

Guidelines improve patient outcomes in specialised mental health care: A systematic review and meta-analysis

Kim Setkowski^{1,2}  | Kelly Boogert² | Adriaan W. Hoogendoorn¹ | Renske Gilissen² | Anton J. L. M. van Balkom¹

¹Department of Psychiatry, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam Public Health Research Institute and GGZ inGeest Specialised Mental Health Care, Amsterdam, The Netherlands

²113 Suicide Prevention, Amsterdam, The Netherlands

Correspondence

Kim Setkowski, Research Department, 113 Suicide Prevention, Paasheuvelweg 25, 1105BP Amsterdam Zuidoost, the Netherlands.

Email: k.setkowski@113.nl

Funding information

This meta-analysis and systematic review is part of the SUPRANET Care study and funded by the Ministry of Health Funding programme for Health Care Efficiency Research ZonMw (537001006). ZonMw is not involved in the design of the study and does not participate in the data collection, statistical analysis of the data and writing of the manuscript.

Abstract

Objective: The uptake of evidence-based guideline recommendations appears to be challenging. In the midst of the discussion on how to overcome these barriers, the question of whether the use of guidelines leads to improved patient outcomes threatens to be overlooked. This study examined the effectiveness of evidence-based guidelines for all psychiatric disorders on patient health outcomes in specialist mental health care. All types of evidence-based guidelines, such as psychological and medication-focused guidelines, were eligible for inclusion. Provider performance was measured as a secondary outcome. Time to remission when treated with the guidelines was also examined.

Method: Six databases were searched until 10 August 2020. Studies were selected, and data were extracted independently according to the PRISMA guidelines. Random effects meta-analyses were used to pool estimates across studies. Risk of bias was assessed according to the Cochrane Effective Practice and Organization of Care Review Group criteria. PROSPERO:CRD42020171311.

Results: The meta-analysis included 18 studies (N = 5380). Guidelines showed a positive significant effect size on the severity of psychopathological symptoms at the patient level when compared to treatment-as-usual (TAU) ($d = 0.29$, 95%-CI = (0.19, 0.40), $p < 0.001$). Removal of a potential outlier gave globally the same results with Cohen's $d = 0.26$. Time to remission was shorter in the guideline treatment compared with TAU (HR = 1.54, 95%-CI = (1.29, 1.84), $p = 0.001$, $n = 3$).

Conclusions: Patients cared for with guideline-adherent treatments improve to a greater degree and more quickly than patients treated with TAU. Knowledge on the mechanisms of change during guideline-adherent treatment needs to be developed further such that we can provide the best possible treatment to patients in routine care.

KEY WORDS

algorithm, evidence-based guideline, implementation, mental health care, meta-analysis

1 | INTRODUCTION

In the 1990s, the majority of people with common mental disorders who visited a mental health professional in the United States, Europe and Australia did not receive effective treatment.¹⁻³ This finding gave a new impetus in the development and implementation of evidence-based guidelines in mental health care. The policy makers and professionals of many countries became aware of the suboptimal quality of mental health care and tried to improve this quality by reducing treatment variation using a set of evidence-based guidelines.^{4,5} By combining the best available scientific evidence from randomised controlled trials (RCTs) with recommendations based on expert opinion, evidence-based guidelines may serve as a tool to deliver optimal mental health care.⁶ This position holds only when three important conditions are met: (i) the organisation of care needs to be adapted in order to enable professionals to deliver care in accordance with the guidelines, (ii) professionals must have the knowledge and capacities to execute the guideline recommendations in clinical practice and (iii) patients must accept the care they receive following the guideline recommendations. The uptake of guideline recommendations appears to be difficult, leading to a gap between the theoretically available evidence-based treatments recommended in existing guidelines and their use in routine care.⁷

Adherence to mental health guidelines may be improved by so-called 'multi-faceted' interventions aimed at guideline implementation in daily clinical practice. The implementation of evidence-based guidelines, however, is complex. Barriers at the organisational, professional and patient levels may prevent everyday use of guidelines.^{4,8-10} In addition, because the implementation of guidelines requires investments in time and resources, policy makers are able to distribute them only with the underuse of these guidelines as a consequence.¹¹ In the midst of the discussion on how to overcome these barriers and see that the use of guidelines increases, the question of whether the use of guidelines leads to improved patient outcomes threatens to be overlooked. This is especially important since the effect size of a certain guideline-adherent treatment delivered in clinical practice may be critically smaller than the effect size of the same treatment on which that guideline was based, evaluated in an efficacy study. This 'efficacy' vs 'effectiveness' study difference may be caused by a higher degree of heterogeneity of the samples included in the latter type of study associated with larger standard deviations and thus smaller effect sizes.¹²⁻¹⁴ It is therefore relevant to examine whether adherence to evidence-based guidelines in everyday mental health clinical practice is superior to TAU at the patient level.

Up until now, most studies in mental health care evaluating the effects of guideline implementation have mainly

Significant outcomes

- Our findings suggest that there is a statistically significant positive effect of evidence-based guidelines on patients' well-being when compared to TAU. These results seem to be robust for potential outliers. We also found that providers in the guideline condition were significantly more adherent to guideline recommendations when compared to providers delivering TAU, but this effect was small.
- Patients receiving guideline-adherent treatments also seemed to improve to a greater degree and more quickly (time to remission) compared to those treated with TAU. Apparently the benefit of delivering guideline-adherent treatments is not only that treatment outcome is superior to TAU, but that this superior improvement is also reached more quickly.
- A small correlation between patients' well-being and provider performance was found, indicating that provider-adherent care can lead to improved patient outcomes. However, it also suggests that other variables also influence outcomes at the patient level. Knowledge on the mechanisms of change during guideline-adherent treatment should therefore be developed further.

Limitations

- Studies are diverse in regard to their design, control conditions and measurements of the primary outcome, making a quantitative summary of the results difficult.
- Our secondary outcome was based on a diverse mix of provider performance variables varying between studies. Future studies should use uniformity in designing and reporting their studies.

focused on change in professional performance. Bauer (2002) conducted a qualitative systematic review with 41 studies published up to 2000 to investigate the extent to which guidelines on various mental disorders were implemented in the daily clinical practice of that time.¹⁵ Depending on study type, adherence rates varied from 27% in cross-sectional studies to 67% in RCTs. Only 11 of the 41 studies presented adherence as well as patient outcome data. Seven of these 11 studies found that improved patient outcome was associated with better adherence to the guidelines in daily practice. Although these conclusions were promising, the qualitative nature of the review precluded an estimate of the magnitude of the effect size between intervention and control

conditions. The review also lacked a quality assessment of the studies included.

A second review followed seven years later¹⁶ and included 18 studies (not statistically pooled) on diverse mental disorders published between 1966 and 2006, of which 13 presented provider performance and patient outcome data. The authors judged the quality of the studies moderate at best. Four of the 13 studies yielded a significant effect in favour of the guidelines at the provider and patient levels. Results suggested that disorder-specific features might determine whether the use of evidence-based guidelines is able to improve outcomes at the patient level. This interesting hypothesis has not yet been validated by the literature.

A few years ago, a third review, by Girlanda et al,¹⁷ was performed to examine the impact of evidence-based guidelines in diverse mental disorders on provider performance. Patient health was measured as a secondary outcome. This meta-analysis included 19 studies up to 2016. The authors judged the methodological quality of three-quarters of those 19 studies to be fair or good. This suggests that the quality of these studies increased over time. Seven out of the 19 studies could be pooled quantitatively. On their primary outcome (provider performance), no significant effect was found ($n = 6$). However, for patient health, they did find a small but significant improved effect on remission ($n = 3$) (OR = 1.5, 95%-CI = (1.03, 2.22)). This conclusion is not easy to interpret; apparently, patient outcomes may improve rather independent of provider performance. Perhaps the small number of analysed studies ($n = 6$ versus $n = 3$) may have influenced these results.

