Going beyond cost-effectiveness: analyzing routine mental healthcare data and stakeholders' perspectives to improve depression care

Kan, Kaying

DOI:
10.33612/diss.183452704

CHAPTER 2

The clinical effectiveness of an algorithm-guided treatment program for depression in specialized mental healthcare: A comparison with efficacy trials

Published as:

Kan, K., Feenstra, T. L., de Vries, S. O., Visser, E., Schoevers, R. A., & Jörg, F. (2020). The clinical effectiveness of an algorithm-guided treatment program for depression in specialized mental healthcare: A comparison with efficacy trials. *Journal of Affective Disorders*, 275, 216-223.

ABSTRACT

Background: Doubts exist on whether effects found in randomized controlled trials (RCTs) are directly generalizable to daily clinical practice. This study aimed (a) to investigate the effectiveness of treatment options within an algorithm-guided treatment (AGT) program for depression and compare their effectiveness with outcomes of efficacy trials and (b) to assess the relation between treatment continuity and outcomes.

Methods: This naturalistic study linked treatment data from January 2012 to November 2014 from a Dutch mental healthcare provider, to routine outcome monitoring (ROM) data (N=351). Effectiveness of the treatment options (pharmacotherapy, psychotherapy and their combination) was compared to the efficacy reported in the meta-analyses. We included treatment continuity as binary variable "early terminators versus completers of the recommended number of treatment sessions".

Results: Remission rates for psychotherapy (38% [95% CI: 32-45]), pharmacotherapy (31% [95% CI: 22-42]) and combination therapy (46% [95% CI: 19-75]) were respectively lower, comparable, and comparable to those reported in the meta-analyses. Similarly, response rates were respectively lower (24% [95% CI: 19-30]), lower (21% [95% CI: 13-31]), and comparable (46% [95% CI: 19-75]) to meta-analyses results. A similar share of early terminators and completers achieved remission and response.

Limitations: A substantial proportion of patients had incomplete ROM data after data linkage. Limited set of patient characteristics to check for selection bias.

Conclusions: Despite the more heterogeneous patient population in clinical practice, the effectiveness of an AGT program, emphasizing strict guideline adherence, approached that found in RCTs. A fixed number of treatment sessions may not suit all individual patients.

INTRODUCTION

To date, a large number of randomized controlled trials (RCTs) has investigated the efficacy of various interventions for depression. Meta-analyses have shown that different types of psychotherapy, pharmacotherapy and their combination are effective in the treatment of depression in adults[1-3]. Compared to efficacy trials, effectiveness in clinical practice settings indicate lower[4,5] to similar results[6,7].

Various reasons explain the differences in outcomes between RCTs and clinical practice. First, only a minority of depressed patients in the "real-world" qualify for participation in efficacy trials[8,9]. Stringent inclusion and exclusion criteria result in relatively homogeneous patient groups. Second, RCTs apply strict protocols to ensure guideline adherence. In clinical practice, clinicians might deviate from practice guidelines, resulting in substantial treatment variance. Clinicians' adherence to treatment guidelines and algorithm-guided treatment (AGT) of depression have been associated with improved treatment outcomes and better quality of care among patients with depression[10-15]. Third, in clinical practice the number of treatment sessions completed might vary, unlike in RCTs. Previous studies have shown inconclusive results on the association between the number of sessions and treatment outcomes[16-22]. Finally, efficacy trials place special emphasis on treatment adherence among patients, resulting in more favorable health outcomes[23-25].

Several medication treatment algorithms studies, like the German Algorithm Project (GAP)[13,26,27] and Texas Medication Algorithm Project (TMAP)[28], investigated the effectiveness of algorithm-guided treatment decisions. In these studies, a more favorable outcome, in terms of symptom reduction, treatment response or remission, was found compared to treatment as usual.

In 2010, GGZ Friesland, a large specialized mental healthcare provider in the north of the Netherlands, introduced an AGT program for depression. One of its main objectives was to optimize guideline adherence by clinicians in clinical practice. The present study aimed to investigate the clinical effectiveness of the AGT program, consisting of both psychotherapeutic and pharmacological interventions, in a naturalistic setting. In contrast to RCTs, this study applied no inclusion or exclusion selection criteria, other than a primary depression diagnosis and availability of both pre-treatment and post-treatment measurement scores. The effectiveness of the different treatments within the AGT program was compared to the efficacy reported in RCTs. Additionally, the present study aimed to assess the relation between treatment continuity and subsequent remission and response rates. We hypothesized that the AGT program leads to comparable effectiveness compared to the efficacy in RCTs due to improved clinician's adherence to the algorithm. We expected that patients that completed their treatment had better treatment outcomes than patients that terminated treatment early.

METHODS

Study Setting

Data for this naturalistic study were provided by GGZ Friesland, a specialized mental healthcare provider with twelve locations in the Netherlands. Data covered the period from January 2012 to November 2014. During this period, GGZ Friesland implemented various specialized mental healthcare programs, including an AGT program for the treatment of depressive disorders. The main objectives of the AGT program were (a) to improve the quality of care and (b) to optimize treatment effectiveness, by deployment of specialist psychiatric personnel and enhancement of guideline adherence by clinicians in daily clinical practice.

The AGT program for depression consisted of a combination of stepped care and matched care, based on the Dutch multidisciplinary guideline for depression treatment[29], matching international guidelines on depression. To support decision-making in daily practice, electronic patient files contained a built-in decision tree reflecting the treatment algorithm, leading to a recommended treatment pathway (for more details on the AGT program and the treatment pathways, see the supplementary material). Treatment pathways were defined by (a) the type of treatment, (b) the frequency of treatment and (c) the maximum number of treatment sessions. Some treatment pathways combined different types of treatment, such as cognitive behavioral therapy and pharmacotherapy.

