

# Cost-Effectiveness of guideline-based stepped and collaborative care versus treatment as usual for patients with depression – a cluster-randomized trial

**Christian Brettschneider** (✉ [c.brettschneider@uke.de](mailto:c.brettschneider@uke.de))

University Medical Center Hamburg-Eppendorf <https://orcid.org/0000-0002-5280-1075>

**Daniela Heddaeus**

University Medical Center Hamburg-Eppendorf, Department of Medical Psychology

**Maya Steinmann**

University Medical Center Hamburg-Eppendorf, Department of Medical Psychology

**Martin Härter**

Universitätsklinikum Hamburg-Eppendorf Institut für Medizinische Soziologie

**Birgit Watzke**

University of Zurich, Institute of Psychology, Clinical Psychology and Psychotherapy Research

**Hans-Helmut König**

Universitätsklinikum Hamburg-Eppendorf Zentrum für Molekulare Neurobiologie Hamburg

---

## Research article

**Keywords:** Depressive Disorder, Costs and Cost Analysis, Quality-Adjusted Life Years, Delivery of Health Care, Integrated

**Posted Date:** March 10th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-16539/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

**Version of Record:** A version of this preprint was published on August 28th, 2020. See the published version at <https://doi.org/10.1186/s12888-020-02829-0>.

## Abstract

Objective Depression is associated with major patient burden. Its treatment requires complex and collaborative approaches. A stepped-care model based on the German National Clinical Practice Guideline “Unipolar Depression” has been shown to be effective. In this study we assess the cost-effectiveness of this guideline based stepped care model versus standard care in depression.

Methods This prospective cluster-randomized controlled trial included 737 depressive adult patients. Primary care practices were randomized to an intervention (IG) or a control group (CG). The intervention consisted of a four-level stepped care model. The CG received treatment as usual. A cost-utility analysis from the societal perspective with a time horizon of 12 months was performed. We used quality-adjusted life years (QALY) based on the EQ-5D as effect measure. Resource utilization was assessed by patient questionnaires. We calculated adjusted group differences in costs and effects, incremental cost-effectiveness ratios and cost-effectiveness acceptability curves. The complete sample and subgroups based on depression severity were considered.

Results In the IG, adjusted mean total costs (+5,016; SE: €2,691) and effects (+0.008 QALY; SE: 0.02) were higher than in the CG; yet, differences were not statistically significant. Significantly increased costs were found in the IG for outpatient physician services and psychiatrist services. Significantly increased total and indirect costs in the IG were found in the group with severe depression. Incremental cost-effectiveness ratios were unfavourable and the probability of cost-effectiveness was low, except for the group with moderate depression (70% for willingness-to-pay threshold of €50,000/QALY).

Conclusions We found no evidence for cost-effectiveness of the intervention. However, we identified indicators that the intervention works according to the aims of the National Treatment Guideline.

## 1. Background

The impact of depression on society is manifold. Research showed that depression occurs frequently (1), often gets diagnosed and treated with great delay (2–4), is associated with a substantial disease burden in terms of loss of quality of life (5), worsens the course and prognosis of somatic diseases (6–8) and causes a high economic burden (9). These challenges have been addressed by the development of systematic care approaches. In the German National Clinical Practice Guideline (NCPG) “Unipolar Depression” (10, 11) a stepped care approach based on collaborative principles is recommended (10). The aim of stepped care is the supply of treatment with the least necessary intensity while constantly monitoring the course of disease (12).

Objective of the presented study was the transfer of the recommendations of the NCPG into a program for clinical practice and, subsequently, the assessment of its effectiveness and cost-effectiveness. The results of the effectiveness assessment have been published recently (13). The guideline-based stepped care model (SCM) showed significantly higher odds of remission and response as well as significant reduction of depression severity in comparison to treatment as usual (TAU) (13). Hence, from a clinical point of view, the intervention can improve depression care. However, the effectiveness assessment did not include the economic consequences of the intervention. To assess these, we performed a cost-effectiveness analysis comparing SCM and TAU from the perspective of the German society.

## 2. Methods

### 2.1. Sample

The details of this study (ClinicalTrials.gov: NCT01731717) have been presented elsewhere (14). In summary, this analysis is based on a prospective cluster-randomized controlled trial. Patient recruitment and inclusion was performed between August 2012 and March 2014 in 49 (SCM: 36; TAU: 13) primary care practices in Hamburg, Germany (Follow-ups: between 2012 and April 2015). The randomization process was not blinded and took place on the practice level. Randomization was performed by a computer program (minimisation based on location and size of practices and the income level of the district the practice is located in). The randomization scheme between intervention group (IG) and control group (CG) was 3:1. Patients were included if they had a score  $\geq 5$  on the Patient Health Questionnaire (PHQ) 9 (indicating a mild depression at minimum), were 18 years or older and gave informed consent. Patients were excluded if they had insufficient German language skills or if a disease or disorder made it impossible to complete the questionnaire. Additionally, patients were excluded if their main treatment focus was on a comorbid mental disorder and not on depression.

