support decision-making, its use remains limited.[11] Indeed, there remain questions about the implementation of MCDA particularly around the selection of the right outcomes, methods of weight elicitation, methods for uncertainty analysis, and whether costs should be an outcome.[11] Thus, there remains a gap in the literature for the methodological implementation of MCDA for evaluation programs in healthcare and social care. This article attempts to present a solution to these methodological problems.

A recent European Horizon 2020–funded project has developed a novel MCDA framework, that is, a particular application of MCDA, with the aim to be applicable to evaluate a wide variety of integrated care programs.[9] This study aimed to implement MCDA to assess the value of the BSiN approach as opposed to usual care for those with complex multiple problems and thereby also to contribute to the methodology and use of MCDA in healthcare decision-making.

Methods

Intervention

BSiN implemented a case management intervention in Amsterdam North, in the Netherlands. The population of Amsterdam North has on average a lower socioeconomic status and relatively high healthcare and social care costs than the rest of the city.[2] BSiN assigns a case manager to each individual. Case managers receive special training and develop individualized care plans together with the individuals enrolled in the program. A typical case management trajectory takes 6 months. After 6 months, the case management could be extended but in practice was provided only for 6 months.

Target Population

Individuals with a potential to be eligible for enrolment in BSiN were identified and referred to a multidisciplinary triage team that includes a general practitioner, a psychiatrist or psychiatric nurse, a district nurse, and a social worker. Only those aged more than 18 years were eligible for the intervention. Individuals entering BSiN were screened using the Self-Sufficiency Matrix (SSM)[12], which assesses an individual's self-sufficiency in 11 life domains: finances, daily activities, housing, relationships at home, mental health, physical health, addiction, activities of daily living, the social network, social participation, and justice.[12] Individuals with limited self-sufficiency or lower on at least 3 of the 11 life domains were assigned to case management.

The control group could not be based on randomization of the intervention population because of ethical reasons. Instead, a control group was recruited among residents aged between 18 and 65 years in Amsterdam who filled out the "Amsterdam Health Monitor Survey" in 2012 or 2016.[13] In the survey of 2012 and 2016, individuals with negative outcomes on the social exclusion index or on the psychological distress scale[14,15] were identified as potentially eligible control group participants. All eligible potential control group participants were requested to take part in a face-to-face interview where the actual SSM was determined by a trained interviewer. If the individual scored limited self-sufficiency on at least 3 of the 11 SSM domains, they were included in the control group and the first measurement for the control group participant was undertaken. Using this method, it was possible to form a control group while complying with medical ethics. Because the control group could not be randomized, statistical techniques were applied to ensure that the intervention and control were more similar on observable characteristics (see the Statistical Analysis section).

MCDA Framework and Decision Context

The MCDA framework was developed around established guidelines and follows the 7 recommended steps (1) establish the decision context, (2) identify and structure criteria, (3) determine the performance on criteria, (4) determine the weights of the criteria, (5) create an overall value score, (6) perform sensitivity analyses, and (7) interpret results.[10,16] In the first step, the aim is to establish what the likely decisions are that need to be made and thus how the MCDA will be used.[10,16] Earlier analysis of the decision context of BSiN revealed that the program context is to provide evidence on the effectiveness of the program, to help establish the long-term sustainability, and potentially widen implementation of the program in the region.[17,18] Hence, the aim of the MCDA was to inform these decisions by comparing the program with usual care. In addition, it is important to identify the relevant stakeholders in this decision-making process. The stakeholders that were considered relevant were 5 groups: patients, partners and other informal caregivers, professionals, payers, and policy makers.[17,18]

Outcomes and Preferences

In the Sustainable Integrated Care Models for Multimorbidity: Delivery, Financing, and Performance (SELFIE) project, we selected a set of 8 outcomes that measure the Triple Aim[10]: improving population health and well-being, improving experience with care, and reducing costs (growth). These 8 outcomes are chosen to be applicable to a wide range of interventions for complex patients. This selection was based on a literature review, workshops with representatives from the 5 stakeholder groups, and focus groups with individuals with multimorbidity.[17] To measure health and well-being, we included outcomes on physical functioning, psychological well-being, social relationships and participation, enjoyment of life, and resilience. To measure experience of care, we included outcomes on person centeredness and continuity of care. Finally, to measure costs we included program intervention costs and total healthcare cost of each individual. The inclusion of cost as an outcome implies that the opportunity costs of resources are determined by those providing the preferences weights[16,18] and not by a (nationally) determined threshold similar to a quality-adjusted life year threshold. This can be justified because the decision context of BSiN and many other integrated care programs is whether to continue or roll out programs. The decision maker in this context

Table 1. Outcomes measured in MCDA.