Based on the findings from previous studies, more clarity on the impact of evidence-based guidelines for the severity of psychopathological symptoms on the patient level in specialist mental health care is greatly needed. We therefore updated and extended the review by Girlanda et al¹⁷ so as to find more robust results for the effectiveness of evidence-based guidelines for all psychiatric disorders on patient health outcomes in specialised mental healthcare settings.

2 | METHODS

2.1 | Primary and secondary outcome measures

The primary outcome measure was the presence and severity of psychopathological symptoms measured at the patient level with a validated and reliable rating scale. If an included study measured patient health outcomes with more than one rating scale, we chose the scale that was most frequently used by other papers included in this meta-analysis and systematic review. Secondary outcome measures were time to remission measured at the patient level and adherence to the

guideline and/or algorithm at the provider level. We also explored whether the effectiveness of these guidelines/algorithms was related to the type of mental disorder,¹⁶ the country in which the study was conducted and the type of control condition.

2.2 | Search strategy

The registration number of this study can be found on PROSPERO (CRD42020171311). For this meta-analysis and systematic review, a literature search was performed in a number of databases, including MEDLINE, Embase, PsycINFO, Web of Science, Scopus and the Cochrane Central Register of Controlled Trials from inception to 10 August 2020 (for the complete search string, see Table S1). The unique records were imported into Covidence Systematic Review Software for the screening process. After importation, two independent assessors (KS and KB) screened the titles and abstracts independently on the inclusion criteria (see below). Next, both assessors conducted a full-text review of the remaining studies in Covidence. The reference lists of studies included in the full-text review were also searched for relevant articles. In case of disagreement, consensus was reached by discussion with a third assessor (AvB). This method section was reported in accordance with the PRISMA guidelines.¹⁸

2.3 | Eligibility criteria

Studies were eligible for inclusion if they evaluated the effectiveness of an evidence-based guideline or an algorithm-based intervention on the severity of psychopathological symptoms, measured at the patient level. Patient health outcomes should be measured with a validated and reliable rating scale. The guideline or algorithm should be mainly delivered in a mental healthcare setting. Included were adults (>18 years) with a primary mental disorder, diagnosed with a structured clinical interview according to ICD or DSM criteria. The following study designs were included: (Cluster-) Randomised Controlled Trial ((C-)RCT), (Cluster-) Non-Randomised Controlled Trial ((C-)NRCT) and Controlled Before and After Design (CBAD). Systematic reviews, meta-analyses, secondary analyses, cost-effectiveness studies, (conference) abstracts and case studies were excluded. We included papers written in English, German, French and Dutch.

2.4 | Definition of guideline

In this study, interventions were considered to be 'evidence-based guidelines' or 'treatment algorithms' when they

consisted of 'systematically developed statements to assist practitioner decisions about appropriate health care for specific clinical circumstances'¹⁹ (p. 38) and when these statements were used to provide treatment steps delivered in a particular order in case of non-response.²⁰ We considered an evidence-based guideline to be developed in a systematic way, when in the manuscript references were found on papers evaluating or developing steps of which the guideline was composed. Studies that met these predefined criteria were included regardless of the explicit use of the term 'guideline' or 'algorithm'. Studies evaluating only one component from a guideline (eg only one type of drug) were excluded.

2.5 | Data extraction and quality assessment

Two reviewers (KS and KB) independently extracted data using a predefined sheet based on the template developed by the Cochrane Effective Practice and Organisation of Care Review Group.²¹ Data extraction started on 9 May 2020. The following data were extracted: name first author, year of publication, country, study design, sample size, dropouts (percentage), type of mental disorder, study duration (weeks), type of intervention and control condition, adverse events, patient demographics (age and gender), rating scales for psychopathology at the patient level, provider performance, type of guideline, description of the intervention and time to remission at the patient level. The methodological quality of the studies was assessed independently by two researchers (KS and KB) using the EPOC data collection checklist.²² RCT, NRCT and CBAD studies were scored 'done', 'not done' or 'not clear' on the included domains from the EPOC data collection checklist (Figure 5). The data extraction and assessment of the methodological quality yielded one coding form per article. In case of differences between the two reviewers, the data in the manuscript were checked and consensus was subsequently reached by discussion with a third assessor (AvB).

2.6 | Effect size calculation

We calculated effect sizes Cohen's d for patient health outcome as well as provider performance for dimensional rating scales.²³⁻²⁵ When means and standard deviations were not reported, we estimated the effect sizes from p -values, t - or F -values, confidence intervals, ranges or other statistics.^{23,26,27} As suggested by Cochrane,²⁸ categorical outcomes were re-expressed as effect sizes by using the formula of Chinn²⁹ and Borenstein²³ (chap. 7). Please refer to Table S4 for the formulas used to estimate Cohen's d from sources other than

means and standard deviations. When a manuscript did not present data to calculate Cohen's d , the corresponding author was asked to provide them.^{30,31} Steinacher et al³⁰ provided us with their data, making it possible to include their study in our statistical analyses. However, we excluded Janssen et al,³¹ because we were unable to contact them and retrieve their data. The magnitude Cohen's d ³² was interpreted as follows: $d \leq 0.2$ small effect, $0.2 < d < 0.5$ medium effect and $d \geq 0.8$ large effect.

One study by Linden et al³³ included three conditions and compared the effects of an 'active' guideline implementation against two types of control condition, namely 'passive' guideline implementation and TAU. One C-NRCT³⁴ included multiple site pairs in their study, resulting in pairwise comparisons between the intervention condition and control condition at all sites. To compute an overall effect size, we used the inverse variance weighting method³⁵ to create a single pairwise comparison²⁷ (chap. 16.5.4).

2.7 | Data synthesis and meta-analysis

We could not pool the effect sizes of all interventions evaluated because some studies ($n = 11$) presented results at post-test, while others ($n = 7$) used change scores²⁷ (chap. 9). For each intervention, Cohen's d was presented in a so-called 'harvest plot'. A harvest plot is an appropriate method for synthesising evidence about differential effects of heterogeneous and complex interventions across different variables of interest, including study design and participant characteristics.^{36,37}

We meta-analysed only studies reporting results at post-test by using Comprehensive Meta-Analysis software (version 3) employing a random effect model. The I^2 statistic was calculated as an indicator of heterogeneity. A non-significant and rather low value of heterogeneity is a prerequisite for pooling effect sizes. We performed sensitivity analyses excluding outliers, defined as studies in which the 95% CI of the effect size did not overlap with the 95% CI of the pooled effect size.³⁸ Finally, before conducting the meta-analysis, preliminary analyses were performed to detect possible differences associated with study designs at both patient level and provider level.

Publication bias was examined with the Duval and Tweedie trim and fill procedure as well as Egger's test for the asymmetry of the funnel plot. Finally, we conducted a subgroup analysis in which studies were categorised into subgroups based on (i) type of mental disorder (schizophrenia vs depressive disorder), (ii) type of control condition (TAU vs 'passive' implementation) and (iii) country of origin (Germany vs United States). Statistical significance between subgroups was tested using formulas from Borenstein et al²³ (chap. 19) to calculate a p -value (for the formulas, see Table S4).