### 2.2. Intervention

SCM: Patients in IG received services from a stratified stepped and collaborative care program, including GPs, psychiatrists, psychotherapists and psychiatric inpatient facilities. The SCM consisted of four steps. Step 1 incorporated active monitoring, Step 2 bibliotherapy, internet-based self-management and telephone-administered psychotherapy. Step 3 consisted of outpatient psychotherapy or antidepressant pharmacotherapy. In Step 4, a combination of psycho- and pharmacotherapy in an out- or inpatient setting was performed. The GP allocated the different interventions according to the guideline recommendations considering depression severity and patient preferences. Depression severity was monitored at regular intervals. Additionally, measures like an online platform displaying vacant treatment capacities in secondary care, a provider network, intensive training of GPs regarding guideline recommendations and quarterly quality circles were introduced.

TAU: Patients in the CG were able to receive every treatment officially approved to be delivered in the German health care system. This covers outpatient as well as inpatient psychotherapeutic or psychiatric services. GPs in the CG underwent the same intensive training regarding the guideline recommendations.

### 2.3. Data collection and measures

#### 2.3.1. Data collection

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

Data were collected at four time points by means of a self-reported questionnaire which were returned by mail: baseline (T0), after 3 months (T1), after 6 months (T2) and after 12 months (T3). Accordingly, the time horizon of the study was one year.

We assessed sociodemographic information, type of health insurance, employment status, social support (F-SOZU-14 (15)), the symptom severity of depression (PHQ-9 (16, 17)) and the physical and mental health status (Physical Component Score (PCS) and Mental Component Score (MCS) of the Short-Form-36 (SF-36) (18, 19)). Main outcomes of the cost-effectiveness analysis were quality-adjusted life years (QALY) in the 12-month period between T0 and T3 (EQ-5D-3L as measure of preference-based health-related quality of life (HRQL) (20)) and total 12-month costs calculated based on service utilization measured by a modified German version of the Client Sociodemographic and Service Receipt Inventory (CSSRI) (21).

### 2.3.2. Measurement of HRQL: EQ-5D-3L

The EQ-5D-3L consists of five items measuring current problems in the dimensions: mobility; self-care; usual activities; pain/discomfort; and anxiety/depression (20). There are three possible answers to each item: 1, no problems; 2, moderate problems; 3, extreme problems. Based on the patient's answers, it is possible to calculate a utility score (EQ-5D index score). These utility scores represent preference-based valuations of HRQL by the general population. In this study, we used EQ-5D index scores from the UK (22), ranging from -0.594 (worst health state) to 1 (best health state).

The EQ-5D has been validated in populations with depression (23, 24).

### 2.3.3. Questionnaire of service utilization

We adopted the societal perspective. In contrast to the assessment of the other instruments, we measured service utilization at T0, T2 and T3, not at T1. The questionnaires covered the periods of 6 months. We considered inpatient services, outpatient physician services (GP + 21 specialists), outpatient non-physician services (e.g. physiotherapy, occupational therapy, and exercise therapy), outpatient psychotherapist services, medication, ambulatory nursing care and informal care. Additionally, indirect costs due to sick leave and treatment appointments (absenteeism) were assessed. Resource utilization of services in Step 2 were extracted from the study documentation.

### 2.3.4. Unit costs

Costs were calculated in Euro at 2012 price level. As the time horizon of the study was one year, costs were not discounted.

Table 1  
Cost categories and sources of applied unit costs

Sector	Service / Goods	Units	Monetary values (unit costs)
Inpatient services	General hospitals, psychiatric hospitals and hospitals for rehabilitation	Days	Type specific mean rates (25)
Outpatient physician services	GP, specialists (e.g. cardiologist, internist, ophthalmologist)	Contacts	Type specific mean rates (25)
Outpatient non-physician services	e.g. physiotherapy, massage, lymph drainage, ergotherapy	Contacts	Reimbursement schedule (25)
Outpatient psychotherapist services	Psychotherapist	Contacts	Reimbursement schedule (25)
Medication	Product	Quantity	Official pharmaceutical index (Rote Liste) (26)
Nursing care	Ambulatory nursing care	Hours	Type specific wage (25)
	Informal care	Hours	Type specific wage (replacement cost approach) (25)
Indirect costs	Productivity losses	Hours	Gross income plus nonwage labor costs (27)

Detailed information regarding the unit costs is shown in Table 1. German standardised unit costs developed by Bock et al (25) were used for all categories, except for medication. The monetary valuation of medication was based on drug codes, dosage and duration and was valued based on the *RoteListe*, a German pharmaceutical database (26). Costs for inpatient services were calculated on a per day basis by hospital type. Outpatient phys-

cial care for older adults and disabled persons' (25). Indirect costs were valued based on the human capital approach by using mean gross income plus nonwage labour costs (27).

### 2.3.5. Intervention costs

Intervention costs were calculated for Step 2 services only. In steps 1, 3 and 4, outpatient physician or psychotherapeutic services, drug prescriptions and inpatient services were delivered. These costs were assessed and presented in the specific categories mentioned above.

As the intervention in step 2 consists of three services (bibliotherapy, internet-based self-management, telephone-administered psychotherapy), intervention costs represent the sum of costs of these three services. Bibliotherapy was valued by the price of the book (15 €). Internet-based self-management was priced by the license fee of the self-management program (250 €). Usually, the validity of the license is limited to 6 months. If a participant used the program

between baseline and T2 and between T2 and T3, we assumed that he or she required two licenses. Costs for telephone-administered psychotherapy were calculated by the product of the number of contacts and a price of 40 € per contact. This corresponds to the wage paid to the psychotherapist per session.