Criteria	Item	Scale
Physical functioning	10 ADL questions of the "Groningen Activity Restriction Scale": Eating and drinking Sitting and standing up from a chair Getting in and out of bed Getting dressed and undressed Moving within the house Moving up and down stairs Entering and existing the house Moving outside the house Washing and face and hands Washing yourself completely	Each item has 4 response options: 1 = only with help of others 2 = with great effort 3 = with some effort 4 = without effort
Psychological well-being	Over the past year, did you ever feel gloomy or sad?	The item has 5 response options: 1 = always 2 = often 3 = sometimes 4 = once in a while 5 = never
Social relations and participation	Do you ever feel alone or lonely?	The item has 3 response options: 1 = yes 2 = more or less 3 = no
Resilience	If something bad happens, can you handle it well?	The item has 3 response options: 3 = able to deal with setbacks 2 = neutral 1 = not able to deal with setbacks
Person centeredness	Three items relating to planning with a case manager: 1. Is the plan available? 2. Is the plan read? 3. Is the plan understood?	Each item has 2 response options: 1 = yes 0 = no
Continuity of care	Do you have 1 care provider whom you can approach with all your problems?	The item has 2 response options: 1 = yes 0 = no
Total costs	Total cost (including primary care, secondary care, and medication) as part of the basic Dutch insurance package	Outcome measured in € between €250 a month and €2083 a month

ADL indicates activity of daily living; MCDA, multi-criteria decision analysis.

is more interested in whether a particular program generates sufficient benefits over an alternative program to continue investing in it rather than whether investment should be made. One benefit of including cost as an outcome is that the additional investment is directly compared with the additional non-cost outcomes, thus incorporating the justification of any investments. Furthermore, because reduction in costs (growth) is one of the aims of integrated care programs, it should be considered alongside other outcomes.

A questionnaire was developed by the Netherlands Organisation for Applied Scientific Research to collect data on the health/well-being and care experience with measurements at baseline, at 6 months, and at 12 months. This questionnaire consisted of a range of existing patient reported validated items (eg, the Groningen Activity Restriction Scale[19] and the Vita-16[20]) and was administered by trained interviewers at the participants' home. Interviews lasted 45 to 60 minutes. All items were piloted before being used in interviews and adapted where needed. A list of outcomes is presented in Table 1.

Cost data were obtained from claims data from a Dutch health insurance firm "Achmea Zilveren Kruis." The company claims database of the Achmea was linked using personal data to data from individuals in the study through a trusted third party (to protect the identity of participants). The claims data contain healthcare cost data from the basic Dutch health insurance. Data on expenditure from individuals on medicine, mental health, primary care, and secondary care are included in this dataset. These claims data do not include social care services because they are not paid for by the insurance company. In addition, cost data from the organizations providing case management were available in the reports of the BSiN alliance of healthcare and social care provider organizations.

For each outcome, the relative importance of each outcome for each stakeholder group was measured using a discrete choice experiment (for full details of how these preference weight were obtained, see[9,21]) with the 5 relevant stakeholder groups in the Netherlands. In the discrete choice experiment, individuals belonging to these stakeholder groups were asked to make choices among programs with different performances on each outcome. Each program was described in terms of 8 outcomes where each outcome had 3 levels of performance. Thus, respondents selected which program they thought would be better for someone with multimorbidity and thereby revealed what outcome was most important to them. Patients, partners, and providers were primarily recruited from an online panel whereas payers and policy makers were recruited using networks of the

study authors and a snowball sampling method. By analyzing their choices using a multinomial logit model, the relative importance of each outcome was obtained.[9,21] Thus, these preference weights reflect the difference in preference between a good performance and poor performance on each outcome.

Statistical Analysis of Performance Score Calculation

In MCDA, for each different outcome that was measured (eg, physical), the performance score (PS) (ie, how patients rate their physical functioning) must be estimated to assess how well each program performed. The performance of BSiN compared with usual care at 6 months and 12 months was estimated. Because there was no randomization and the control and intervention groups were recruited in different ways, the participants in the 2 groups were likely to differ in relevant characteristics. To increase comparability of characteristics in the 2 groups and reduce confounding, we performed inverse probability weighting.[22,23]

Inverse probability weighting was used in an attempt to rebalance the composition of the control group to match that of the intervention group by weighting those in the control group most similar to the intervention group more heavily.[22,23] To calculate the inverse probability weighting, we included the following variables: gender, age, education, living situation, marital status, financial problems, work status, volunteer work, housing problems, number of chronic diseases, number of SSM domains on which the score is less than 3, and self-perceived health. This results in a weight that reflects the estimated probability that someone would be in the intervention group. We estimate the average treatment effect on the treated by setting the weights for each individual in the intervention group to 1 and for each individual in the control group to p/(1-p), with p being the estimated probability that an individual is in the intervention group. To assess whether the inverse probability weighting improved the comparability of the intervention and control group, a set of matching statistics were used. Three matching statistics were calculated: Rubin's B (the standardized difference of the means of the linear index of the propensity score in the treated and non-treated group), Rubin's R (the ratio of treated to non-treated variances of the linear index of the propensity score), and the median absolute standardized bias (which is the median of the ratios of the difference of the sample means in the treated and non-treated groups over the square root of the average of the variances in both groups).[24]

To analyze the performance on outcomes, we used repeated measurement models with individual-level random effects, to account for the fact that individuals were followed up over time, and we used an intention to treat analysis. We use models assuming continuous outcomes for ease of interpretation, only for continuity of care did we use a binary logit regression to constrain the predicted values to a 0 to 1 range. The following regression was estimated for each outcome separately:

$$Y_{jt} = \beta_0 + \beta_{1*} C_n + \delta_t * T_t + \gamma_t * T_t * C_n + u_j + e_{jt}$$

where Y_{jt} is the outcome for person j at time t, is the random intercept variance term, C_n indicates whether the observation is from the intervention or control cohort, and T_t indicates whether the observation is from time point 0, 6, or 12 months. β and δ are a set of regression coefficients. The treatment effects are given by the regression coefficients γ_t. It should be noted that all outcomes were included in the MCDA regardless of statistical significance, because a treatment effect that is not statistically significant is not zero,[25] and because later described MCDA uncertainty analysis takes the uncertainty in treatments effects into consideration.