Selection of Patients

All outpatients from GGZ Friesland with unipolar depression as primary diagnosis (≥5 symptoms according to a clinical diagnostic interview based on DSM-IV[30]) at intake were selected for this study. Patients with bipolar disorder were excluded. Among those selected, only those patients with both pre-treatment and post-treatment routine outcome monitoring (ROM) scores available were included in the analysis (see Data Sources). ROM scores were considered appropriate if assessed at both the start and end of treatment (+/- 3 months).

The Medical Ethics Review Board (METc UMCG) concluded that the current research was exempted from full review according to the Dutch Medical Research with Human Subjects Law (WMO) as data were taken from the medical files of a group of patients. Patients were given the opportunity to opt out of the use of their anonymized data in the research database.

Data Sources

Treatment data

Treatment data were obtained on 920 patients treated for unipolar depression. Patients

were offered one of the following treatments: psychotherapy alone, pharmacotherapy alone or a combination of both psychotherapy and pharmacotherapy. Psychotherapy consisted of cognitive behavioral therapy (CBT) or interpersonal psychotherapy (IPT). Pharmacotherapy consisted of treatment with antidepressants alone or with additional coaching and support. The combination therapy consisted of treatment with antidepressants and CBT or IPT. The available demographic variables were age and gender. The available treatment information consisted of treatment type, starting and ending dates of treatment, number of treatment sessions recommended and completed, and treatment duration.

Outcome data

GGZ Friesland used ROM questionnaires to evaluate treatment and measure patients' progress. Within the AGT program for depression, the Inventory of Depressive Symptomatology (IDS-SR30) and the Outcome Questionnaire (OQ-45) were used.

The *Inventory of Depressive Symptomatology* is a self-report instrument to assess the severity of depression symptoms[31,32]. The IDS-SR30 contains 30 questions; summing responses of the items yields a total score. A maximum score of 84 can be obtained, with scores at the high end indicating more severe depression (mild: 14-25; moderate: 26-38; severe: 39-84). We used the cut-off score of the original English version to assess remission (scores <14), as the psychometric properties of the Dutch version have not been investigated.

The *Outcome Questionnaire* is a 45-item self-report instrument designed for repeated measurement of a patient's status throughout a course of treatment and upon treatment termination[33]. It measures functioning in three domains (symptom distress, interpersonal functioning and social role) and has proved sensitive to changes in psychological distress over short periods of time. A higher total score indicates more psychological symptoms, difficulties in interpersonal functioning and inadequacy in tasks related to a patient's employment, family roles or leisure. The OQ-45 was validated in the Dutch population; hence, we used the cut-off score for the Dutch OQ-45 to assess remission (scores <55).

Outcome Measures

The main outcome measures were remission rates and response rates for the three treatment types within the AGT program. Patients were remitted when they no longer met the clinical cut-off point of the relevant questionnaire upon treatment termination. Patients who experienced at least a 50% reduction in score compared to baseline were considered responders. The IDS-SR30 score was chosen as the preferred instrument. If no IDS-SR30 score was available, we used the total score of the OQ-45 questionnaire.

To compare the effectiveness of the AGT program for depression with the efficacy reported by meta-analyses, recent meta-analyses were identified. A graphical illustration

of the criteria used for the selection of studies for efficacy calculation is displayed in Figure 1. We searched for meta-analyses in PubMed, PsychINFO and the Cochrane database starting in 2012. We included studies on the following treatments: (a) the most frequently used pharmacological interventions of GGZ Friesland, that is, one of three classes of drugs (selective serotonin reuptake inhibitors, selective serotonin and noradrenalin reuptake inhibitors, and tricyclic antidepressants); (b) cognitive behavioral therapies; (c) interpersonal psychotherapies; and (d) combination therapies (CBT or IPT combined with pharmacotherapy).

Many meta-analyses reported effects in terms of odds ratios (ORs) or standardized mean differences (Cohen's *d* or Hedges' *g*). We selected only meta-analyses that reported the proportion of remitters and/or responders as outcome, and in which the individual studies and their sample sizes were reported. To prevent overlap in studies, we removed all duplicate studies. For each treatment type, an overall remission and response rate was calculated using the total sample size of the studies found in the meta-analyses.