3 | RESULTS

3.1 | Study selection

The study selection process is illustrated in the PRISMA flow chart (Figure 1). A total of 7231 articles were retrieved from all databases. After the removal of 2909 duplicate articles, 4322 unique articles remained for the title and abstract review. Additionally, six articles were added from reference lists and assessed for full-text eligibility. A total of 107 articles were retrieved in full text and assessed for inclusion. Subsequently, 88 studies were excluded. Also, one study had to be excluded because we could not retrieve the necessary data,³⁰ leaving 18 studies which met the inclusion criteria.

3.2 | Characteristics of the studies

The 18 included studies were published between 2003 and 2017. The majority of the studies were performed in the

United States ($n = 7$) and Germany ($n = 6$). We included a total of five RCTs, three C-RCTs, six NRCTs, one C-NRCT and three CBADs, with, in total, 5380 patients. The studies had a duration of 12 to 156 weeks, with a median of 24 weeks. Five studies pertained to a total of 1073 patients with schizophrenia,^{16,31,34,39,40} another two studies included 573 patients with a bipolar disorder,^{41,42} seven studies included 1900 patients with a depressive disorder,^{33,43-48} one study focused on 537 patients with an alcohol use disorder⁴⁹ and three studies included mixed patient samples ($n = 767$).⁵⁰⁻⁵² Table S2 presents the main study characteristics. In 16 studies, the intervention condition was compared with TAU,^{16,31,33,39-49,51,52} while four studies evaluated head-to-head comparisons of an active with a passive implementation strategy.^{31,33,34,50} Because two studies^{31,33} presented two head-to-head comparisons (by comparing an active strategy vs TAU as well as an active strategy vs a passive strategy), the total number of comparisons is 20.

Various types of guidelines were evaluated. For example, eight studies were included investigating evidence-based

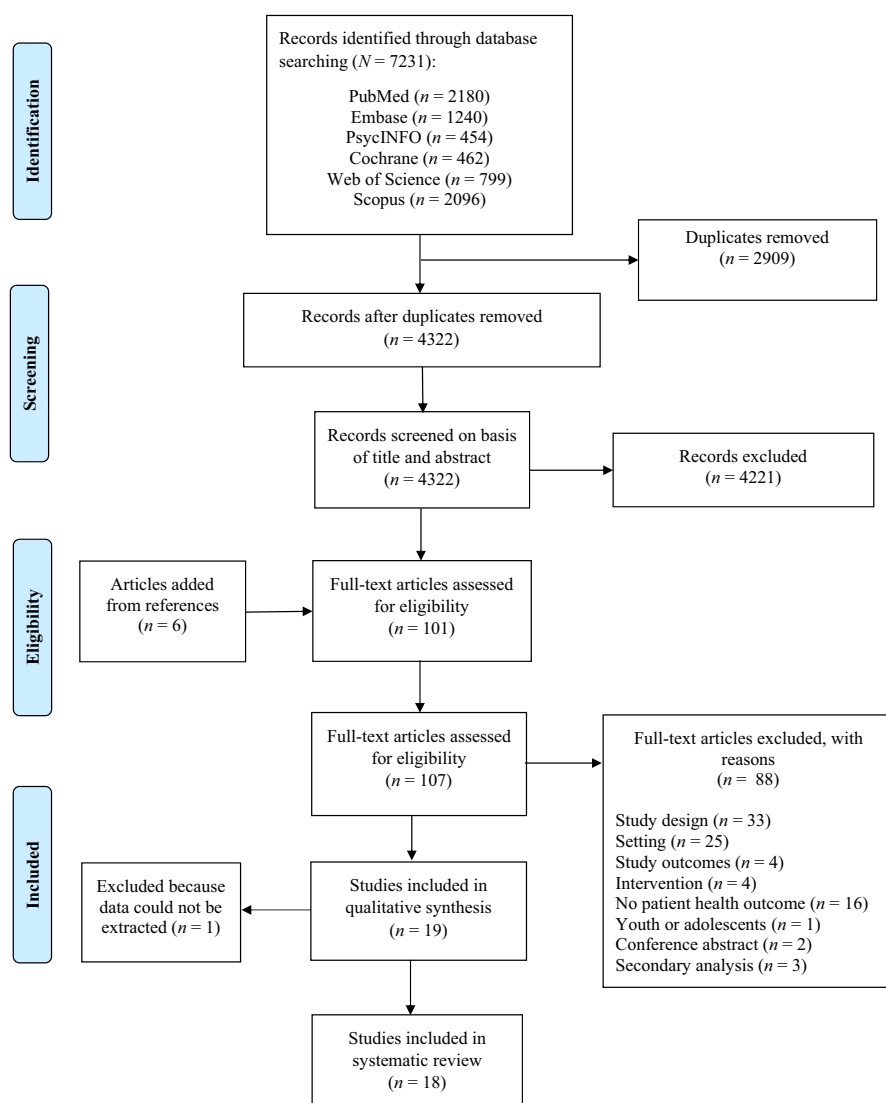


FIGURE 1 PRISMA flow diagram for study selection

(multidisciplinary) practice guidelines using at least four or more components at the same time.^{16,31,33,34,44,49,50,52} Furthermore, the other 10 included studies investigated evidence-based treatment algorithms developed with a structured methodology, using tailored strategies, based on expertise from national consensus groups and expert panels and written according to the latest scientific evidence.^{39-43,45-48,51} The content of the guidelines and algorithms evaluated in the included studies is presented in Table S3. As shown in Table S3, the majority of evidence-based guidelines were medication-focused.

Most guidelines (n = 8) pertained to medication algorithms, evaluated the status of the patient regularly and recommended that providers change the treatment module in case of non-response or deterioration.^{40,42,43,45-48,51} One guideline, investigated in the study by Bauer et al,⁴¹ included a psychotherapy algorithm in addition to a medication algorithm. Furthermore, a study by Choi et al³⁹ investigated the effects of a computerised decisional tool with algorithms using psychosocial strategies in addition to medication management. All included studies implemented the guidelines at a professional and a patient level, such as by training clinical staff; providing educational meetings and visits; distribution

of educational material via computer or in printed form; patient and family education; educational packages; information cards and booklets; and providing reminders, audits and written feedback to providers.

3.3 | Magnitude of the effect size Cohen's *d* for each intervention separately at patient and provider levels

3.3.1 | Patient level

Effect sizes for all included studies on outcomes at the patient level are presented in the harvest plot in Figure 2. Of 20 effect sizes Cohen's *d* on rating scales measuring severity of psychopathology, 17 were positive,^{16,33,34,39-46,48-51} with nine small,^{33,34,39,42-45,49,51} seven moderate^{16,33,40,41,46,48,50} and one large effect size.⁴⁷ Since the 95% CI crossed the null-line in 11 cases, eight effect sizes could be considered significantly positive in favour of the guideline.^{16,40-42,46,47,49,50} When these results were broken down into diagnostic categories, it appeared that two of six effect sizes in the treatment of patients with schizophrenia could be considered significantly