To calculate PSs, we predict the mean score of the intervention group at 6 months and at 12 months based on the regressions results. In addition, we calculate the mean score of the control group assuming they had the same baseline as the intervention group. In this way, the calculated PSs can be directly compared between the intervention and control group. This is done separately for each outcome.

$$\text{Performance score}_{nt} = \beta_0 + \beta_1 * C_{n=1} + \delta_t * T_t + \gamma_t * T_t * C_n$$

To incorporate the performance scores with weights, we first normalize the PSs to a range of 0 to 1. This removes the difference in scales among all PSs. Normalization is done by:

$$\text{Normalized performance scores} = \frac{\text{performance scores} - \text{minimum score}}{\text{maximum score} - \text{minimum score}}$$

We take the range of each outcome as the minimum and maximum. For costs, we took a range based on the existing costs of care for this group of patients and program cost from the literature (€250–€2083 a month). We then multiplied the outcome weight of stakeholder group with the normalized performance to obtain the partial value, and we sum the partial value to obtain the overall scores.

Overall Value Calculation

Finally, to combine preference weights and PSs of each outcome into an overall value score considering all outcomes, we use a weighted sum approach:

$$\text{Overall score} = \sum \text{Weight}_c \text{Normalized performance score}_c$$

Probabilistic sensitivity analysis (PSA) using Monte-Carlo simulation was conducted to assess the joint uncertainty of preference weights and PSs. We use the Cholesky decomposition to draw correlated draws[26] from the distributions of both performance estimates and the weights for 10 000 repetitions. The overall scores are then combined as previously described. From this, we ascertain how often the integrated care program has an overall score better than the usual care, and we obtain the confidence interval (CI) of the overall scores based on the 95% percentile method. All analysis was conducted in Stata 15 (StataCorp. 2017. *Stata Statistical Software: Release 15*. StataCorp LLC, College Station, TX).

Results

Sample Characteristics

The background characteristics of the sample are presented in Table 2. Before inverse probability weighting, there are differences between the groups. The intervention group scores low on SSM with over 6 domains being declared insufficient, indicating that on a wide range of life domains they are not self-sufficient. In general, the control group has better baseline outcomes, for example, the control group scores better on the SSM with less domains in crisis and has less financial problems. After inverse probability weighting, many of these differences are reduced and indeed the matching statistics of Rubin's B and median absolute standardized

Table 2. Descriptive statistics of baseline background characteristics and outcomes.

Variables	Before matching		After matching
	Intervention	Control	Control
Data used for the non-cost outcomes			
Subject characteristics			
Sample size	74	161	
Men, %	35	41	39
Age, mean	53 (13)	58 (11)	58 (9)
High education level, %	31	54	47
Living together, %	59	60	45
Financial problems, %	64	20	65
Paid employment, pension, of student, %	20	46	39
Volunteer work, %	18	29	24
Housing problems, %	9	1	3
Number of chronic diseases	3.5 (2.3)	3.3 (2.3)	4.5 (2.4)
Perceived health	3.6 (0.8)	3.1 (0.8)	3.7 (0.8)
Total SSM score, mean	3.3 (0.4)	4.2 (0.5)	3.6 (0.5)
Number of SSM domains that are vulnerable or in crisis, mean	6.3 (1.8)	2.5 (2)	5.5 (2.3)
Outcomes			
Baseline outcome on physical functioning scale (SD)	2.4 (1.1)	2.7 (1.3)	2.1 (1)
Baseline outcome on psychological well-being (SD)	2.7 (1.2)	3.5 (1.1)	2.5 (1.1)
Baseline outcome on social relations and participation (SD)	1.9 (0.9)	2.4 (0.8)	1.8 (0.9)
Baseline outcome on resilience (SD)	2.3 (0.8)	2.5 (0.7)	1.9 (0.9)
Baseline outcome on continuity of care (SD)	0.9 (0.3)	0.2 (0.4)	0.4 (0.5)
Baseline outcome on person centeredness (SD)	0.8 (1.3)	0.5 (1.1)	0.1 (0.5)
Matching statistics			
Rubin's B	228		104
Rubin's R	1.1		2.4
Median absolute standardized bias	46.5		32.1
Data used for the cost outcome			
Subject characteristics			
N at baseline	46	111	
Men, %	0.39	0.41	0.40
Age, mean	55 (13)	61 (11)	55 (13)
Outcomes			
Baseline total cost in first month	€432.44	€679.88	€410.09
Matching statistics			
Rubin's B	52		5
Rubin's R	0.9		0.68
Median absolute standardized bias	20.9		2.4

SD indicates standard deviation; SSM, Self-Sufficiency Matrix.

bias show this for both the cost and non-cost outcomes. However, Rubin's R values are now further away from the ideal value of 1.