### 2.3.6. Calculation of QALY

QALY were calculated separately for each period between measurements. These values were summed up to gain 12-month QALY. The calculation was based on the assumption that the development of quality of life between two time points follows a linear trend. This means that the EQ-5D indices of two following time points, e.g. T2 and T3, were added and afterwards divided by 2 to gain the mean HRQL for this period. This mean HRQL value was multiplied with the observation time of the specific patient to calculate the QALY.

## 2.4. Statistical analysis

Analyses were performed based on the complete sample (base case analysis) as well as for subgroups of patients with different depression severity. Groups were defined by the baseline values of the PHQ-9. According to cut-off values extracted from the literature (28), a score of 5–9 constituted mild depression, a score of 10–14 moderate depression, a score of 15–19 moderately severe depression and a score of 20–27 severe depression.

All analyses were performed with STATA 15 (StataCorp, College Station, USA). Results were considered statistically significant at  $p \leq .05$ .

### 2.4.1. Imputation of missing values

Missing values were imputed on item level by 'multiple imputation using chained equations' (MICE) (29–32). 236 variables were included in the analysis. The proportion of missing values at baseline ranged from 0% (Age) and 27% (Number of hours absent from work due to physician appointments). 48% of the participants (IG: 48%; CG: 49%) had no missing values. Loss to follow-up was 39% (IG: 40%; CG: 36%).

Imputation was based on sociodemographic, clinical and economic data assessed at baseline as well as at T1 and T2 and was performed under fully conditional specification (32, 33). Summarizing, the following analyses are based on 30 datasets with  $N = 737$  participants per data set (IG: 569; CG: 168).

### 2.4.2. Comparison of baseline characteristics

We used linear and logistic mixed-effects regression models to identify baseline differences between IG and CG. The analyses were unadjusted considering only the treatment group as independent variable and the primary practice as random effect.

### 2.4.3. Comparison of total costs, cost categories and effects after 12 months

The analyses in the complete sample and the subgroups were adjusted for baseline variables with differences at a  $p$ -value of 0.1. This implies:

Complete sample: Age, employment status, PCS, MCS.

Mild depression: Age, social support.

Moderate depression: Age, employment status, PCS, MCS, social support, baseline HRQL.

Moderately severe depression: Type of health insurance, depression severity.

Severe depression: baseline HRQL.

Additionally, we considered the specific baseline costs in all analytical models.

We constructed linear mixed models with the aforementioned covariates as fixed effects and the primary care practice as random effect. To address the issue of the skewness of cost data, we calculated bootstrapped standard errors based on 1,000 replications.

After inspecting the results, we decided to perform an additional post-hoc analysis. In this analysis, we constructed mixed effects logistic models to assess differences in the probability of resource utilization in mental health care services (inpatient psychiatric, outpatient psychiatric and outpatient psychotherapeutic services). The models were adjusted by the aforementioned variables.

### 2.4.4. Calculation of the ICER as point estimate of cost-effectiveness

We calculated the incremental cost effectiveness ratio (ICER) as a point estimate of cost-effectiveness. The ICER is a ratio and consists of the difference between IG and CG in mean total costs (C) in the numerator and mean effects (E) in the denominator:

$$ICER = \frac{\bar{C}_{IG} - \bar{C}_{CG}}{\bar{E}_{IG} - \bar{E}_{CG}} = \frac{\Delta \bar{C}}{\Delta \bar{E}}$$

As there is no official German threshold to consider an ICER cost-effective, we applied the widely used threshold of 50,000 €/QALY (34).

### 2.4.5. Calculation of the CEAC as assessment of uncertainty

As the ICER is a point estimate considering only mean values of costs and effects, it provides no information on the uncertainty in the analysis. For this reason, we constructed cost-effectiveness acceptability curves (CEAC) based on a series of net-benefit regressions (35, 36).

In a first step, the patient-specific net benefit (NB)  $NB_i = E_i \times \lambda - C_i$  was calculated. The NB consists of the individual 12-month costs in € ( $C_i$ ), the individual 12-month effect in QALY ( $E_i$ ) and a willingness-to-pay (WTP) margin in €/QALY ( $\lambda$ ). To construct a CEAC, the individual NB is used as dependent variable in a regression model, while group is used as independent variable. This procedure is repeated for different WTP margins. In case of our study, we used WTP margins ranging from 0 €/QALY to 130.000 €/QALY and proceeded in 10.000/QALY steps. To present the CEAC graphically, the different WTP margins are plotted on the x-axis and the probabilities of cost-effectiveness are plotted on the y-axis. The probability of cost-effectiveness at a WTP margin corresponds to the 0.5 x the p-value of the coefficient of the group difference in the net-benefit regressions, in case the coefficient is negative and 1-0.5 x the p-value if the coefficient is positive. For a rationale of this approach, please see Hoch et al (35).

We used the same regression approach and adjusted for the same covariates as for the comparison of costs and effects (step 3).

### 3. Results

#### 3.1. Characteristics of the study population at baseline

The mean age of the population was 42.9 years (SD: 14.0; Range: 18–88), the majority was female (73%). 59% of the participants were living with a partner. The mean symptom severity of depression was moderately severe (mean PHQ-9: 15.0; SD: 4.8). 93 patients were mildly, 232 moderately depressed. 271 patients suffered from a moderately severe depression and 141 from a severe depression. Mean HRQL (EQ-5D index) was 0.57 (SD: 0.27). Patients in the IG were more frequently employed than patients in the CG (IG: 78%; CG: 69%;  $p < .05$ ). No other differences reached statistical significance at a level of  $p \leq .05$  (Table 2).