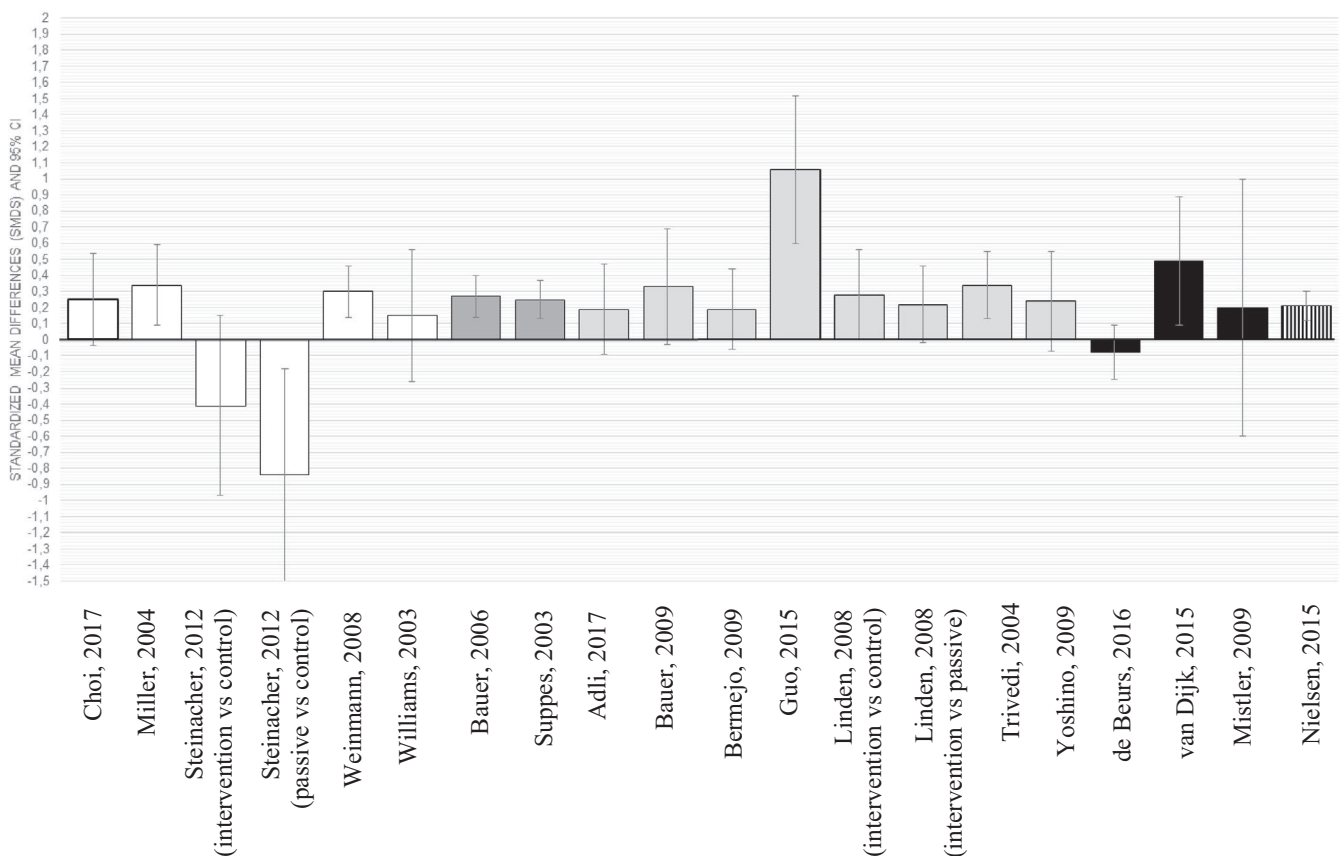


FIGURE 2 Harvest plot of all included studies (N = 18) which presented outcomes at the patient level. Depicted are the effect sizes Cohen's *d* (standard mean differences with 95% confidence intervals). Colours of bars are categorised as follows: (i) white: schizophrenia, (ii) dark grey: bipolar disorder, (iii) light grey: depressive disorder, (iv) black: mixed disorders and (v) vertical stripes: alcohol use disorder

positive^{16,40} vs both effect sizes in the treatment of bipolar disorder,^{41,42} two of eight effect sizes in depressive disorder,^{46,47} one of three in the mixed patient groups⁵⁰ and one study in the treatment of alcohol use disorder.⁴⁹

3.3.2 | Provider level

The harvest plots on provider behaviour (adherence to process indicators during guideline treatment vs control treatment) suggest that of 11 effect sizes Cohen's *d*, only three favoured adherence behaviour in the guideline treatment.^{33,50,51} In the remaining eight comparisons, the 95% CI included the null (please refer to Figure 3). The effect sizes at the patient and provider levels correlated only moderately (Spearman's *Rho* = 0.32, *n* = 11, *p* = 0.34).

3.4 | Meta-analysis at patient and provider levels: Evidence-based guidelines vs TAU

As previously described in the methods section, preliminary analyses were performed to detect possible differences associated with study designs at the patient level and provider level. For the patient level, 11 studies were included (three NRCTs, six RCTs and two CBADs), finding a significant difference between CBAD studies and NRCTs (mean difference = -0.618, 95%-CI = (-1.23, -0.005), *p* = 0.048) and between CBAD studies and RCTs (mean difference = -0.691,

95%-CI = (-1.33, -0.05), *p* = 0.015). CBAD studies were therefore excluded, remaining a total of nine studies for the meta-analysis (*n* = 9). At the provider level, seven studies were included in the preliminary analyses (two NRCTs, two RCTs and three CBADs), but no significant differences were found. Therefore, all studies (*n* = 7) were included in the meta-analysis for provider performance.

3.4.1 | Patient level

A total of nine studies comparing an intervention with a TAU condition were meta-analysed.^{33,39-41,43-45,47,48} As follows from Figure 4a, meta-analysis at the patient level yielded a statistically significant effect in favour of evidence-based guidelines (summary Cohen's *d* = 0.29, 95%-CI = (0.19, 0.40), *p* < 0.001). Among the pooled studies, a low and non-significant level of heterogeneity was found (*I*² = 34%, *p* = 0.145). However, as illustrated in Figure 4a, the effect size of the Guo et al study⁴⁷ showed no overlap with the overall pooled effect size and thus could be a potential outlier. Repeating the meta-analysis after the removal of this potential outlier gave globally the same results (*d* = 0.26, 95%-CI = (0.18, 0.34), *p* < 0.001) for between-study heterogeneity (*I*² = 0%, *p* = 0.988).

The funnel plot (Figure S1a) and Duvall and Tweedie's trim and fill procedure as well as Egger's test suggested no influence of publication bias among these nine studies. Egger's test was not significant (*p* = 0.228). Because this did not

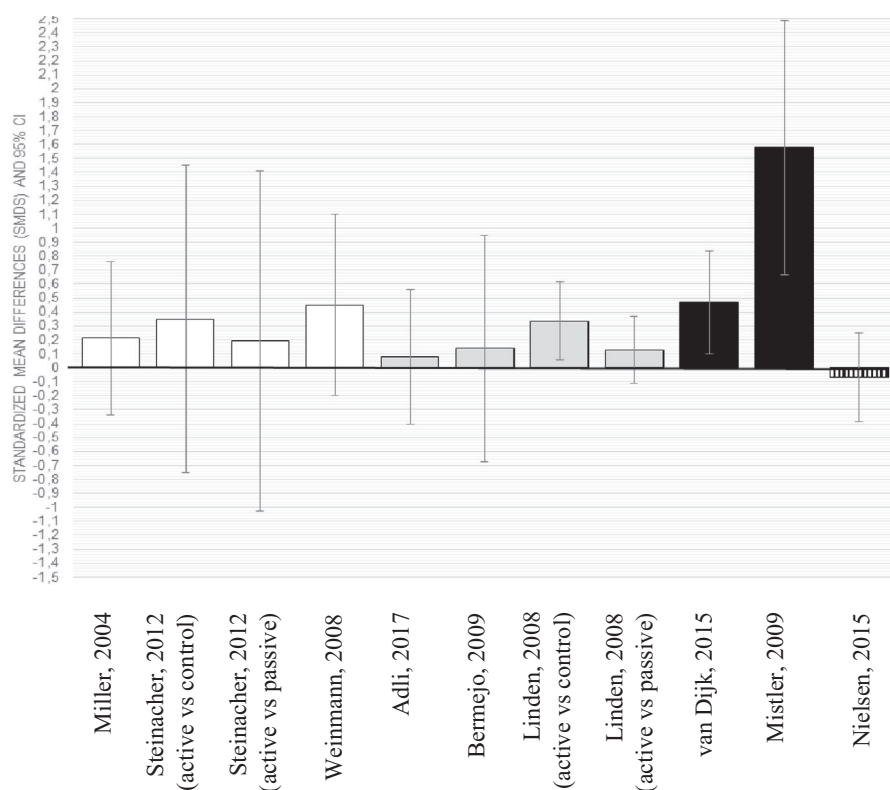


FIGURE 3 Harvest plot of all included studies (*n* = 9) which presented outcomes at the provider level. Depicted are the effect sizes Cohen's *d* (standard mean differences with 95% confidence intervals). Colours of bars are categorised as follows: (i) white: schizophrenia, (ii) dark grey: bipolar, (iii) black: mixed disorders and (iv) vertical stripes: alcohol use disorder