Treatment Effects

The treatment effects are presented in Table 3. At the measurement after 6 months of intervention, we find a significant effect for psychological well-being and the costs outcomes, with BSiN performing better than usual care for psychological well-being but worse for costs. The improvement on psychological well-being indicates that in the first 6-month period, those in BSiN improved a full step on the 5 point scale and felt less "gloomy or sad." On costs the treatment effect is €1368, indicating that BSiN is more costly than usual care. For other outcomes, there are 3 positive treatment effects and 2 negative treatment effects. The treatment effects of physical functioning, social participation, and continuity of care are positive, indicating BSiN has led to an

Table 3. Regression estimates per outcome.

Variable	Physical functioning		Psychological well-being		Social relations and participation		Resilience		Person centeredness		Continuity of care		Total costs	
	Est	95% CI	Est	95% CI	Est	95% CI	Est	95% CI	Est	95% CI	Est	95% CI	Est	95% CI
Constant	2.10	(1.78–2.41)	2.51	(2.01–3.01)	1.74	(1.39–2.1)	1.90	(1.51–2.28)	0.08	(–0.04 to 0.19)	–1.17	(–3.25 to 0.9)	€410.09	(289.98–530.19)
Intervention cohort	0.31	(–0.1 to 0.71)	0.22	(–0.35 to 0.79)	0.11	(–0.3 to 0.52)	0.37	(–0.05 to 0.8)	0.71	(0.38–1.04)	4.93	(1.82–8.03)	€22.36	(–222.37 to 267.09)
Time at 6 mo	–0.27	(–0.53 to –0.01)	–0.36	(-0.9 to 0.18)	0.00	(–0.17 to 0.16)	–0.18	(–0.49 to 0.13)	0.97	(–0.53 to 2.46)	0.72	(–1.21 to 2.64)	€–3.22	(–105.44 to 99)
Time at 12 mo	–0.11	(–0.34 to 0.11)	–0.20	(–0.84 to 0.44)	0.00	(–0.15 to 0.14)	0.15	(–0.14 to 0.44)	0.50	(–0.45 to 1.45)	1.91	(–0.23, 4.05)	€4.11	(–147.82 to 156.04)
Intervention at 6 mo	0.24	(–0.19 to 0.66)	1.00	(0.34–1.66)	0.36	(0.06–0.65)	–0.06	(–0.46 to 0.34)	–0.25	(–1.83 to 1.32)	1.55	(–1.99 to 5.09)	€1367.38	(1189.56 to 1545.2)
Intervention at 12 mo	0.21	(–0.19 to 0.62)	0.48	(–0.24 to 1.2)	0.30	(–0.02 to 0.62)	–0.38	(–0.74 to –0.01)	0.04	(–1.05 to 1.13)	–1.89	(–4.61 to 0.83)	€–11.31	(–206.63 to 184.01)
Observations	427		422		420		416		226		427		471	
Sample size	233		233		234		232		133		235		157	

CI indicates confidence interval; Est, beta coefficient.

improvement on these 3 outcomes. On resilience and person centeredness, there is a negative treatment effect, indicating BSiN has led to a deterioration on these 2 outcomes.

At the third measurement point after 12 months (6 months of intervention and 6 months of follow-up), we only obtained significant negative (worse) results for resilience. For the other 6 outcomes no statistically significant effects were found, with 5 of the treatment effects being positive whereas 1 was negative. Similar to the second measurement, physical functioning, psychological well-being, and social participation showed positive treatment effects. Cost and person centeredness had positive although essentially zero treatment effects, opposite to their 6-month treatment effects. At the second measurements, continuity of care had a negative treatment effect, in opposite to the 6-month measurement.

MCDA Overall Scores

The weights for the SELFIE outcomes elicited from the 5 Dutch stakeholder groups are reported in Figure 1 and in the Appendix I in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.02.007. The predicted scores for the MCDA, their standardized scores, and the weighted scores that are presented in Table 4 (after 6 months) and 5 (after 12 months). After 6 months, BSiN was preferred to usual care in the weighted score for all 5 stakeholders, and this is mostly influenced by psychological well-being and social relationships and participation. In particular at 6 months, it is patients who prefer BSiN over usual care the most and policy makers the least (Table 4).

After 12 months, BSiN is still preferred to usual care, which is again mostly driven by psychological well-being. For patients, the difference in total scores between BsiN and usual care is reduced compared with the difference after 6 months. Thus, patients had a stronger preference for BSiN than usual care at 6 months than at 12 months. For providers, payers, and policy makers, the reverse is true, and their preference for BSiN compared with usual care is improved at 12 months compared with at 6 months (Table 5).