Table 2  
Sociodemographic characteristics of the complete sample at baseline

Characteristic	Intervention group (n = 569)	Control group (n = 168)	p-value
Age (years)			
mean (SD)	42.09 (13.45)	45.60 (15.45)	0.07
Female: %	72.41	76.19	0.38
Single: %	59.03	56.83	0.64
Private HI: %	5.55	8.73	0.25
Employed: %	77.56	68.53	0.04
Severity of depression (PHQ-9)			
mean (SD)	15.29 (4.68)	14.09 (4.91)	0.17
Total costs (€)			
mean (SD)	5,636 (8,297)	7,688 (10,764)	0.10
EQ-5D Index			
mean (SD)	0.58 (0.26)	0.53 (0.28)	0.22
Physical Health Status (SF-12)			
mean (SD)	44.48 (10.52)	42.15 (10.31)	0.06
Mental Health Status (SF-12)			
mean (SD)	28.56 (8.41)	30.56	0.07

#### 3.2. Differences in Costs and Effects

The results of the group comparison in terms of costs and effects are displayed in Table 3.

Table 3  
Differences in costs and effects in the complete sample and per subgroup (€ 2012)

Cost category	Total sample (n = 737)		Subgroup: Mild depression (n = 93) (PHQ-9: 5–9)		Subgroup: Moderate depression (n = 232) (PHQ-9: 10–14)		Subgroup: Moderately severe depression (n = 271) (PHQ-9: 15–19)		Subgroup: Severe depression (n = 141) (PHQ-9: 20–27)	
	Mean (€)	SE (€) <sup>#</sup>	Mean (€)	SE (€) <sup>#</sup>	Mean (€)	SE (€) <sup>#</sup>	Mean (€)	SE (€) <sup>#</sup>	Mean (€)	SE (€) <sup>#</sup>
Total costs	5,016	2,691	9,731	6,504	-628	3,476	6,047	4,659	14,579*	6,017
Direct costs	2,719	1,790	6,435	4,391	-233	2,399	3,676	2,580	4,306	3,950
Inpatient services	1,769	1,589	3,468	3,601	-453	2,263	2,487	2,155	2,911	3,629
Psychiatric hospital	-202	417	1,101	1,403	-660	854	196	638	-1,695	1,875
Outpatient physician services	467*	174	723*	300	347	239	832*	252	266	322
Psychiatrist	95*	30	84	76	79	62	77	55	123	81
Psychotherapist	239	141	250	203	244	230	242	220	75	30
Outpatient non-physician services	37	49	46	121	-57	81	50	75	51	117
Drugs	-43	179	256	419	49	180	-199	355	-67	155
Ambulatory care	-8	20	32	45	6	66	-24	32	124	206
Informal care	82	332	543	710	-162	496	-72	520	434	1,029
Intervention	218*	25	285*	64	195*	32	204*	36	203*	44
Indirect costs	2,238	1,483	2,754	3,159	-279	2,053	1,855	3,172	10,646*	3,582
QALY (EQ-5D: UK-Tarif)	0.008	0.020	-0.010	0.057	0.028	0.037	0.013	0.036	0.007	0.053

\*p < .05; <sup>#</sup> bootstrapped standard error (1,000 replications)

In the complete sample, total costs as well as direct and indirect costs increased in the IG. However, these differences did not achieve statistical significance. Significantly increased cost in the IG were found for outpatient physician services and intervention services in step 2. In reference to QALY, there was a non-significant gain in the IG.

Cost differences in the groups with mild and moderately severe depression were comparable to the results in the complete sample. Total, direct and indirect costs were non-significantly increased in the IG while the increases in outpatient physician services and intervention services were statistically significant. On the effect side, the IG suffered a loss of QALY in the group with mild depression and reached a gain in the group with moderately severe depression. Both difference did not reach statistical significance.

The group with moderate depression showed a different pattern. Total, direct and indirect costs were lower in the IG. Additionally there was a QALY gain in the IG. Both differences did not reach statistical significance.

The group with severe depression was the only group showing a significant difference in total costs. Patients in the IG caused higher total costs over the course of 12 months (mean: +14,579 €; SE: 6,017; p < .01). This difference was mainly driven by a significant increase of indirect costs (mean: +10,646 €; SE: 3,582; p < .01). This increase was mainly caused by significantly higher costs in the IG between baseline and T2, i.e. in the first six months (mean: +7,593 €; SE: 2,781; p < .01). The difference in the time period between T2 and T3 (six months) was not significant (mean: +3,068 €; SE: €1,733; ns). On the effect side, there was a small non-significant gain in QALY in the IG.

### 3.3. Differences in costs and utilization of mental health services

For outpatient psychiatrist services we found significantly increased costs (mean: +95€; SE: 30 €; p < .01) in the IG in the complete sample. Significant differences in outpatient psychotherapist services were not identified. However, there were trends for higher costs for outpatient psychiatrist and psychotherapeutic services in all subgroups of the IG.

For psychiatric hospital services, we found a trend for lower cost in the complete sample and the groups with moderate and severe depression, but a trend for higher costs in the groups with mild and moderately severe depression.

Table 4 presents the utilization rates for the mental health services. There were no significant differences between the groups. However, we identified different trends. In the total sample, outpatient services were more frequently utilized in the IG, while inpatient services were more frequently used in the CG. This was also observed for all subgroups, except for inpatient services in the group with mild depression.