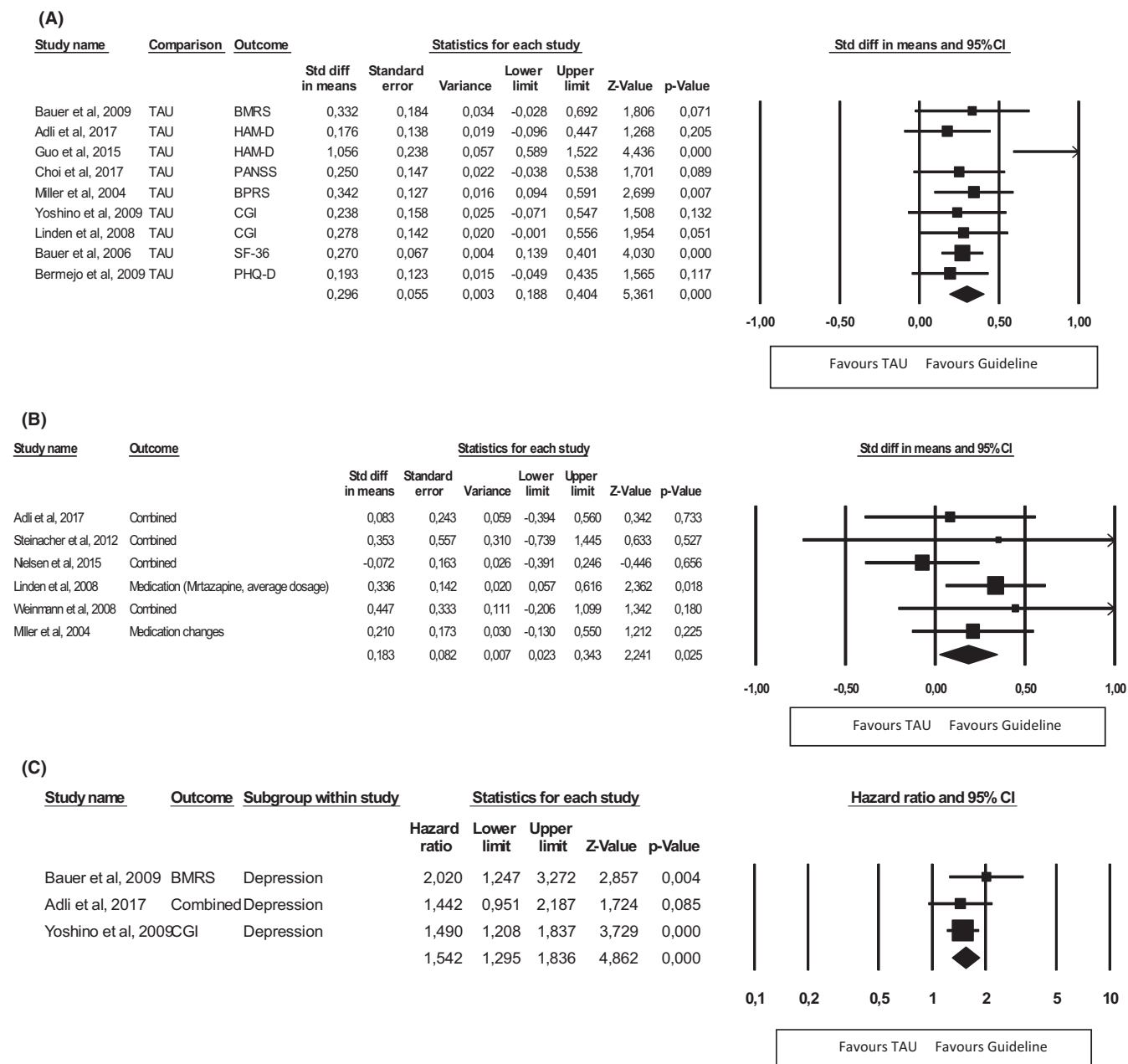


FIGURE 4 (A). Forest plot representing effect sizes Cohen's *d* (standard mean differences with 95% confidence intervals) comparing evidence-based guidelines vs TAU (*n* = 9) on severity of psychopathological symptoms at the patient level. (B). Forest plot representing effect sizes Cohen's *d* (standard mean differences with 95% confidence intervals) comparing evidence-based guidelines vs TAU (*n* = 6) on provider performance. (C). Forest plot of meta-analysis compared evidence-based guidelines versus TAU on hazard ratios

largely influence the effect size, adjustment for publication bias was therefore not needed.

3.4.2 | Provider level

Meta-analysis was performed with seven studies to quantify the impact of evidence-based guidelines vs TAU on provider performance (see Figure 4b for forest plot). Because heterogeneity was significant ($I^2 = 54\%$, $p = 0.04$), we could not

pool these data. The study by Mistler et al⁵¹ was identified as an outlier. After the removal of this outlier, heterogeneity decreased and became non-significant ($I^2 = 0\%$, $p = 0.474$). Without the Mistler et al study,⁵¹ the overall effect size was $d = 0.18$, 95%-CI = (0.02, 0.34), $p = 0.025$.

Both the funnel plot (Figure S1b) and Duvall and Tweedie's trim and fill procedure suggested the influence of publication bias among these six studies. Using trim and fill, the imputed point estimate under the random effects model was $d = 0.17$, 95%-CI = (0.01, 0.32). Because Egger's test

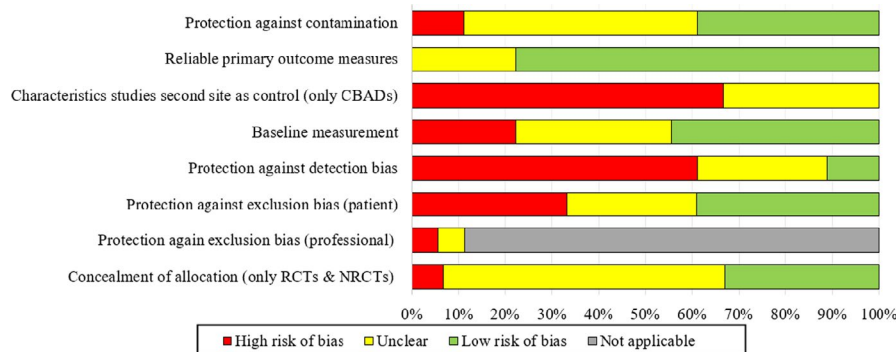


FIGURE 5 Summary of risk of bias assessment for RCT and NRCT studies according to EPOC checklist

appeared non-significant ($p = 0.736$), adjustment for publication bias was not needed as this did not greatly influence the overall effect size.