Sensitivity Analysis

The PSA provides the 95% CI for the total value score. There is in all cases an overlap between the CI of the intervention and control. The PSA reveals that the intervention is preferred both after 6 month and after 12 months in at least 55% of the runs. After 6 months for patients in 84% of the MCDA iterations, the integrated care program has a higher overall value score than usual care for patients. This is at the minimum of 55% for policy makers. After 12 months, payers prefer BSiN the most (with 82% of cases) and partners the least (with 65%).

Discussion

This study found that BSiN, an integrated care case management program of 6 months for individuals with complex multiple problems, showed an overall improvement, although the improvement was not large. Most of the improvement was driven by the psychological well-being and some by social relationships and participation. BSiN performed the worst on costs, where there were high costs of the intervention and this had not led to short term savings that could offset those costs. Although all stakeholders preferred BSiN to usual care, it was patients who preferred it the most and policy makers the least. In the 6-month follow-up period, where there was no case management provided, some of the benefit was reduced, but this was also offset by improvements in costs. Overall, BSiN was still preferred by all stakeholders to usual care, which means that according to stakeholders the BSiN program should continue to be implemented.

It is difficult to directly compare the results of BSiN with other programs. Integrated care programs are highly complex interventions, and comparing them with each other presents difficulties owing to population and intervention heterogeneity.[27,28] In addition, there is little standardization of outcomes for these sort of evaluations, and indeed many studies may not even focus or include quality of life outcomes, although many do include care utilization or costs.[27] Overall previous research has shown that case management delivers mixed results.[29,30] Indeed, it could be argued that also in BSiN the results are somewhat mixed, because

Figure 1. Relative preferences of different stakeholders across different outcome.

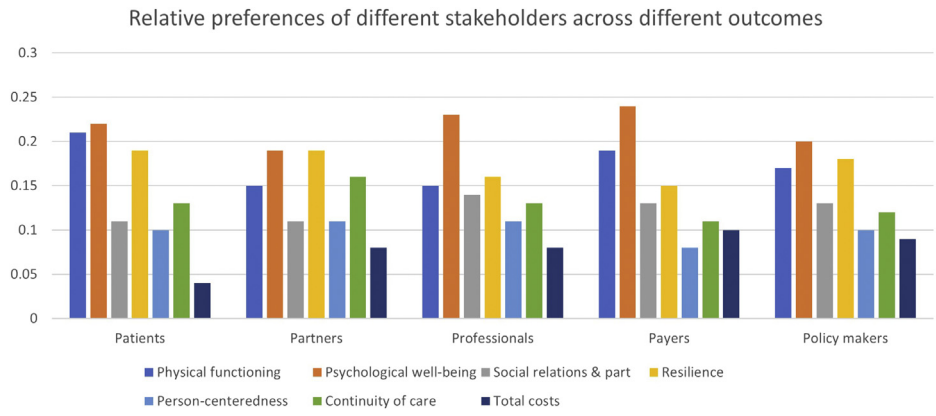

Table 4. Total and partial value scores at 6 months.

Variable	Predicted scores		Normalized performance score		Patients		Partners		Professionals		Payers		Policy makers	
					Weighted score		Weighted score		Weighted score		Weighted score		Weighted score	
	BSiN	UC	BSiN	UC	BSiN	UC	BSiN	UC	BSiN	UC	BSiN	UC	BSiN	UC
Health/ well-being														
Physical functioning	2.37	2.14	0.46	0.38	0.09	0.08	0.07	0.06	0.07	0.06	0.09	0.07	0.08	0.07
Psychological well-being	3.37	2.33	0.59	0.33	0.13	0.07	0.11	0.07	0.13	0.08	0.14	0.08	0.12	0.07
Social relationships and participation	2.21	1.84	0.60	0.42	0.07	0.05	0.07	0.05	0.08	0.06	0.08	0.06	0.08	0.05
Resilience	2.03	2.11	0.52	0.55	0.10	0.10	0.10	0.10	0.08	0.09	0.08	0.08	0.09	0.10
Experience of care														
Person centeredness	1.50	1.76	0.50	0.59	0.05	0.06	0.05	0.06	0.05	0.06	0.04	0.05	0.05	0.06
Continuity of care	0.98	0.97	0.98	0.97	0.13	0.13	0.16	0.16	0.13	0.13	0.10	0.10	0.12	0.12
Costs														
Total costs	€1796	€430	0.16	0.90	0.01	0.04	0.01	0.07	0.01	0.07	0.02	0.09	0.01	0.08
Overall value scores					0.57	0.53	0.58	0.57	0.57	0.55	0.55	0.53	0.55	0.55
mean (95% confidence interval)					(0.53–0.62)	(0.45–0.61)	(0.53–0.62)	(0.48–0.65)	(0.52–0.61)	(0.47–0.63)	(0.5–0.59)	(0.46–0.61)	(0.51–0.6)	(0.47–0.63)
% BSiN > UC					84		58		69		64		55	

BSiN indicates Better Together in Amsterdam North; UC, usual care.

Table 5. Total and partial value scores at 12 months.

