

A National Quality Improvement Collaborative for the clinical use of outcome measurement in specialised mental healthcare: results from a parallel group design and a nested cluster randomised controlled trial

Margot J. Metz, Marjolein A. Veerbeek, Gerdien C. Franx, Christina M. van der Feltz-Cornelis, Edwin de Beurs and Aartjan T. F. Beekman

Background

Although the importance and advantages of measurement-based care in mental healthcare are well established, implementation in daily practice is complex and far from optimal.

Aims

To accelerate the implementation of outcome measurement in routine clinical practice, a government-sponsored National Quality Improvement Collaborative was initiated in Dutch-specialised mental healthcare.

Method

To investigate the effects of this initiative, we combined a matched-pair parallel group design (21 teams) with a cluster randomised controlled trial (RCT) (6 teams). At the beginning and end, the primary outcome ‘actual use and perceived clinical utility of outcome measurement’ was assessed.

Results

In both designs, intervention teams demonstrated a significant higher level of implementation of outcome measurement than control teams. Overall effects were large (parallel group $d=0.99$; RCT $d=1.25$).

Conclusions

The National Collaborative successfully improved the use of outcome measurement in routine clinical practice.

Declaration of interest

None.

Copyright and usage

© The Royal College of Psychiatrists 2017. This is an open access article distributed under the terms of the Creative Commons Non-Commercial, No Derivatives (CC BY-NC-ND) license.

Measurement-based care (MBC)^{1,2} has beneficial effects on achieving response and remission of mental health disorders, such as depression.^{1–6} In addition, MBC can enhance effective communication between patients and clinicians and involvement of patients in clinical decision-making.^{1,5,7–9} Despite these promising prospects of MBC, the progress in the application of outcome measurement in routine mental healthcare is slow,^{10,11} because of the complexity of its implementation.^{12–16}

To promote outcome measurement in routine clinical practice in Dutch-specialised mental healthcare, a government-sponsored National Quality Improvement Collaborative (QIC) was initiated.^{17–20} This National Collaborative gives the unique opportunity to investigate the actual use of outcome measurement in clinical practice and assess the perceived utility of this so-called routine outcome monitoring (ROM).^{5,9,14,21} The results of this evaluation study, conducted within this National Collaborative, are presented in this paper.

Method

Study design

This evaluation study, conducted within the National ROM QIC, aimed at accelerating the implementation of ROM in clinical practice (for details see ‘Intervention’). The study included a parallel group design with matched pairs of participating teams, in which a cluster randomised controlled trial (RCT) was embedded (Fig. 1). In both groups, we investigated the primary outcome: the actual use of ROM in clinical practice and the perceived clinical

utility of outcome measurement. In addition, we tested whether there were differences among three groups of clinicians (physicians, psychologists and nurses).

The participating specialised mental healthcare providers were each requested to enrol two similar teams. In total, 21 intervention teams across the country participated in the Collaborative and survey. Fourteen of them had a matched control team from the same provider, treating the same patient group (age, diagnosis and setting) in the same geographical catchment area. Of the 14 matched pairs, 6 pairs were randomly and 8 were non-randomly assigned to either the intervention or the control condition. The randomisation of six matched pairs was conducted by an independent data manager²⁰ (Dutch Trial Register, NTR5262) (Fig. 1). The 14 control teams conducted ROM ‘as usual’ and implemented the best practice, only after the end of the study.

In the teams not participating in the RCT, the participating mental health organisations were allowed to choose which of their two parallel teams was assigned to the experimental arm of the study and which was assigned to the control condition. Both teams were treating similar patient groups in the same geographical catchment area, just as the matched pairs of the randomised teams. In this paper, we present the results of the parallel group design and the nested RCT.

The teams consisted of three groups of clinicians: physicians, psychologists and nurses. The exact multidisciplinary composition depended on the patient group to be treated (i.e. nurses typically work in chronic care and psychologists in short-term curative out-patient treatment). For the study, no patient involvement was required; thus, no informed consent was needed.

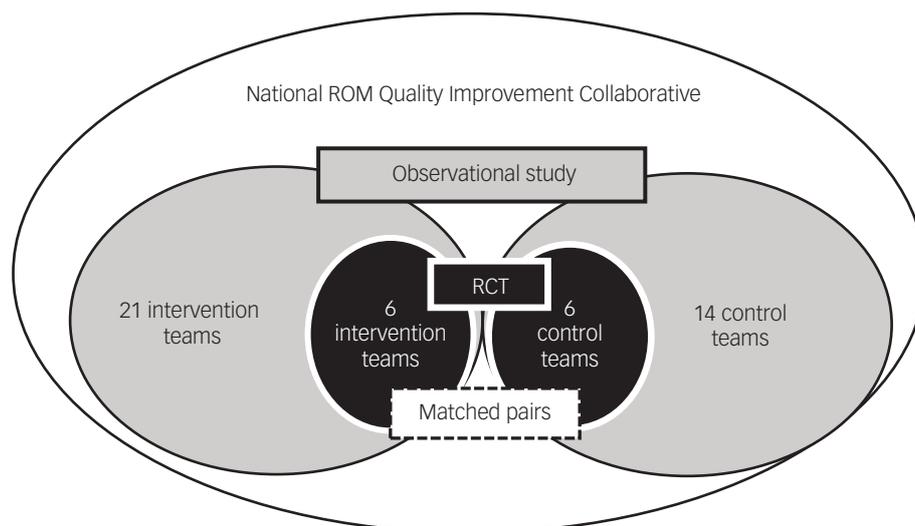


Fig. 1 Parallel group design with nested randomised controlled trial (RCT). ROM, routine outcome monitoring.