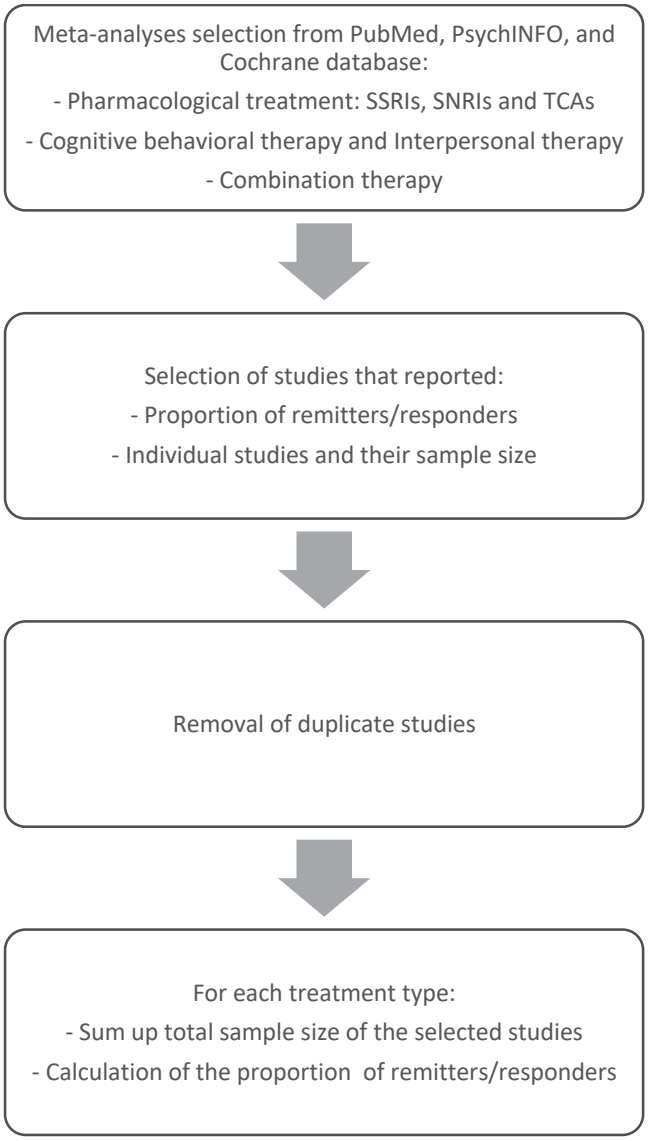
Treatment continuity was included as the binary variable "early terminators versus completers of the recommended number of treatment sessions." Patients were considered early terminators if they did not complete the recommended number of treatment sessions. Patients were considered completers if they had completed the recommended number of treatment sessions or more.

Statistical Analysis

Treatment data were linked to ROM data. The researchers received the anonymized dataset without personal identifiers. Remission and response rates were calculated as the percentage of patients achieving remission and response upon treatment termination. 95% Confidence intervals were calculated for the outcomes of the treatments within the AGT program and the outcomes found in the meta-analyses and we checked for overlap. To investigate possible selection bias, we first compared the baseline characteristics of the patients included in the study with those of patients with no or one ROM measurement available at the start or end of treatment. Independent *t*-tests, chi-squared tests or Mann Whitney two-sample tests were used to compare the groups. A two-sided test with a p value of <0.05 was considered statistically significant.

Second, we investigated whether the completers of the recommended number of treatment sessions had higher remission and response rates compared to the early terminators. The Kolmogorov-Smirnov test was used to determine whether the distribution of completers differed between the remitters/non-remitters and the responders/non-responders. All data were analyzed with STATA/SE version 14.1 (StataCorp LLC, College Station, Texas, USA).

Figure 1. Graphical illustration of the criteria used for the selection of studies for efficacy calculation



RESULTS

Patient Characteristics

Of the 920 patients who were offered one of the three treatments, 351 patients had both pre-treatment and post-treatment ROM scores and could thus be included in our analyses. The baseline characteristics of the included group were compared with the

baseline characteristics of the group with no or one ROM score (n = 569) (Table 1). The included group was on average younger (t(918)=2.37, p = 0.02) and had completed a significantly higher number of therapy sessions (t(768.42)=-5.56, p<0.01) compared to the group with incomplete ROM scores. Gender distribution, depression severity at baseline and treatment duration were comparable between the two groups. The average treatment duration was 232 days and the average time period between the assessment of the end of treatment scores and the actual end of treatment date was 33 days.

In the included group of patients, 71.5% received psychotherapy, 24.8% pharmacotherapy and 3.7% combination therapy, versus 57.8%, 36.6% and 5.6%, respectively, in the excluded group. These proportions differed significantly between the groups of patients (χ^2 (2, N = 920)=17.47, p<0.01).

Table 1. Baseline characteristics of patients with and without complete pre-treatment and post-treatment measurements

	Patients with complete pre-treatment and post-treatment measurements			Patients without complete pre- treatment and post-treatment measurements			*p*-value
	N	**Mean (s.d.)**	**Median**	***N***	**Mean (s.d.)**	**Median**	
Age at study entry	351	40.3 (13.0)	42	569	42.5 (13.2)	45	*0.02*
Gender, % male, (*n*)	351	44.4 (156)		569	45.5 (259)		0.75
Baseline severity IDS-SR30	108	33.2 (12.7)	33.5	63[a]	31.7 (11.4)	34	0.43
Baseline severity OQ-45	243	89.3 (21.0)	90	237[a]	87.8 (21.8)	88	0.42
Number of sessions	351	13.4 (7.2)	13	569	10.6 (7.5)	9	*<0.01*
Treatment duration, days	351	232.4 (115.4)	219	569	246.3 (149.8)	220	0.11

IDS-SR30, Inventory of Depressive Symptomatology self-rated; OQ-45, Outcome Questionnaire
a. Of patients with an available baseline score

Effectiveness of the AGT Program and Efficacy derived from the Meta-Analyses

The overall effectiveness of the AGT program was 36.8% (95% CI: 31.7-42.0) in terms of remission and 23.9% (95% CI: 19.5-28.4) in terms of response. Table 2 presents the patient characteristics for each treatment type, and Table 3 compares the effectiveness of the AGT program to the overall efficacy derived from the meta-analyses. Nine meta-analyses reported remission and/or response rates. Six studies focused on the efficacy of pharmacotherapy[34-39]. Four studies focused on the efficacy of psychotherapy[39-42], and one study reported the efficacy of combination therapy[39]. To achieve a more reliable comparison for combination therapy, four additional meta-analyses were added dating from before 2012[43-46].

The remission rates for pharmacotherapy and combination therapy in the AGT program were comparable to the remission rates found in the meta-analyses: 31.0% (95% CI: 21.5-41.9) versus 41.6% (95% CI: 40.5-42.7) for pharmacotherapy and 46.2% (95% CI: 19.2-74.9) versus 42.9% (95% CI: 40.2-45.7) for combination therapy. For psychotherapy,

we found a lower remission rate in the AGT program compared to the meta-analyses, that is, 38.3% (95% CI: 32.2-44.6) versus 52.7% (95% CI: 50.8-54.5).