Variable	Predicted scores		Normalized performance score		Patients		Partners		Professionals		Payers		Policy makers	
					Weighted score		Weighted score		Weighted score		Weighted score		Weighted score	
	IC	UC	IC	UC	IC	UC	IC	UC	IC	UC	IC	UC	IC	UC
Health/ well-being														
Physical functioning	2.50	2.29	0.50	0.43	0.10	0.09	0.08	0.07	0.08	0.06	0.09	0.08	0.09	0.08
Psychological well-being	3.02	2.53	0.50	0.38	0.11	0.08	0.10	0.07	0.11	0.09	0.12	0.09	0.10	0.08
Social relationships and participation	2.15	1.85	0.57	0.42	0.06	0.05	0.07	0.05	0.08	0.06	0.07	0.06	0.07	0.05
Resilience	2.05	2.42	0.52	0.71	0.10	0.14	0.10	0.13	0.09	0.12	0.08	0.11	0.10	0.13
Experience of care														
Person centeredness	1.33	1.28	0.44	0.43	0.04	0.04	0.05	0.05	0.05	0.05	0.04	0.04	0.05	0.04
Continuity of care	0.96	0.99	0.96	0.99	0.13	0.13	0.16	0.16	0.13	0.13	0.10	0.11	0.12	0.12
Costs														
Total costs	€426.80	€437.67	0.90	0.90	0.04	0.04	0.07	0.07	0.07	0.07	0.09	0.09	0.08	0.08
Overall value scores					0.59	0.57	0.61	0.60	0.60	0.58	0.60	0.57	0.60	0.58
Mean (95% confidence interval)					(0.54–0.63)	(0.5–0.63)	(0.57–0.66)	(0.53–0.67)	(0.56–0.65)	(0.51–0.64)	(0.55–0.64)	(0.5–0.63)	(0.56–0.65)	(0.52–0.65)
% IC > UC					70		65		77		82		71	

BSiN indicates Better Together in Amsterdam North; UC, usual care.

although some outcomes were improved, this improvement was not maintained when the 6-month trajectory of case management was stopped. This indicates that the intervention period might have been too short.

This study of BSiN also served as a case study to display the potential of MCDA. Although theoretical discussion about MCDA has been conducted,[10,11,16,31] the number of real-life applications remains limited.[11] The SELFIE MCDA framework[9] should be useful for healthcare decision makers for the evaluation of complex programs.[32] The availability of the SELFIE MCDA framework can provide a useful framework for designing the evaluation of these programs. Particularly, MCDA provides 3 key benefits. First, it provides a structured and explicit approach for evaluation of complex programs. For example, in this case study, it was apparent that on certain outcomes BSiN performed better than usual care and in some it did not. MCDA allows decision makers to use an explicit decision-making technique to aggregate the various outcomes and to determine an overall score while considering the preferences of stakeholders. If only the individual treatment effects were available, an overall decision could not have been made transparently. This MCDA evaluation process can be applied to different integrated care interventions and thus can ensure some consistency in valuation. It remains the case that although there are many innovative integrated care initiatives ongoing, program evaluation is often not included. Second, MCDA also provides insight into disaggregated effectiveness on a set of outcomes beneficial to a complex interventions. It can disentangle the overall improvement by outcome and allows the use of a broad range of outcomes, which is particularly important for complex interventions that target complex patients with problems in multiple life domains. Indeed, it is even more important in programs where one may expect improvements in 1 set of outcomes but deteriorations in others, as was the case for BSiN. Many existing evaluations tend to focus on health only and lack critical aspects important for complex interventions. This was apparent in the current evaluation of BSiN where one of the outcomes of the core set of outcomes in the SELFIE project, namely, enjoyment of life, was not measured as an outcome. Third, this MCDA framework allowed the evaluation of BSiN using appropriate and comprehensive outcomes on the Triple Aim. In a quality-adjusted life year-based approach, for example, only the direct health benefits and the costs would be included, and aspects such as social relations and participation where BSiN differed from usual care could not easily be included.

The literature has discussed the potential of MCDA but many methodological issues are still outstanding,[11,33] ranging from which outcomes to include to how to assess uncertainty. The SELFIE MCDA framework has systemically worked through several of these issues. First, we made sure that deterioration in one outcome could be compensated by improvement in other

outcomes, and lack of overlap between outcome measures was avoided to avoid double counting. Second, the PSs were determined using an evaluation with real patient data and not expert elicitation. Lack of data sources need not be a problem because existing questionnaires that have been validated for relevant populations can be used to measure the recommended outcomes. Third, a discrete choice experiment was used for weights because of the strong theoretical basis involving trade-offs between outcomes[34] and because it takes account of the entire range of performance[11] of integrated care programs. Other weight elicitation methods often do not take into account the full range of performance possible and thus are criticized for their lack of validity.[11] Fourth, uncertainty was formally incorporated using PSA Monte-Carlo simulations on both weights and scores. Finally, the importance of including all stakeholders has been noted,[31] which was addressed by having weights from 5 important stakeholder groups.

This study has some limitations. Owing to the nature of the study population, a randomized trial was not possible. We have attempted to reduce confounding and selection bias by introducing propensity score matching although even this is limited by the small control group. Within the sample, we saw quite some dropout, which was analyzed using intention to treat method, which assumes the observations are missing at random.[35] The nature of a study with a vulnerable complex population makes it unlikely that a study could be conducted with little dropout or a very comparable control group. The results should be interpreted considering these limitations, but it is difficult to speculate whether the results would be biased in a particular direction because of these limitations.