Table 4  
Utilization of mental health care services in the complete sample and per subgroup

Mental health services	Total sample (n = 737)		Subgroup: Mild depression (n = 93) (PHQ-9: 5–9)		Subgroup: Moderate depression (n = 232) (PHQ-9: 10–14)		Subgroup: Moderately severe depression (n = 271) (PHQ-9: 15–19)		Subgroup: Severe depression (n = 141) (PHQ-9: 20–27)	
	IG	CG	IG	CG	IG	CG	IG	CG	IG	CG
Psychiatric hospital	6%	10%	8%	3%	3%	7%	6%	13%	10%	18%
Psychiatrist	42%	35%	42%	25%	38%	31%	40%	42%	51%	40%
Psychotherapist	60%	52%	47%	34%	56%	48%	61%	58%	69%	67%

### 3.4. Point estimates of cost-effectiveness

In the complete sample, the ICER was unfavourable (627,000 €/QALY). In the group with mild depression, the intervention was dominated which means that the intervention caused higher costs but led to fewer QALY. In the group with moderate depression, the intervention was dominant as it reduced costs and increased effects. In the remaining groups there were unfavourable ICER of 465,154 €/QALY (moderately severe depression) and 2,082,714 €/QALY (severe depression).

### 3.5. Uncertainty analyses of cost-effectiveness

Pattern one (complete sample and severe depression) shows a rather flat slope on a very low level of probability for cost-effectiveness. This indicates that it is - independent from the WTP- highly unlikely that the intervention is cost-effective. By using the commonly employed WTP margin of 50,000 €/QALY, we can even derive that in the group with severe depression the intervention was not cost-effective as the probability of cost effectiveness was 2.5%, which means that the probability of the control being cost-effective was 97.5%. The second pattern (mild and moderately severe depression) shows also a rather low probability of cost-effectiveness of the intervention (between 10% and 30%). However, the slope of these curves is steeper, which indicates a higher degree of uncertainty in the data of these groups. The third pattern, found in the group with moderate depression, shows an already elevated probability of 57% at the minimum WTP, which increases to 78%. Using the WTP margin of 50,000 €/QALY, the CEAC indicates a 70% probability of cost-effectiveness of the intervention.

## 4. Discussion

Our analysis failed to provide sufficient evidence to decision-makers that the intervention is cost-effective. In case of severe depression, the evidence even indicates that routine care is economically advantageous. These results are not in line with the findings by Härter et al, who observed for patients in the SCM a pronounced improvement of symptom burden as well as increased odds of response and remission (13). Nevertheless, some indicators for an impact of the intervention on health care delivery warrant further research on the intervention.

The development of the NCPG and hence of this intervention was triggered by the need to improve identification of depression cases and to bring patients into adequate treatment (10, 11). Several of our results can be interpreted as in accord with this aim. The NCPG recommends low intensity treatments for patients with mild depression (10). In our study, these interventional measures (bibliotherapy, web-based self-management, telephone psychotherapy) showed the highest incremental costs in this group of patients in comparison to other degrees of severity. Furthermore, the NCPG lays a strong emphasis on treatment in the outpatient sector by mental health professionals (10). In the complete sample, we found that the costs for psychiatric outpatient services were significantly increased in the IG. The same trend was found for all subgroups and the psychotherapeutic services. Considering the share of patients utilizing these services, we found that these services were more often utilized in the IG, especially in the mildly and moderately affected subgroups.

A different pattern appeared with respect to psychiatric inpatient services. For these services, we found a trend for cost reductions (moderate and severe depression) and negligibly higher costs (moderately severe depression) in the IG. As the NCPG fosters outpatient treatment, this can be interpreted as in line with the NCPG. However, there are two intuitively contradicting observations. IG patients in the group with mild depression showed the largest increase in costs for psychiatric hospital services and those in the group with severe depression the largest decrease of these costs. It is not to expect that mildly affected patients are in need of a care-intensive inpatient treatment, while severely affected patients appear to be the main target population for these services. The observation in the severely affected group could be interpreted as an intervention effect. It is possible that the SCM provided the coordinating physician with a variety of different treatment options making inpatient treatment a last line treatment. The observation in the group of mildly affected patients is part of a larger utilization pattern. This group exhibits not only the highest cost difference between IG and CG in psychiatric hospital services but also in drugs and hospital services in general. On the one hand, it is possible that this result was affected by the small sample size in this group (IG: 63; CG 30) and is hence not generalizable for the group of all patients with mild depression. On the other hand, there could be a systematic effect. For example, the SCM could increase the availability of mental health care services. Furthermore, by including the patient in the process of decision-making the demand for health care services could grow. This could be an explanation for the higher costs for drugs and inpatient psychiatric services. The trend towards higher costs for hospital services in

general is not only a phenomenon observable in the group with mild depression but also in other subgroups. These findings and the corresponding observations for outpatient services (a trend for an increase in mental health unspecific services) can be interpreted as indicators of an activating effect of the SCM. Depressive disorders are associated with a lack of motivation and energy and strong feelings of hope- and helplessness (37). Furthermore, they have a relationship of mutual influence with somatic disorders (3, 38, 39). The analysis by Härter et al showed that the symptom status of patients in the SCM group improved significantly (13). Additionally, this program is based on regular monitoring and frequent contacts to health care professionals. This results in a higher probability of the treatment of previously untreated health issues and is an explanation for the higher utilization of mental health unspecific services.