Table 2. Characteristics of patients in the three treatment groups (*n*=351)

	Psychotherapy	Pharmacotherapy	Combination therapy
N	251	87	13
Age, mean (s.d.)	38.1 (12.9)	46.5 (11.3)	42.2 (12.0)
Gender, % male (*n*)	39.0 (98)	57.5 (50)	61.5 (8)
Baseline score IDS-SR30			
mean (s.d.)	33.2 (12.6)	31.1 (10.4)	47.0 (24.8)
% mildly depressed	27.9	36.8	33.3
% moderately depressed	33.7	42.1	0
% severely depressed	38.4	21.1	66.7
Baseline score OQ-45[a]			
mean (s.d.)	86.9 (19.8)	93.7 (22.1)	99.5 (26.7)
Treatment duration, days			
mean (s.d.)	228.5 (106.8)	235.9 (137.4)	283.2 (109.3)
median	220	204	259
% recommended sessions completed (s.d.)	82.6 (41.4)	82.7 (40.7)	91.6 (49.5)

IDS-SR30, Inventory of Depressive Symptomatology self-rated; OQ-45, Outcome Questionnaire
a. If no baseline score of IDS-SR30 was available, the OQ-45 was used to assess baseline severity

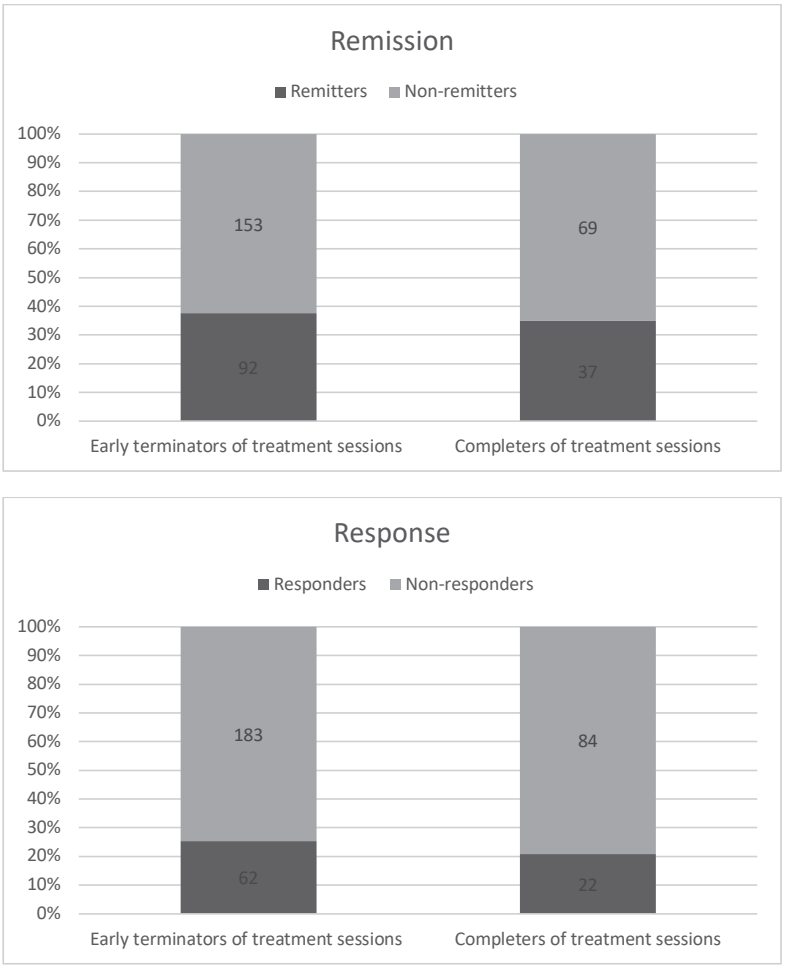
Table 3. Effectiveness of the AGT program and the efficacy results derived from meta-analyses[a]

	Effectiveness AGT program		Efficacy meta-analyses	
	Remission rates (%)	95% CI	Remission rates (%)	95% CI
Psychotherapy	38.3	32.2-44.6	52.7	50.8-54.5
Pharmacotherapy	31.0	21.5-41.9	41.6	40.5-42.7
Combination therapy	46.2	19.2-74.9	42.9	40.2-45.7
	Effectiveness AGT program		**Efficacy from meta-analyses**	
	Response rates (%)	95% CI	Response rates (%)	95% CI
Psychotherapy	23.9	18.8-29.7	49.3	45.8-52.9
Pharmacotherapy	20.7	12.7-30.7	56.1	55.3-56.8
Combination therapy	46.2	19.2-74.9	50.8	44.7-56.8

a. Rates derived from all studies mentioned in the meta-analyses cited in section Effectiveness of the AGT Program and Efficacy derived from the Meta-Analyses

For all three treatments, the response rates in the AGT program were lower than the remission rates in the AGT program. For psychotherapy and pharmacotherapy, the response rates in the AGT program were lower than those reported in the meta-analyses: 23.9% (95% CI: 18.8-29.7) versus 49.3% (95% CI: 45.8-52.9) for psychotherapy and 20.7% (95% CI: 12.7-30.7) versus 56.1% (95% CI: 55.3-56.8) for pharmacotherapy. The response

Figure 2. Frequencies and proportion of patients achieving remission and response among early terminators and among completers



rate for combination therapy in the AGT program was comparable to that found in the meta-analyses, that is, 46.2% (95% CI: 19.2-74.9) versus 50.8% (95% CI: 44.7-56.8).