3.5 | Time to remission

Four studies presented time to remission as an outcome variable. These four studies showed significant levels of heterogeneity ($I^2 = 74\%$, $p = 0.01$). After the removal of one outlier,⁴⁷ time to remission appeared significant in favour of the intervention condition (hazard ratio = 1.54, 95%-CI = (1.29, 1.84), $p = 0.001$, $n = 3$) as can be shown in Figure 4c.

3.6 | Subgroup analyses

None of the exploratory subgroup analyses were significant, indicating that the type of mental disorder, the country in which the study was conducted or the type of control condition did not critically affect the primary outcome (patient health). Following are the specifics: type of mental disorder: depressive disorder ($n = 5$) vs schizophrenia ($n = 2$) (mean difference = 0.052, 95%-CI = (-0.42, 0.53), $p = 0.829$); country: United States ($n = 3$) vs Germany ($n = 4$) (mean difference = 0.05, 95%-CI = (-0.48, 0.58), $p = 0.854$); type of control condition: TAU ($n = 8$) vs passive implementation ($n = 3$) (mean difference = 0.10, 95%-CI = (-0.77, 0.98), $p = 0.821$).

3.7 | Adverse events

Adverse events were reported in six studies. One study⁴¹ reported that the number of deaths did not differ between intervention (12 deaths; 7%) and control conditions (8 deaths; 5%). Another study⁴³ reported one suicide in TAU vs no suicides in the intervention conditions. The study by Guo et al⁴⁷ also reported on adverse events but described these events as common side effects, including a dry mouth, loss of appetite and headache. The authors reported no significant differences

on these events between both conditions. Furthermore, Bauer et al⁴⁸ found a higher rate of adverse drug events in the intervention condition compared to TAU, leading to more dropouts in this group. A study by Weinmann et al¹⁶ reported a decrease in adverse drug events after the implementation of the guideline. Finally, a study by Linden et al³³ found no significant difference on adverse drug reactions between conditions.

3.8 | Dropouts

Dropout rates were reported in 13 of the 18 included studies. For the intervention condition, dropout rates ranged from 4.6% to 52.7%. For the control condition, a range of 11.4% to 58.9% was reported between studies. Based on the analyses ($n = 11$), we found a non-significant difference for dropout rates between the intervention condition and control condition ($p = 0.253$). Two studies were not included in this analysis because they reported dropout rates only for the total sample.^{41,46} Dropout rates for each included study are presented in Table S2.

3.9 | Methodological quality

Overall, the quality of the included studies fluctuated (Figure 5). For RCT and NRCT studies, concealment of allocation was often not described ($n = 9$, 60%). Protection against contamination was not clear in nine studies (60%). Given the nature of the included trials, blinding of patients and providers was often not possible, resulting in a high risk of bias on this domain (protection against detection bias) for 11 studies. Most studies did not perform follow-up assessments at the professionals' level (protection against exclusion bias). In CBAD studies ($n = 3$), characteristics of control sites were often not mentioned ($n = 2$). In one-third of the studies, follow-up measurements were obtained in at least 80% of the patients ($n = 7$). The majority used reliable primary outcome measures (90%). We conducted a meta-regression to analyse whether the magnitude Cohen's d systematically

co-varied with methodological quality. This was not the case ($b = -0.003$, $SD_b = 0.07$, $z = -0.05$, $p = 0.962$).

4 | DISCUSSION

This systematic review and meta-analysis examined the effectiveness of evidence-based guidelines at the patient level and found a statistically significant effect for these guidelines over TAU ($d = 0.29$, 95%-CI = (0.19, 0.40), $p < 0.001$). We tried to relate these patient-level outcomes to changes in the guideline adherence rates of the mental health professionals and found that providers in the guideline condition were significantly more adherent to guideline recommendations than providers delivering treatment in TAU ($d = 0.18$, 95%-CI = (0.02, 0.34), $p = 0.025$, $n = 6$), suggesting that more adherent provider behaviour may contribute to improved patient outcomes. However that may be, outcomes at both levels only correlated by 0.3, indicating that other factors may influence outcomes at the patient level as well. An important finding was that guideline-adherent treatment led to remission significantly earlier than TAU (HR = 1.54, 95%-CI = (1.29, 1.84)).

We could not confirm the hypothesis that the type of mental disorder affected the magnitude of the effect sizes. Additionally, the country in which the study was performed did not critically influence the outcome, nor did the type of control condition. The dropout rates of intervention vs control condition did not differ significantly, suggesting that the tolerability of the guidelines is not a major clinical issue. Adverse events were almost never reported in the studies.

5 | STRENGTHS AND WEAKNESSES

We could verify and extend the findings of the meta-analysis of Girlanda et al¹⁷ using patient-level outcomes as the primary outcome measure and including four additional studies, using a more conservative random effects model in the analyses, assessing publication bias and the quality of the studies included, evaluating dropout rates and reviewing adverse events. Compared to Girlanda et al,¹⁷ we were able to include additional studies measuring psychiatric patients' well-being in relation to evidence-based guidelines, and therefore, it was possible to extend the already-existing evidence from the literature.

A problem with the set-up of the trials is that the studies are very diverse in regard to the design, duration, control condition, measurement of patient outcomes and assessment of professional adherence to the guidelines. Since we included outcome trials with a variety of disorders with their

own disorder-specific rating scales, diversity in these ratings comes as no surprise. The measurement of the outcome of provider performance, however, could be improved by developing a standardised global evaluation method, not based solely on idiosyncratic process indicators derived from the guideline evaluated. We calculated effect sizes on a diverse mix of provider performance variables differing between studies on number and content. The choice of these provider performance variables may have affected the magnitude of the effect sizes derived. Because of this, we suggest that future researchers use uniformity in designing and reporting their studies since this would help to further develop the field of evidence-based mental health.

Comparing the quality judgements from two earlier published reviews,^{16,17} we may conclude that the quality of the studies improves gradually. The nature of studies evaluating guidelines holds that efficacy studies are virtually impossible to execute (see for example Feinstein 53). First, the duration of a guideline treatment in the case of non-response may last even more than one year, while efficacy studies usually have a short-term duration in order to avoid occurrences interfering with the effect of the treatment evaluated. Second, creating a sound control condition is difficult, especially when the guideline under study has been published by the authorities. In these circumstances, the only control condition possible may be 'passive' dissemination. In addition, the complex content of the guideline evaluated precludes simple but classic control conditions, such as 'pill-placebo', decreasing the possibility of demonstrating the superiority of the guideline. Third, the premise of the RCT may be compromised since neither patients, nor professionals nor research staff can be blinded. Moreover, sometimes it may be impossible to randomise because certain teams or institutions have implemented the guideline and can no longer serve as the control condition. Indeed, we had to include NRCTs as alternatives for RCTs in this meta-analysis. This implies that we must be satisfied with the less-than-optimal quality of this type of necessary evaluation research.

5.1 | Gaps in knowledge

We may conclude that delivering guideline-adherent treatments in mental health care yields superior outcomes at the patient level compared with non-adherent treatments. Up until now, this conclusion was drawn only in so-called modelling studies.^{54,55} From these studies, derived from 'real-life' epidemiological data combined with assumptions such as optimal coverage for every person and evidence-based treatments delivered to everyone, it appeared that delivering only guideline-adherent treatment in mental health care was cost-effective⁵⁴ and led to earlier remission.⁵⁵ Unfortunately,

the studies included in this review did not present economic data. Other sparse research suggests that guideline-based treatment may be cost-effective compared with TAU in patients with various mental disorders.^{48,56-58} Although these data are important and promising, we need further economic evaluations of RCTs comparing guideline-based treatment with TAU.