The last aspect of the results that need to be discussed are the significantly increased costs for absenteeism from work caused by participants with severe depression in the IG. As only the cost difference between T0 and T2 was significant, we can conclude that physicians tend to attest long-term sick leave right at the beginning of the intervention to facilitate the systematic treatment of the severely affected patients. After six months and successfully initiating treatment, physicians might start to reintegrate patients into their daily routines again.

In view of the results and the discussion it becomes obvious that follow-up time is the most limiting factor in our study. The time horizon of our study was one year. This implies that it is possible that not all effects and health care consequences of the intervention were observed. Due to the natural course of depression, the duration and number of episodes, the duration of remission and the risk of relapses, one year might be too short to observe all differences between the interventions (40, 41). According to our results, the costs go up in the first year and there is no notable effect on HRQL. However, we cannot rule out that the increase in costs is an investment in the future and will result in a decrease in later years. On the effects side, we know that the intervention reduces symptom severity and leads to more remissions, but not to better HRQL. The conceptual literature on patient reported outcomes shows that symptom status and HRQL are related and part of the same outcome continuum (40, 41). Symptom status can affect HRQL directly, if it is especially burdensome, or via its influence in functioning and general health perception. Supposing that the change in symptom status was not pronounced enough to affect HRQL directly, it has to influence functioning and general health perception. Considering this, we can assume that the time horizon of one year could not have been long enough to lead to changes in functioning and perception and hence to a substantial change in HRQL.

Furthermore, we have to consider that the use of patient questionnaires is associated with a risk of missing values and recall bias. The degree of missing values was manageable and was handled by an elaborated approach. The presence of a recall bias cannot be ruled out or controlled. Additionally, in the interpretation of the results, we have to keep in mind that we identified mostly trends and that the randomization was not stratified for the subgroups. This means that the composition of the subgroups was not necessarily randomized. For this reason, we adjusted the analyses in the subgroups for the group specific significant baseline differences.

## 5. Conclusion

We found no evidence that the SCM is cost-effective over a one-year period. However, we identified indicators that the intervention works according to the aims of the NCPG. For this reason, it is theoretically possible that the SCM proves its societal value over the long term. The results show the need for the further evaluation of the SCM. Specifically, relevant research fields are the assessment of the long-term cost-effectiveness; the in-depth analysis of the processes in the group with mild depression; and the adaption of the SCM to the specific needs of patients with moderate depression, as this group showed the most promising results.

## Abbreviations

C: mean total Costs

CEAC: Cost-Effectiveness Acceptability Curve

CG: Control Group

CSSRI: Client Sociodemographic and Service Receipt Inventory

E: mean Effects

GP: General Practitioner

HRQL: Health-related Quality of Life

ICER: Incremental Cost Effectiveness Ratio

IG: Intervention Group

MCS: Mental Component Score

MICE: Multiple Imputation using Chained Equations

NB: Net Benefit

NCPG: National Clinical Practice Guideline

PCS: Physical Component Score  
Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

PHQ: Patient Health Questionnaire

QALY: Quality Adjusted Life Year

SCM: Stepped Care Model

SF-36: Short Form 36

TAU: Treatment As Usual

## Declarations

Ethics approval and consent to participate

The study was approved by the Ethics Committee of the Hamburg Chamber of Psychotherapists. The study was conducted according to the principles of the Declaration of Helsinki (2013 version). Written informed consent was obtained from all participants before inclusion into the study.

Consent for publication

Not applicable

Availability of data and materials

Data are available from the corresponding author on reasonable request.

Competing interests

Christian Brettschneider is an Associate Editor of BMC Psychiatry. All other authors declare that there are no competing interests.

Funding

The study was funded by the German Federal Ministry of Education and Research (Grant number: 01KQ1002B). The funding source of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Authors' contributions

MH and BW conceptualized and designed the clinical parts of the study. CB and HHK conceptualized and designed the economic parts. MH, BW, MS and DH contributed to the acquisition of the data. CB and HHK analysed the data. All authors contributed to the interpretation of the results. CB drafted the manuscript. All authors revised the manuscript critically and approved the final version of the manuscript. All authors agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Acknowledgements