Treatment Continuity

Figure 2 presents the frequencies and proportions of patients achieving remission/non-remission and response/non-response among the early terminators and completers. The completers did not differ significantly from the early terminators in remission and response rates ($\chi^2(1)=0.22$, $p = 0.64$ and $\chi^2(1)=0.84$, $p = 0.36$, respectively). The Kolmogorov-Smirnov test showed no significant differences in the distribution of treatment continuity between the remitters/non-remitters group ($D = 0.08$, $p = 0.61$) and the responders/non-responders group ($D = 0.09$, $p = 0.64$).

We investigated whether remitters differed in characteristics from the non-remitters within the groups of early terminators and completers (Table 4). Remitters in the early terminators group had a significantly lower baseline score compared to the non-remitters in this group (IDS-SR30: $t(77)=4.32$, $p<0.01$; OQ-45: $t(164)=3.26$, $p<0.01$), while the distribution of age and gender were not significantly different. In the completers group, the remitters were on average younger ($t(104)=2.27$, $p = 0.03$) and had lower baseline severity scores (IDS-SR30: $t(27)=2.50$, $p = 0.02$; OQ-45: $t(75)=1.50$, $p = 0.14$) compared to the non-remitters.

Table 4. Characteristics of patients achieving remission/non-remission in the groups of early terminators and completers

	Early Terminators			Completers		
	Remission (*n*=92)	**No Remission (*n*=153)**	***p*-value**	**Remission (n=37)**	**No remission (n=69)**	***p*-value**
Age mean (s.d.)	38.3 (11.9)	41.1 (13.5)	0.10	37.2 (13.9)	43.1 (12.1)	*0.03*
Gender, % male	42.4	46.4	0.54	32.4	49.3	0.10
Baseline IDS-SR30 severity mean (s.d.)	24.4 (10.3)	36.4 (12.0)	<0.01	27.5 (14.4)	38.8 (9.9)	*0.02*
Baseline OQ-45 severity mean (s.d.)	81.7 (22.7)	92.1 (18.4)	*<0.01*	87.6 (20.6)	95.2 (21.4)	0.14
Proportion of patients with very severe depression, %	8.0	46.3	*<0.01*	20.00	52.63	0.13

IDS-SR30, Inventory of Depressive Symptomatology self-rated; OQ-45, Outcome Questionnaire

DISCUSSION

This study examined the effectiveness of an AGT program for depression in regular specialized care. The overall remission rate of the AGT program was 36.8%, and the overall response rate was 23.9%. Comparing these results with those of meta-analyses, we found lower to comparable treatment effects in clinical practice. Moreover, among both the early terminators and the completers of the prescribed number of treatment sessions, a similar proportion of patients achieved remission and response.

In the current study, all patients were treated regardless of their depression severity, disease duration or other criteria that could make them ineligible for participation in RCTs. Patients in our sample might therefore not completely resemble patients in RCTs. In our study population, patients had moderate depression at baseline, while most meta-analyses reported a mean baseline score of moderate to severe depression. An

additional analysis of our data excluding patients with mild depression showed slightly lower remission (32.8%) and response rates (23.4%).

In comparison with another study in which ROM data were used to assess treatment effectiveness[5], remission rates for the different treatments within the AGT program were higher in our population (31-46% vs. 17-27%), while response rates were comparable (21-46% vs. 29-32%). The AGT program put strong emphasis on guideline adherence, which might explain its higher remission rates. Our overall remission rate (37%) was comparable to the remission rate found in a clinical practice study (35%), where patients in a naturalistic setting were given their choice of treatment[7].

The remission rate for pharmacotherapy found in this study corresponds to other algorithm-based studies. For instance, the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) trial, using a randomized controlled design resembling clinical practice, found a 28-33% remission rate [6], which is comparable to the 31% for pharmacotherapy in this study. A naturalistic study reporting on a pharmacologic treatment algorithm, found an overall remission rate of 30%[47].

Reasons are unclear for our finding of a lower remission rate for psychotherapy (38%) compared to the meta-analyses. For psychotherapy, the meta-analyses reported remission rates in the range of 33-59%. In general, we found lower response rates compared to remission rates for the treatment options within the AGT program. This was somewhat surprising, as response rates are usually higher than remission rates. We used the conventional response definition of a ≥ 50% change compared to the baseline, as this is best-known and widely used in trials[48]. However, in our study sample, response was relatively more difficult to achieve in patients with low to moderate baseline scores. This may explain why the response rates found were lower than remission rates. Remitted patients were not included in our definition of response, in accordance with the definition of response as the critical endpoint for defining improvement in acute treatment studies[49]. Including remitted patients in the definition of response resulted in a response rate of 38.5% (95% CI: 33.4-43.6). The use of different measurement scales for depression by GGZ Friesland and RCTs might complicate comparison, although studies show similar responsiveness between disorder-specific (and generic specific) instruments for depression[50,51] and equal responsiveness of IDS-SR total score and OQ-45 subscale [50].

Finally, we hypothesized that patients that completed their treatment sessions had better treatment outcomes than patients that did not complete their treatment sessions. In present study, some patients achieved remission/response with a lower number of recommended treatment sessions, while other patients extended their number of treatment sessions in order to achieve remission/response. This is in line with the previous dose-response literature, which indicates that responsiveness to treatment can differ for different groups of patients[52,53].