Future research should focus on developing stronger links between outcomes and weights in the MCDA. In this study, we have assumed that the performance scales and weight scales are linearly related. It could be the case that the relation is not linear. To investigate this may involve, for example, mapping between the discrete choice experiment levels and the performance scale. However, changing our assumption of linearity is unlikely to affect the current results because the PSs of the intervention and control group do not differ very much. Further work could explore the role of cost if the decision context changes. The current local context was to inform potentially wider implementation of the program in the region. However, for a more traditional health technology assessment approach to inform allocation of resources on national level, costs could be removed but a threshold for the maximum costs per unit gained in the overall value score would have to be elicited or estimated. Finally, further work could be obtained on the actual use of the MCDA tool by decision makers to elucidate how stakeholders want to engage with MCDA and precisely when and how it contributes the most.[33] Although the SELFIE MCDA framework had a clear decision context attached, it may be used in others, and potentially qualitative work with decision makers would enable the literature to understand the best application of MCDA in practice.

This study applied an MCDA to the evaluation of an innovative health and social care program aimed at achieving self-sufficiency with a group of people with multiple problems in the Netherlands. The BSiN program showed an overall increase in value, but this was mostly driven by a couple outcomes and the improvement was reduced during follow-up. This case study shows that the MCDA framework can be used to evaluate complex healthcare programs and the strength of using multiple outcomes, valid weights, and uncertainty analysis. Therefore, the SELFIE MCDA framework should be able to achieve greater transparency and better decisions in healthcare decision-making.

Supplemental Materials

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2021.02.007.

Article and Author Information

Accepted for Publication: February 12, 2021

Published Online: May 27, 2021

doi: https://doi.org/10.1016/j.jval.2021.02.007

Author Affiliations: Erasmus School of Health Policy & Management, Erasmus University, Rotterdam, the Netherlands (Karimi, Mölken); Netherlands Organization for Applied Scientific Research TNO, Unit Healthy Living, Leiden, The Netherlands (der Zwaan, van Genabeek); Department of Economics, University of Bergen, Bergen, Norway (Islam); NORCE Norwegian Research Centre, Bergen, Norway (Islam); Institute for Medical Technology Assessment (iMTA), Erasmus University Rotterdam, Rotterdam, The Netherlands (Mölken).

Correspondence: Milad Karimi, PhD, OPEN Health, Marten Meesweg 107, 3068 AV Rotterdam, The Netherlands. Email: miladkarimi@openhealthgroup.com

Author Contributions: *Concept and design*: Karimi, Islam, van Genabeek, Rutten-van Mölken
Acquisition of data: van der Zwaan, van Genabeek
Analysis and interpretation of data: Karimi, van der Zwaan, van Genabeek, Rutten-van Mölken
Drafting of the manuscript: Karimi, van der Zwaan, Islam, van Genabeek, Rutten-van Mölken
Critical revision of the paper for important intellectual content: Karimi, Islam, van Genabeek, Rutten-van Mölken
Statistical analysis: Karimi, van der Zwaan, Islam, Rutten-van Mölken
Provision of study materials or patients: van der Zwaan, van Genabeek
Obtaining funding: van Genabeek, Rutten-van Mölken
Administrative, technical, or logistic support: van der Zwaan
Supervision: Islam, van Genabeek, Rutten-van Mölken

Conflict of Interest Disclosures: The authors reported no conflict of interest.

Funding/Support: This work was supported by grants 634288 from the European Union's Horizon 2020 research and innovation program.

Role of the Funder/Sponsor: This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no 634288. The funder had no role in the design of the study; in the collection, analysis, and interpretation of data; and in writing the manuscript. The content of this publication reflects only the project group's views, and the European Commission is not liable for any use that may be made of the information contained herein.

REFERENCES

1. Nolte E, Knai C, Hofmarcher M, et al. Overcoming fragmentation in health care: chronic care in Austria, Germany and The Netherlands. *Health Econ Policy Law*. 2012;7(1):125–146.
2. Kuluski K, Ho JW, Hans PK, Nelson M. Community care for people with complex care needs: bridging the gap between health and social care. *Int J Integr Care*. 2017;17(4):2.
3. Bjorkquist C, Hansen GV. Coordination of services for dual diagnosis clients in the interface between specialist and community care. *J Multidiscip Healthc*. 2018;11:233–243.
4. Feinglass J, Norman G, Golden RL, N Muramatsu, Gelder M, Cornwell T. Integrating social services and home-based primary care for high-risk patients. *Popul Health Manag*. 2018;21(2):96–101.
5. Wagner EH, Austin BT, Von Korff M. Organizing care for patients with chronic illness. *Milbank Q*. 1996;74(4):511–544.
6. Crossman S, Ohde A. A case study exploring the effectiveness of an innovative "5Q Care Test" to determine whether patients with complex needs require health or social care. *Health Soc Care Community*. 2019;27(2):409–414.
7. Looman WM, Huijsman R, Fabbricotti IN. The (cost-)effectiveness of preventive, integrated care for community-dwelling frail older people: a systematic review. *Health Soc Care Community*. 2019;27(1):1–30.