We found a small correlation (Spearman's $Rho = 0.32$) between patient improvement and provider adherence. Probably, provider-adherent care induces better patient outcomes. Nevertheless, this correlation was not 1.0, suggesting that other variables also influence outcomes at the patient level. It has previously been found that less adequate treatment is not only dependent on the adherence behaviour of the professional but also depends on factors related to the organisation of care (primary vs secondary care) and to patient factors (sex, income, urbanicity).² Another explanation for this rather low association between adherent provider behaviour and patient outcome may be that the process indicators with which provider behaviour is measured might be too 'distal' from outcome at the patient level to expect a higher correlation. Finally, measurement of the degree of satisfaction of the patient with the guideline-adherent treatment lacked in almost all studies. Although dropout rates, which did not differ critically between guideline treatment and control treatment, and to a lesser extent adverse events, do not suggest that the feasibility of the guideline treatments is low, a more explicit way of measuring patient satisfaction would serve in the development of the mental health service in general and specific modules in particular.⁵⁹

Finally, we do not have data on the mechanism of action of the effectiveness of guideline-adherent treatment. It has been previously hypothesised that the superiority of guidelines would not depend on the delivery of evidence-based treatments but rather on the structured method of repeated evaluations of response and tolerability,²⁰ sometimes termed 'measurement-based health care'.⁶⁰ The improvement of guideline-based care is only possible when we know which element(s) from the 'treatment package' is responsible for the superiority of guideline-adherent treatment.

6 | CONCLUSION AND IMPLICATIONS FOR CLINICAL PRACTICE

To elevate the proportion of patients in mental healthcare settings receiving evidence-based treatments, guideline development started almost 30 years ago. Since the uptake of guideline recommendations remained rather low, implementation strategies were developed in an effort to optimise the

quality of mental health care. We may conclude that the evaluation of the effectiveness of evidence-based guidelines is improving, but still finds itself in its infancy. The good news is that the quality of the evaluation studies has improved over the years. Furthermore, findings from our study are highly relevant to current clinical practice as they give a first indication that patients receiving guideline-adherent treatment improve to a larger extent and more quickly compared to patients treated with TAU. Disappointing is the rather large diversity of studies, making a quantitative summary of the findings only possible for a minority of the studies. Be that as it may, guideline-adherent treatment in mental health care is acceptable and tolerable for our patients and appears to be feasible to deliver for our professionals. Knowledge on the mechanisms of change during guideline-adherent treatment needs to be developed further so that we can provide patients with the best possible treatment in routine care, achieving better clinical outcomes.

ACKNOWLEDGEMENTS

We thank Caroline Planting of the VU University Medical Center library for her help with conducting the literature search.

CONFLICT OF INTEREST

All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf, they have no financial relationships with any organisations that might have an interest in the submitted work. No other relationships or activities that could appear to have influenced the submitted work.

AUTHOR CONTRIBUTIONS

KS, KB, RG and AvB have written the concept and design of the study. KS, KB and AvB assessed studies for eligibility and extracted the information from all articles. KS and AH conducted the statistical analyses and interpreted the data. AvB supervised the study and reviewed and edited the main manuscript text. All authors revised the text critically and all approved the version to be published.

PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1111/acps.13332>.

DATA AVAILABILITY STATEMENT

All records and data of this study are saved in a separate database. If researchers are interested in the data, they can contact us by using the contact information of the corresponding author (KS).

ORCID

Kim Setkowski  <https://orcid.org/0000-0002-3122-5451>

REFERENCES

1. Fernandez A, Haro JM, Martinez-Alonso M, et al. Treatment adequacy for anxiety and depressive disorders in six European countries. *Br J Psychiatry*. 2007;190:172-173.
2. Wang PS, Lane M, Olfson M, Pincus HA, Wells KB, Kessler RC. Twelve-month use of mental health services in the United States: results from the National Comorbidity Survey Replication. *Arch Gen Psychiatry*. 2005;62: 629-640.
3. Young AS, Klap R, Sherbourne CD, Wells KB. The quality of care for depressive and anxiety disorders in the United States. *Arch Gen Psychiatry*. 2001;58: 55-61.
4. Barth JH, Misra S, Aakre KM, et al. Why are clinical practice guidelines not followed? *Clin Chem Lab Med*. 2016;54: 1133-1139.
5. Burgers J, van der Weijden T, Grol R. Richtlijnen als hulpmiddel bij de verbetering van zorg. In: Grol R, Wensing M, eds. *Implementatie*, 7th edn. Houten: Bohn Stafleu van Loghum; 2017.
6. Grimshaw JM, Thomas RE, Maclennan G, et al. Effectiveness and efficiency of guideline dissemination and implementation strategies. *Health Technol Assess*. 2004;8: iii-iv, 1-72.
7. Grol R, Grimshaw J. From best evidence to best practice: effective implementation of change in patients' care. *Lancet*. 2003;362: 1225-1230.
8. Cochrane LJ, Olson CA, Murray S, Dupuis M, Tooman T, Hayes S. Gaps between knowing and doing: understanding and assessing the barriers to optimal health care. *J Contin Educ Health Prof*. 2007;27: 94-102.
9. Mokkenstorm J, Franx G, Gilissen R, Kerkhof A, Smit JH. Suicide prevention guideline implementation in specialist mental healthcare institutions in The Netherlands. *Int J Environ Res Public Health*. 2018;15: 1-12.
10. Setkowski K, Gilissen R, Franx G, Mokkenstorm JK, Van Den Ouwelant A, Van Balkom A. Practice variation in the field of suicide prevention in Dutch mental healthcare. *Tijdschr Psychiatr*. 2020;62: 439-447.
11. Gagliardi AR, Brouwers MC, Palda VA, Lemieux-Charles L, Grimshaw JM. How can we improve guideline use? A conceptual framework of implementability. *Implement Sci*. 2011;6: 1-11.
12. Streiner DL. The 2 "Es" of research: efficacy and effectiveness trials. *Can J Psychiatry*. 2002;47:552-556.
13. Higgins JPT, Green S. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.2* (updated September 2009). The Cochrane Collaboration; 2008. <https://www.cochrane-handbook.org>. Accessed 6 May 2020.
14. Singal AG, Higgins PD, Waljee AK. A primer on effectiveness and efficacy trials. *Clin Transl Gastroenterol*. 2014;5: e45.
15. Bauer MS. A review of quantitative studies of adherence to mental health clinical practice guidelines. *Harv Rev Psychiatry*. 2002;10:138-153.
16. Weinmann S, Koesters M, Becker T. Effects of implementation of psychiatric guidelines on provider performance and patient outcome: systematic review. *Acta Psychiatr Scand*. 2007;115:420-433.
17. Giralanda F, Fiedler I, Becker T, Barbui C, Koesters M. The evidence-practice gap in specialist mental healthcare: systematic review and meta-analysis of guideline implementation studies. *Br J Psychiatry*. 2017;210:24-30.
18. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Medicine*. 2009;6: e1000097.
19. Field MJ, Lohr KN. *Clinical Practice Guidelines: Directions for a New Program*. Washington DC (USA): National Academies Press; 1990.
20. Adli M, Bauer M, Rush AJ. Algorithms and collaborative-care systems for depression: are they effective and why? A systematic review. *Biol Psychiatry*. 2006;59:1029-1038.
21. Cochrane Effective Practice and Organisation of Care Review (EPOC). Data collection form. EPOC Resources for review authors. 2017. <https://epoccochrane.org/resources/epoc-specific-resources-review-authors>. Accessed 3 February 2020.
22. Cochrane Effective Practice and Organisation of Care Review (EPOC). Data collection checklist. Ottawa, Canada: Cochrane Effective Practice and Organisation of Care Review Group; 2002.
23. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. *Introduction to Meta-Analysis*. United Kingdom: John Wiley & Sons Ltd; 2009.
24. Rosenthal R, DiMatteo MR. Meta-Analysis: recent developments in quantitative methods for literature reviews. *Annu Rev Psychol*. 2001;52:59-82.
25. Ellis PD. *The Essential Guide to Effect Sizes: An Introduction to Statistical Power, Meta-Analysis and the Interpretation of Research Results*. Cambridge: Cambridge University Press; 2010.
26. Morriss R. Implementing clinical guidelines for bipolar disorder. *Psychol Psychother*. 2008;81:437-458.
27. Higgins JPT, Green S. *Cochrane Handbook for Systematic Reviews of Intervention Version 5.1.0* [updated March 2011]. The Cochrane Collaboration. 2011. <http://www.handbook.cochrane.org>. Accessed 28 July 2020.
28. Higgins JPT, Thomas J, Chandler J, et al. *Cochrane Handbook for Systematic Reviews of Interventions version 6.0* (updated August 2019). The Cochrane Collaboration; 2019. <https://www.training.cochrane.org/handbook>. Accessed 28 July 2020.
29. Chinn S. A simple method for converting an odds ratio to effect size for use in meta-analysis. *Stat Med*. 2000;19: 3127-3131.
30. Janssen B, Ludwig S, Eustermann H, et al. Improving outpatient treatment in schizophrenia: effects of computerized guideline implementation—results of a multicenter-study within the German research network on schizophrenia. *Eur Arch Psychiatry Clin Neurosci*. 2010;260: 51-57.
31. Steinacher B, Mausolff L, Gusy B. The effects of a clinical care pathway for schizophrenia: a before and after study in 114 patients. *Dtsch Arztebl Int*. 2012;109:788-794.
32. Cohen J. *Statistical power analysis for the behavioral sciences*. New York: Academic Press; 1969.
33. Linden M, Westram A, Schmidt LG, Haag C. Impact of the WHO depression guideline on patient care by psychiatrists: a randomized controlled trial. *Eur Psychiatry*. 2008;23:403-408.
34. Williams DK, Thrush CR, Armitage TL, Owen RR, Hudson TJ, Thapa P. The effect of guideline implementation strategies on akathisia outcomes in Schizophrenia. *The Journal of Applied Research*. 2003;3: 470-482.
35. Lee CH, Cook S, Lee JS, Han B. Comparison of two meta-analysis methods: inverse-variance-weighted average and weighted sum of Z-scores. *Genomics Inform*. 2016;14: 173-180.
36. Ogilvie D, Fayer D, Petticrew M, et al. The harvest plot: a method for synthesising evidence about the differential effects of interventions. *BMC Med Res Methodol*. 2008;8: 1-7.
37. Crowther M, Avenell A, Maclennan G, Mowatt G. A further use for the Harvest plot: a novel method for the presentation of data synthesis. *Res Synth Methods*. 2011;2: 79-83.