Limitations

The use of a naturalistic data sample has the advantage that it yields important information about real-world effectiveness. However, it is also accompanied by several limitations. First, the allocation to treatment was not randomized in this naturalistic sample. No inclusion or exclusion selection criteria were applied, other than a primary depression diagnosis and availability of both pre-treatment and post-treatment measurement scores. Besides, there was a lack of medical histories and other demographic data, which made it difficult to control for confounding. This restricted our investigation of the comparability of patient characteristics, both between the included treatment groups and between the groups with and without both pre-treatment and post-treatment ROM scores.

The second limitation concerns the substantial proportion of patients who were lost to follow-up and for whom no reason for treatment termination could be identified. We observed a higher proportion of patients receiving pharmacotherapy in the group with incomplete pre-treatment and post-treatment ROM measurements (36.6% vs. 24.8%). This matches findings of a recent meta-analysis which reported that patients prescribed pharmacotherapy were more likely to drop out than those who received psychotherapy[54]. This is possibly a reason for the unavailability of ROM measurements upon treatment termination.

A third limitation is the unavailability of data on antidepressant dosage and specification of drug type of patients receiving pharmacotherapy. Therefore, we could not investigate the effectiveness of drug dosages and combinations of drugs.

Finally, only a small group of patients received combination therapy. Therefore, these results should be interpreted with caution. In the AGT program, the majority of patients with unipolar depression were moderately depressed, and according to the algorithm, only patients with complicating factors (e.g., personality disorders and psychotic features) should receive combination therapy. Furthermore, most of the meta-analyses identified for combination therapy reported remission rates only, and few studies reported the response rate for combination therapy.

Clinical Implications

An AGT program with a strong emphasis on guideline adherence can approach the efficacy found in RCTs because it leads clinicians to choose the right treatment type for the patient concerned. By making use of treatment algorithms, inappropriate treatment variance between clinicians may be reduced and treatment outcomes enhanced.

Linking ROM data to different types of treatment enables treatment outcomes to be assessed in a naturalistic setting and in a heterogeneous population, without exclusion of patients who might normally be ineligible for inclusion in RCTs. Acknowledging the value of ROM data in clinical practice is relevant for both clinicians and patients. After all, better monitoring practices can provide opportunities for improving care

and treatment outcomes[55]. Routinely collected administrative data can yield valuable results if collected in a structured and consistent way.

A fixed number of treatment sessions does not seem to suit all individual patients. Although the majority of treatment algorithms (and RCTs) are developed with a pre-defined treatment duration for pragmatic reasons, the optimal treatment length can vary between patients.

Overall, this study found that an AGT program for specialized treatment of depression in daily practice, combining stepped and matched care, and with an emphasis on guideline adherence by clinicians, can approach the efficacy reported in RCTs.

REFERENCES

1. Cuijpers P, Berking M, Andersson G, Quigley L, Kleiboer A, Dobson KS. A meta-analysis of cognitive-behavioural therapy for adult depression, alone and in comparison with other treatments. *Can J Psychiatry*. 2013;58(7):376-385.
2. Karyotaki E, Smit Y, Holdt Henningsen K, et al. Combining pharmacotherapy and psychotherapy or monotherapy for major depression? A meta-analysis on the long-term effects. *J Affect Disord*. 2016;194:144-152.
3. Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: A systematic review and network meta-analysis. *Lancet*. 2018.
4. Gaynes BN, Warden D, Trivedi MH, Wisniewski SR, Fava M, Rush AJ. What did STAR*D teach us? results from a large-scale, practical, clinical trial for patients with depression. *Psychiatr Serv*. 2009;60(11):1439-1445.
5. van der Lem R, van der Wee NJ, van Veen T, Zitman FG. Efficacy versus effectiveness: A direct comparison of the outcome of treatment for mild to moderate depression in randomized controlled trials and daily practice. *Psychother Psychosom*. 2012;81(4):226-234.
6. Trivedi MH, Rush AJ, Wisniewski SR, et al. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: Implications for clinical practice. *Am J Psychiatry*. 2006;163(1):28-40.
7. Peeters F, Huibers M, Roelofs J, et al. The clinical effectiveness of evidence-based interventions for depression: A pragmatic trial in routine practice. *J Affect Disord*. 2013;145(3):349-355.
8. Keitner GI, Posternak MA, Ryan CE. How many subjects with major depressive disorder meet eligibility requirements of an antidepressant efficacy trial? *J Clin Psychiatry*. 2003;64(9):1091-1093.
9. Zimmerman M, Mattia JI, Posternak MA. Are subjects in pharmacological treatment trials of depression representative of patients in routine clinical practice? *Am J Psychiatry*. 2002;159(3):469-473.
10. Hepner KA, Rowe M, Rost K, et al. The effect of adherence to practice guidelines on depression outcomes. *Ann Intern Med*. 2007;147(5):320-329.
11. Guo T, Xiang YT, Xiao L, et al. Measurement-based care versus standard care for major depression: A randomized controlled trial with blind raters. *Am J Psychiatry*. 2015;172(10):1004-1013.
12. Adli M, Bauer M, Rush AJ. Algorithms and collaborative-care systems for depression: Are they effective and why? A systematic review. *Biol Psychiatry*. 2006;59(11):1029-1038.
13. Bauer M, Pfennig A, Linden M, Smolka MN, Neu P, Adli M. Efficacy of an algorithm-guided treatment compared with treatment as usual: A randomized, controlled study of inpatients with depression. *J Clin Psychopharmacol*. 2009;29(4):327-333.
14. Melfi CA, Chawla AJ, Croghan TW, Hanna MP, Kennedy S, Sredl K. The effects of adherence to antidepressant treatment guidelines on relapse and recurrence of depression. *Arch Gen Psychiatry*. 1998;55(12):1128-1132.
15. Schneider F, Harter M, Brand S, et al. Adherence to guidelines for treatment of depression in in-patients. *Br J Psychiatry*. 2005;187:462-469.
16. Dekker J, Molenaar PJ, Kool S, Van Aalst G, Peen J, de Jonghe F. Dose-effect relations in time-limited combined psycho-pharmacological treatment for depression. *Psychol Med*. 2005;35(1):47-58.
17. Forde F, Frame M, Hanlon P, et al. Optimum number of sessions for depression and anxiety. *Nurs Times*. 2005;101(43):36-40.
18. Kachele H. How long does psychotherapy last? *Psychother Psychosom Med Psychol*. 1990;40(5):148-151.