Not applicable

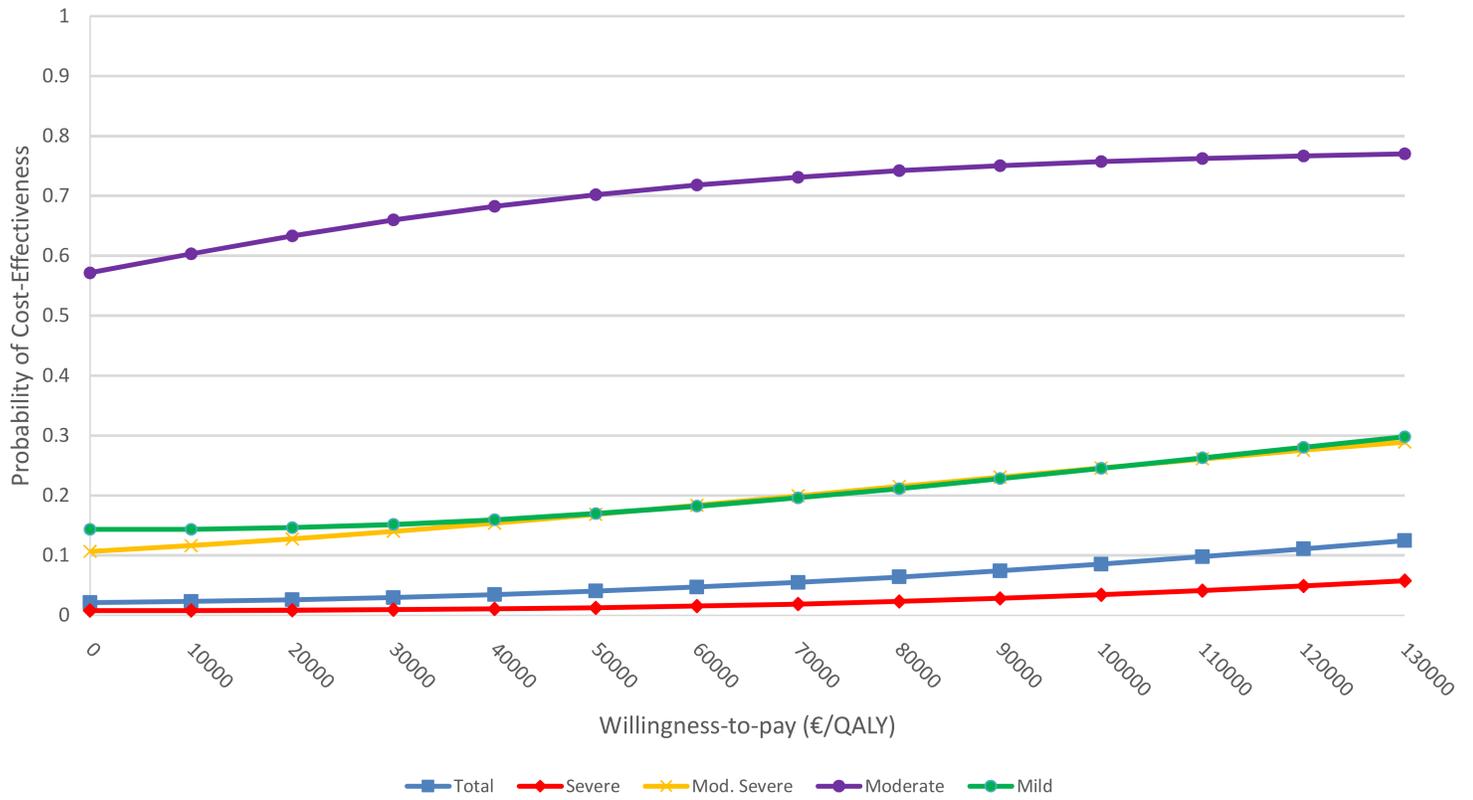
## References

1. Wittchen HU, Jacobi F, Rehm J, Gustavsson A, Svensson M, Jönsson B, et al. The size and burden of mental disorders and other disorders of the brain in Europe 2010. *Eur Neuropsychopharmacol.* 2011;21(9):655-79.
2. Duhoux A, Fournier L, Gauvin L, Roberge P. Quality of care for major depression and its determinants: a multilevel analysis. *BMC Psychiatry.* 2012;12:142.
3. Katon WJ. Epidemiology and treatment of depression in patients with chronic medical illness. *Dialogues Clin Neurosci.* 2011;13(1):7-23.
4. Jacobi F, Höfler M, Meister W, Wittchen HU. [Prevalence, detection and prescribing behavior in depressive syndromes. A German federal family physician study]. *Nervenarzt.* 2002;73(7):651-8.
5. Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet.* 2012;380(9859):2163-96.
6. Stafford L, Berk M, Reddy P, Jackson HJ. Comorbid depression and health-related quality of life in patients with coronary artery disease. *J Psychosom Res.* 2007;62(4):401-10.
7. Leung YW, Flora DB, Gravely S, Irvine J, Carney RM, Grace SL. The impact of premorbid and postmorbid depression onset on mortality and cardiac morbidity among patients with coronary heart disease: meta-analysis. *Psychosom Med.* 2012;74(8):786-801.
8. Cuijpers P, Vogelzangs N, Twisk J, Kleiboer A, Li J, Penninx BW. Differential mortality rates in major and subthreshold depression: meta-analysis of studies that measured both. *Br J Psychiatry.* 2013;202(1):22-7.
9. Luppá M, Heinrich S, Angermeyer MC, König HH, Riedel-Heller SG. Cost-of-illness studies of depression: a systematic review. *J Affect Disord.* 2007;98(1-2):29-43.
10. DGPPN B, KBV, AWMF (Hrsg.) für die Leitliniengruppe Unipolare Depression., S3-Leitlinie/Nationale VersorgungsLeitlinie Unipolare Depression –

Loading  $\text{[MathJax]}/\text{jax/output/CommonHTML/fonts/TeX/fontdata.js}$  [2: [Available from: [www.depression.versorgungsleitlinien.de](http://www.depression.versorgungsleitlinien.de).