38. Cuijpers P. Meta-analyses in mental health research: A practical guide. Amsterdam: Vrije Universiteit Amsterdam; 2016.
39. Choi J, Lysaker PH, Bell MD, et al. Decisional informatics for psychosocial rehabilitation: a feasibility pilot on tailored and fluid treatment algorithms for serious mental illness. *J Nerv Ment Dis.* 2017;205: 867-872.
40. Miller AL, Crismon ML, Rush AJ, et al. The Texas medication algorithm project: clinical results for schizophrenia. *Schizophr Bull.* 2004;30: 627-647.
41. Bauer MS, McBride L, Williford WO, et al. Collaborative care for bipolar disorder: Part II. Impact on clinical outcome, function, and costs. *Psychiatr Serv.* 2006;57: 937-945.
42. Suppes T, Rush AJ, Dennehy EB, et al. Texas Medication Algorithm Project, phase 3 (TMAP-3): clinical results for patients with a history of mania. *J Clin Psychiatry.* 2003;64: 370-382.
43. 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: 721-730.
44. Bermejo I, Schneider F, Kriston L, et al. Improving outpatient care of depression by implementing practice guidelines: a controlled clinical trial. *Int J Qual Health Care.* 2009;21: 29-36.
45. Yoshino A, Sawamura T, Kobayashi N, Kurauchi S, Matsumoto A, Nomura S. Algorithm-guided treatment versus treatment as usual for major depression. *Psychiatry Clin Neurosci.* 2009;63: 652-657.
46. 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: 669-680.
47. Guo T, Xiang Y-T, Xiao LE, et al. Measurement-based care versus standard care for major depression: a randomized controlled trial with blind raters. *Am J Psychiatry.* 2015;172: 1004-1013.
48. 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: 327-333.
49. Nielsen AS, Nielsen B. Implementation of a clinical pathway may improve alcohol treatment outcome. *Addict Sci Clin Pract.* 2015;10: 1-7.
50. van Dijk MK, Oosterbaan DB, Verbraak MJ, Hoogendoorn AW, Penninx BW, van Balkom AJ. Effectiveness of the implementation of guidelines for anxiety disorders in specialized mental health care. *Acta Psychiatr Scand.* 2015;132: 69-80.
51. Mistler LA, Mellman TA, Drake RE. A pilot study testing a medication algorithm to reduce polypharmacy. *Qual Saf Health Care.* 2009;18: 55-58.
52. de Beurs DP, de Groot MH, de Keijser J, van Duijn E, de Winter RF, Kerkhof AJ. Evaluation of benefit to patients of training mental health professionals in suicide guidelines: cluster randomised trial. *Br J Psychiatry.* 2016;208: 477-483.
53. Feinstein AR. Additional basic approaches in clinical research. *Clin Res.* 1985;33:111-114.
54. Andrews G, Issakidis C, Sanderson K, Corry J, Lapsley H. Utilising survey data to inform public policy: comparison of the cost-effectiveness of treatment of ten mental disorders. *Br J Psychiatry.* 2004;184:526-533.
55. Meeuwissen JAC, Feenstra TL, Smit F, et al. The cost-utility of stepped-care algorithms according to depression guideline recommendations - Results of a state-transition model analysis. *J Affect Disord.* 2019;242: 244-254.
56. Bauer M, Rush A, Ricken R, Pilhatsch M, Adli M. Algorithms for treatment of major depressive disorder: efficacy and cost-effectiveness. *Pharmacopsychiatry.* 2019;52: 117-125.
57. Kashner TM, Rush AJ, Crismon ML, et al. An empirical analysis of cost outcomes of the Texas Medication Algorithm Project. *Psychiatr Serv.* 2006;57: 648-659.
58. Ricken R, Wiethoff K, Reinhold T, et al. Algorithm-guided treatment of depression reduces treatment costs—results from the randomized controlled German Algorithm Project (GAPII). *J Affect Disord.* 2011;134: 249-256.
59. van Dijk MK, Oosterbaan DB, Verbraak MJ, van Balkom AJ. The effectiveness of adhering to clinical-practice guidelines for anxiety disorders in secondary mental health care: the results of a cohort study in the Netherlands. *J Eval Clin Pract.* 2013;19: 791-797.
60. 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: 28-40.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Setkowski K, Boogert K, Hoogendoorn AW, Gilissen R, van Balkom AJLM. Guidelines improve patient outcomes in specialised mental health care: A systematic review and meta-analysis. *Acta Psychiatr Scand.* 2021;144:246–258. <https://doi.org/10.1111/acps.13332>