19. Kadera SW, Lambert MJ, Andrews AA. How much therapy is really enough? : A session-by-session analysis of the psychotherapy dose-effect relationship. *J Psychother Pract Res*. 1996;5(2):132-151.

20. Molenaar PJ, Boom Y, Peen J, Schoevers RA, Van R, Dekker JJ. Is there a dose-effect relationship between the number of psychotherapy sessions and improvement of social functioning? *Br J Clin Psychol*. 2011;50(3):268-282.

21. Shapiro DA, Barkham M, Rees A, Hardy GE, Reynolds S, Startup M. Effects of treatment duration and severity of depression on the effectiveness of cognitive-behavioral and psychodynamic-interpersonal psychotherapy. *J Consult Clin Psychol*. 1994;62(3):522-534.

22. Stulz N, Lutz W, Kopta SM, Minami T, Saunders SM. Dose-effect relationship in routine outpatient psychotherapy: Does treatment duration matter? *J Couns Psychol*. 2013;60(4):593-600.

23. Akerblad AC, Bengtsson F, Ekselius L, von Knorring L. Effects of an educational compliance enhancement programme and therapeutic drug monitoring on treatment adherence in depressed patients managed by general practitioners. *Int Clin Psychopharmacol*. 2003;18(6):347-354.

24. Cahill J, Barkham M, Hardy G, et al. Outcomes of patients completing and not completing cognitive therapy for depression. *Br J Clin Psychol*. 2003;42(Pt 2):133-143.

25. von Knorring L, Åkerblad A, Bengtsson F, Carlsson Å, Ekselius L. Cost of depression: Effect of adherence and treatment response. *European psychiatry*. 2006;21(6):349-354.

26. Adli M, Berghofer A, Linden M, et al. Effectiveness and feasibility of a standardized stepwise drug treatment regimen algorithm for inpatients with depressive disorders: Results of a 2-year observational algorithm study. *J Clin Psychiatry*. 2002;63(9):782-790.

27. Adli M, Wiethoff K, Baghai TC, et al. How effective is algorithm-guided treatment for depressed inpatients? results from the randomized controlled multicenter german algorithm project 3 trial. *Int J Neuropsychopharmacol*. 2017;20(9):721-730.

28. Trivedi MH, Rush AJ, Crismon ML, et al. Clinical results for patients with major depressive disorder in the texas medication algorithm project. *Arch Gen Psychiatry*. 2004;61(7):669-680.

29. Spijker J, Bockting CLH, Meeuwissen JAC, Vliet IM van, Emmelkamp PMG, Hermens MLM, Balkom ALJM on behalf of the Guideline Development Group Multidisciplinary guideline development for anxiety/depressive disorder. [Multidisciplinary guideline depression (second revision): Guideline for diagnostics and treatment of adults with a major depressive disorder]. [[in Dutch]]. 2011;Utrecht: Trimbos-instituut.[in Dutch].

30. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders (DSM-IV)*. 4th edition ed. Washington DC: ; 1994.

31. Rush AJ, Giles DE, Schlesser MA, Fulton CL, Weissenburger J, Burns C. The inventory for depressive symptomatology (IDS): Preliminary findings. *Psychiatry Res*. 1986;18(1):65-87.

32. Rush AJ, Gullion CM, Basco MR, Jarrett RB, Trivedi MH. The inventory of depressive symptomatology (IDS): Psychometric properties. *Psychol Med*. 1996;26(03):477-486.

33. Lambert MJ, Burlingame GM, Umphress V, et al. The reliability and validity of the outcome questionnaire. *Clinical Psychology & Psychotherapy*. 1996;3(4):249-258.

34. Cipriani A, Purgato M, Furukawa TA, et al. Citalopram versus other anti-depressive agents for depression. *Cochrane Database Syst Rev*. 2012;(7):CD006534. doi(7):CD006534.

35. Cipriani A, Koesters M, Furukawa TA, et al. Duloxetine versus other anti-depressive agents for depression. *Cochrane Database Syst Rev*. 2012;10:CD006533.

36. Leucht C, Huhn M, Leucht S. Amitriptyline versus placebo for major depressive disorder. *Cochrane Database Syst Rev*. 2012;12:CD009138.

37. Magni LR, Purgato M, Gastaldon C, et al. Fluoxetine versus other types of pharmacotherapy for depression. *Cochrane Database Syst Rev*. 2013;(7)(7):CD004185.

38. Purgato M, Papola D, Gastaldon C, et al. Paroxetine versus other anti-depressive agents for depression. *Cochrane Database Syst Rev*. 2014;(4):CD006531. doi(4):CD006531.

39. Linde K, Rücker G, Sigterman K, et al. Comparative effectiveness of psychological treatments for depressive disorders in primary care: Network meta-analysis. *BMC Fam Pract*. 2015;16:10.1186/s12875-015-0314-x.