11. Härter M, Bermejo I, Ollenschläger G, Schneider F, Gaebel W, Hegerl U, et al. Improving quality of care for depression: the German Action Programme for the implementation of evidence-based guidelines. *Int J Qual Health Care*. 2006;18(2):113-9.
12. Bower P, Gilbody S. Stepped care in psychological therapies: access, effectiveness and efficiency. Narrative literature review. *Br J Psychiatry*. 2005;186:11-7.
13. Härter M, Watzke B, Daubmann A, Wegscheider K, König HH, Brettschneider C, et al. Guideline-based stepped and collaborative care for patients with depression in a cluster-randomised trial. *Sci Rep*. 2018;8(1):9389.
14. Watzke B, Heddaeus D, Steinmann M, König HH, Wegscheider K, Schulz H, et al. Effectiveness and cost-effectiveness of a guideline-based stepped care model for patients with depression: study protocol of a cluster-randomized controlled trial in routine care. *BMC Psychiatry*. 2014;14:230.
15. Fydrich T, Sommer G, Tydecks S, Brähler E. Fragebogen zur sozialen Unterstützung (F-SozU): Normierung der Kurzform (K-14). *Z Med Psychol*. 2009;18(1):43-8.
16. Löwe B, Spitzer RL, Zipfel J, Herzog W. PHQ-D Gesundheitsfragebogen für Patienten; Manual und Kurzform: Pfizer; 2002.
17. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: Validity of a Brief Depression Severity Measure. *J Gen Intern Med*. 2001;16(9):606-13.
18. Ware JE, Sherbourne CD. The MOS 36-Item Short-Form Health Survey (SF-36): I. Conceptual Framework and Item Selection *Med Care*. 1992;30(6):473-83.
19. Bullinger M. German translation and psychometric testing of the SF-36 Health Survey: Preliminary results from the IQOLA Project. *Soc Sci Med*. 1995;41(10):1359-66.
20. EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy*. 1990;16(3):199-208.
21. Roick C, Kilian R, Matschinger H, Bernert S, Mory C, Angermeyer MC. [German adaptation of the client sociodemographic and service receipt inventory - an instrument for the cost of mental health care]. *Psychiatr Prax*. 2001;Suppl 2:S84-S90.
22. Dolan P. Modeling valuations for EuroQol health states. *Med Care*. 1997;35(11):1095-108.
23. Sapin C, Fantino B, Nowicki ML, Kind P. Usefulness of EQ-5D in assessing health status in primary care patients with major depressive disorder. *Health Qual Life Outcomes*. 2004;2:20.
24. König HH, Bernert S, Angermeyer MC. [Measuring preferences for depressive health states]. *Psychiatr Prax*. 2005;32(3):122-31.
25. Bock JO, Brettschneider C, Seidl H, Bowles D, Holle R, Greiner W, et al. [Calculation of standardised unit costs from a societal perspective for health economic evaluation]. *Gesundheitswesen*. 2015;77(1):53-61.
26. Rote Liste Service GmbH. Rote Liste 2012. Frankfurt/Main: Rote Liste Service GmbH; 2012.
27. German Statistical Office. Earnings and labour costs Wiesbaden: German Statistical Office; 2012 [Available from: <https://www.destatis.de/EN/FactsFigures/NationalEconomyEnvironment/EarningsLabourCosts/EarningsLabourCosts.html>].
28. Kroenke K, Spitzer RL. The PHQ-9: A New Depression Diagnostic and Severity Measure. *Psychiatr Ann*. 2002;32:509-21.
29. 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;20:40-9.
30. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med*. 2011;30:37-399.
31. Rubin DB. Multiple imputation for nonresponse in surveys: John Wiley & Sons; 2004.
32. van Buuren S, Brand JPL, Groothuis-Oudshoorn CGM, Rubin DB. Fully conditional specification in multivariate imputation. *J Stat Comput Simul*. 2006;76(12):1049-64.
33. van Buuren S. Flexible Imputation of Missing Data Boca Raton: CRC Press; 2012.
34. Grosse SD. Assessing cost-effectiveness in healthcare: history of the \$50,000 per QALY threshold. *Expert Rev Pharmacoecon Outcomes Res*. 2008;8:165-78.
35. Hoch JS, Briggs AH, Willan AR. Something old, something new, something borrowed, something blue: a framework for the marriage of health econometrics and cost-effectiveness analysis. *Health Econ*. 2002;11(5):415-30.
36. Briggs AH, O'Brien BJ, Blackhouse G. Thinking outside the box: recent advances in the analysis and presentation of uncertainty in cost-effectiveness studies. *Annu Rev Public Health*. 2002;23:377-401.
37. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5). Arlington: American Psychiatric Association; 2013.
38. Rutledge T, Reis VA, Linke SE, Greenberg BH, Mills PJ. Depression in heart failure a meta-analytic review of prevalence, intervention effects, and associations with clinical outcomes. *J Am Coll Cardiol*. 2006;48(8):1527-37.
39. Thombs BD, Bass EB, Ford DE, Stewart KJ, Tsilidis KK, Patel U, et al. Prevalence of depression in survivors of acute myocardial infarction. *J Gen Intern Med*. 2006;21(1):30-8.
40. Eaton WW, Shao H, Nestadt G, Lee HB, Bienvenu OJ, Zandi P. Population-based study of first onset and chronicity in major depressive disorder. *Arch Gen Psychiatry*. 2008;65:513-20.
41. Solomon DA, Keller MB, Leon AC, Mueller TI, Lavori PW, Shea MT, et al. Multiple recurrences of major depressive disorder. *Am J Psychiatry*. 2000;157:229-33.

## Figures



**Figure 1**  
 Cost-Effectiveness Acceptability Curves for the complete sample and the subgroups

### Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [CHEERSStatement.pdf](#)