8. Tsiachristas A, Stein KV, Evers S, Rutten-van Mölken M. Performing economic evaluation of integrated care: highway to hell or stairway to heaven? *Int J Integr Care*. 2016;16(4):3.
9. 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.
10. 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.
11. 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.
12. Fassaert T, Lauriks S, van de Weerd S, et al. Psychometric properties of the Dutch version of the self-sufficiency matrix (SSM-D). *Commun Ment Health J*. 2014;50(5):583–590.
13. GGD Amsterdam. Gezondheid in Beeld. Resultaten Amsterdamse Gezondheidsmonitor. Amsterdam: GGD Amsterdam; 2016. Available at: https://www.ggd.amsterdam.nl/beleid-onderzoek/gezondheidsmonitors/amsterdamse-0/. Accessed April 23, 2021.
14. Donker T, Comijs H, Cuijpers P, et al. The validity of the Dutch K10 and extended K10 screening scales for depressive and anxiety disorders. *Psychiatry Res*. 2010;176(1):45–50.
15. Fassaert T, De Wit MA, Tuinebreijer WC, et al. Psychometric properties of an interviewer-administered version of the Kessler psychological distress scale (K10) among Dutch, Moroccan and Turkish respondents. *Int J Methods Psychiatr Res*. 2009;18(3):159–168.
16. 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.
17. Leijten FRM, Hoedemakers M, Struckmann V, et al. Defining good health and care from the perspective of persons with multimorbidity: results from a qualitative study of focus groups in eight European countries. *BMJ Open*. 2018;8(8):e021072.
18. Leijten FRM, Struckmann V, van Ginneken E, et al. The SELFIE framework for integrated care for multi-morbidity: development and description. *Health Policy*. 2018;122(1):12–22.
19. Kempen GI, Miedema I, Ormel J, Molenaar W. The assessment of disability with the groningen activity restriction scale. Conceptual framework and psychometric properties. *Soc Sci Med*. 1996;43(11):1601–1610.
20. Strijk JE, van Dongen JM, van Steenbergen E, Wendel-Vos GCW, Hildebrandt V. Towards a positive view on health by gaining insights into the concept of vitality: associations with societal participation and costs among Dutch adults: Ingrid Hendriksen. *Eur J Public Health*. 2014;24(Suppl_2):cku163-108.
21. Rutten-van Mölken M, Karimi M, Leijten F, et al. 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.
22. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat Med*. 2015;34(28):3661–3679.
23. Baser O. Too much ado about propensity score models? Comparing methods of propensity score matching. *Value Health*. 2006;9(6):377–385.
24. 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.
25. McShane BB, Gal D, Gelman A, Robert C, Tackett JL. Abandon statistical significance. *Am Stat*. 2019;73(Sup 1):235–245.
26. Naversnik K, Rojnik K. Handling input correlations in pharmacoeconomic models. *Value Health*. 2012;15(3):540–549.
27. Wodchis WP, Dixon A, Anderson GM, Goodwin N. Integrating care for older people with complex needs: key insights and lessons from a seven-country cross-case analysis. *Int J Integr Care*. 2015;15:e021.
28. Lukersmith S, Millington M, Salvador-Carulla L. What is case management? A scoping and mapping review. *Int J Integr Care*. 2016;16(4):2.
29. Khanassov V, Pluye P, Descoteaux S, et al. Organizational interventions improving access to community-based primary health care for vulnerable populations: a scoping review. *Int J Equity Health*. 2016;15(1):168.
30. Hudon C, Chouinard MC, Lambert M, Dufour I, Krieg C. Effectiveness of case management interventions for frequent users of healthcare services: a scoping review. *BMJ Open*. 2016;6(9):e012353.
31. Hall W, Mitton C, Levy A. Realizing the "great potential" of MCDA in HTA. *Value Health*. 2018;21(12):1355–1356.
32. Marsh K, Thokala P, Youngkong S, Chalkidou K. Incorporating MCDA into HTA: challenges and potential solutions, with a focus on lower income settings. *Cost Eff Resour Alloc*. 2018;16(Suppl 1):43.
33. 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.
34. Bridges JF, Hauber AB, Marshall D, et al. Conjoint analysis applications in health–a checklist: a report of the ISPOR good research practices for conjoint analysis task force. *Value Health*. 2011;14(4):403–413.
35. Bhaskaran K, Smeeth L. What is the difference between missing completely at random and missing at random? *Int J Epidemiol*. 2014;43(4):1336–1339.

Appendix Preference weights per outcome per stakeholder

	Patients	**Partners**	**Professionals**	**Payers**	**Policy makers**
Physical functioning	0.21	0.15	0.15	0.19	0.17
Psychological well-being	0.22	0.19	0.23	0.24	0.20
Social relations & part	0.11	0.11	0.14	0.13	0.13
Resilience	0.19	0.19	0.16	0.15	0.18
Person-centeredness	0.10	0.11	0.11	0.08	0.10
Continuity of care	0.13	0.16	0.13	0.11	0.12
Total costs	0.04	0.08	0.08	0.10	0.09