40. Jakobsen JC, Hansen JL, Simonsen S, Simonsen E, Gluud C. Effects of cognitive therapy versus interpersonal psychotherapy in patients with major depressive disorder: A systematic review of randomized clinical trials with meta-analyses and trial sequential analyses. *Psychol Med*. 2012;42(7):1343-1357.

41. Shinohara K, Honyashiki M, Imai H, et al. Behavioural therapies versus other psychological therapies for depression. *Cochrane Database Syst Rev*. 2013;(10):CD008696. doi(10):CD008696.

42. Johnsen TJ, Friborg O. The effects of cognitive behavioral therapy as an anti-depressive treatment is falling: A meta-analysis. *Psychol Bull*. 2015;141(4):747-768.

43. de Maat SM, Dekker J, Schoevers RA, de Jonghe F. Relative efficacy of psychotherapy and combined therapy in the treatment of depression: A meta-analysis. *Eur Psychiatry*. 2007;22(1):1-8.

44. Pampallona S, Bollini P, Tibaldi G, Kupelnick B, Munizza C. Combined pharmacotherapy and psychological treatment for depression: A systematic review. *Arch Gen Psychiatry*. 2004;61(7):714-719.

45. Thase ME, Greenhouse JB, Frank E, et al. Treatment of major depression with psychotherapy or psychotherapy-pharmacotherapy combinations. *Arch Gen Psychiatry*. 1997;54(11):1009-1015.

46. Wexler BE, Cicchetti DV. The outpatient treatment of depression. implications of outcome research for clinical practice. *J Nerv Ment Dis*. 1992;180(5):277-286.

47. Hawley CJ, Pattinson HA, Quick SJ, et al. A protocol for the pharmacologic treatment of major depression. A field test of a potential prototype. *J Affect Disord*. 1998;47(1-3):87-96.

48. Nierenberg AA, DeCecco LM. Definitions of antidepressant treatment response, remission, nonresponse, partial response, and other relevant outcomes: A focus on treatment-resistant depression. *J Clin Psychiatry*. 2001;62 Suppl 16:5-9.

49. Keller MB. Past, present, and future directions for defining optimal treatment outcome in depression: Remission and beyond. *JAMA*. 2003;289(23):3152-3160.

50. de Beurs E, Vissers E, Schoevers R, Carlier IVE, van Hemert AM, Meesters Y. Comparative responsiveness of generic versus disorder-specific instruments for depression: An assessment in three longitudinal datasets. *Depress Anxiety*. 2018.

51. Corruble E, Legrand JM, Duret C, Charles G, Guelfi JD. IDS-C and IDS-sr: Psychometric properties in depressed in-patients. *J Affect Disord*. 1999;56(2-3):95-101.

52. Hansen NB, Lambert MJ, Forman EM. The psychotherapy dose-response effect and its implications for treatment delivery services. *Clinical Psychology: Science and Practice*. 2002;9(3):329-343.

53. Baldwin SA, Berkeljon A, Atkins DC, Olsen JA, Nielsen SL. Rates of change in naturalistic psychotherapy: Contrasting dose-effect and good-enough level models of change. *J Consult Clin Psychol*. 2009;77(2):203-211.

54. Swift JK, Greenberg RP, Tompkins KA, Parkin SR. Treatment refusal and premature termination in psychotherapy, pharmacotherapy, and their combination: A meta-analysis of head-to-head comparisons. *Psychotherapy (Chic)*. 2017;54(1):47-57.

55. Lambert MJ. *Prevention of treatment failure: The use of measuring, monitoring, and feedback in clinical practice.* American Psychological Association; 2015. https://books.google.nl/books?id=_TfUsgEACAAJ.

SUPPLEMENTARY MATERIAL

Supplemental information about the algorithm-guided treatment program for depression

In the algorithm-guided treatment (AGT) program for depression, patients diagnosed with depression were referred to small specialist outpatient teams consisting of a psychiatrist, one or two psychologists/psychotherapists, and several psychiatric nurses, all specialized in the treatment of depressive disorders.

The AGT program for depression consisted of a combination of stepped care and matched care, based on the Dutch multidisciplinary guideline for depression (2nd revision)[2], matching international guidelines for depression. The treatment algorithm derived from this guideline took into account several characteristics of the depressive disorder, including the number of previous episodes, treatment history, and the duration and severity of the current episode. For example, for a mild first onset depressive episode that has not responded to supportive interventions during the first three months, a short problem solving therapy of five sessions is recommended. In contrast, patients with a severe depression with psychomotor retardation receive a combination of pharmacotherapy and weekly coaching and support by a psychiatric nurse.

The treatment algorithm thus leads to a recommended treatment pathway. These treatment pathways were the main building blocks of the AGT program for depression and consisted of different treatments or combinations of treatments. Examples of treatment pathways are: stand-alone pharmacotherapy; stand-alone cognitive behavioral therapy; combination of pharmacotherapy and cognitive behavioral therapy. To optimize guideline adherence, deviation from the recommended treatment pathway was only allowed after consulting a senior team member.

Pharmacotherapy was based on the 5-step model from the above mentioned Dutch multidisciplinary guideline for depression. It recommends lithium addition after two consecutive non-successful 4-6 weeks trials with an antidepressant, switching to a nonselective MAO inhibitor as step 4, and electroconvulsive therapy as step 5.

2 Spijker J, Bockting CLH, Meeuwissen JAC, Vliet IM van, Emmelkamp PMG, Hermens MLM, Balkom ALJM on behalf of the Guideline Development Group Multidisciplinary guideline development for anxiety/depressive disorder. (2011). [Multidisciplinary guideline depression (second revision): Guideline for diagnostics and treatment of adults with a major depressive disorder]. Utrecht: Trimbos-instituut. ([in Dutch])