Intervention: National ROM QIC

The Collaborative promoted the routine use of clinical outcome questionnaires or rating scales at the beginning, during and at the end of the treatment. Clinicians were asked to discuss the ROM results with their patients to guide treatment decisions jointly. To help implement this ROM practice, the participating teams followed a National QIC programme of 1-year duration. A QIC is a multifaceted implementation strategy.^{17–19} It comprised a mix of improvement methods, applied both nationally and locally (in the teams). Conference days, training and booster sessions for exchange and learning, with experts and patient representatives present, were important national components of the improvement strategy. Moreover, the local teams, with involvement of patient representatives and supported by their management, determined their own improvement plans, specified in goals, actions and indicators. The multidisciplinary local teams organised meetings at their own location to work on their improvement plans. The teams planned, implemented, evaluated and adjusted their plans to improve the application of ROM in clinical practice in Plan-Do-Check-Act cycles.^{19,22,23} After the Collaborative's ending, the control teams are all offered the intervention.

Measurements: primary outcome

The primary outcome, the actual use and perceived clinical utility of ROM in clinical practice, was assessed by a survey²⁴ for clinicians at two moments: at the beginning (T_0) and at the end (T_1) of the QIC (after 1 year). Data collection took place independent of the Collaborative, by a data management team. Clinicians were invited and received a reminder by email to fill out the survey. The results were processed anonymously, and respondents were only labelled by team.

The survey had previously been developed by the Trimbos Institute, Netherlands Institute of Mental Health and Addiction.²⁴ Commissioned by the Ministry of Health, Welfare and Sport, the survey aims to identify the degree of implementation of ROM. The items of the survey were based on a systematic literature search of studies into influencing factors to ROM implementation and on expert meetings that assessed and rated the relevance of the identified factors. After a pilot test among clinicians, this development process resulted in a survey with 22 statements measuring the use of ROM in clinical practice from the perspective of the clinician. All statements had five response categories, ranging from 'strongly disagree' (score 1) to 'strongly agree' (score 5). A higher score meant a better implementation and use of ROM in clinical

practice.²⁴ Exploratory factor analysis demonstrated a four-factor structure of the instrument:

- Individual use and perceived utility of ROM in daily practice, consisting of eight items, for example 'I use the ROM scores to evaluate the course of treatment'
- Use of ROM in the team and organisational preconditions (seven items), for example 'ROM scores are used in multi-disciplinary consultations'
- Usefulness of the ROM questionnaires (four items), for example 'The questionnaires are suitable for measuring change'
- Accessibility of ROM for patient and clinician (three items), for example 'The output of ROM is simple and attractive'

In addition, a total scale score is calculated by summing all the items. The internal consistency of the total scale and the domain 'Individual use and perceived utility of ROM in daily practice' is very good, respectively, $\alpha=0.93$ and $\alpha=0.91$. The Cronbach's alphas of the two other domains are good: 'Use of ROM in the team and organisational preconditions' $\alpha=0.86$ and 'Usefulness of the ROM questionnaires' $\alpha=0.86$. The internal consistency of the domain 'Accessibility ROM for patient and clinician' is less adequate ($\alpha=0.51$). However, this scale was maintained in the survey, first because of the importance of the content of these items. According to implementation literature^{1,5,12,14,16,21} and experiences in the intervention teams, the accessibility of ROM results for patients and clinicians is an important precondition in using ROM in clinical practice (i.e. giving feedback on outcome data to patients and clinicians, communicating about the results, validating and using the information for (changes in) treatment plans). Second, a Cronbach's alpha >0.5 is deemed just acceptable, with a minimum of three items contributing to the domain.²⁵

Statistical analysis

Analysis was performed on the four subdomains of the survey and the total scale score. Data were analysed by SPSS for Windows, version 22. First, the number of teams, the number of drop-outs of the study, response to the survey and the composition of teams who responded to the survey, were described. Chi-squared tests were used to test potential differences in the composition of teams between the intervention and control groups. To calculate differences between T_0 and T_1 and the difference at T_1 between the intervention and control groups, independent sample *t*-tests were used, because

clinicians of the participating teams, who filled out the survey at T_0 and T_1 , were not always the same. Mean, standard deviation, confidence intervals and effect sizes were computed. The effect sizes were calculated by the following formula: $M_{\text{post}} - M_{\text{pre}} / SD_{\text{pooled}}$ (because of independent groups).

$SD_{\text{pooled}} = \sqrt{((SD1^2 + SD2^2) / 2)}$ using the effect size calculator for separate groups of L. Becker, University of Colorado (www.uccs.edu/lbecker/index.html). The thresholds for interpreting the effect size were small (0.00–0.32), medium (0.33–0.55) and large (0.56–1.20).²⁶ We repeated the analysis described above for the randomised teams (the nested RCT). Finally, in the intervention group of the parallel group design we looked at the differences between three main groups of clinicians (physicians, psychologists and nurses). Independent sample *t*-tests were used to calculate differences between T_0 and T_1 for each group of clinicians separately. Differences between the groups of clinicians on T_0 and T_1 were tested with analysis of variance (ANOVA) and *post hoc* tests (Bonferroni).

Power calculation

This study was designed to detect, in the intervention teams of the parallel group design, a medium effect size of $d=0.5$ on the primary outcome ‘actual use and the perceived clinical utility of ROM in clinical practice’ comparing T_1 with T_0 . With $\alpha=0.05$ and a power $\beta=0.80$, the required sample size was 65 clinicians in the intervention group.²⁷

Results

In each paragraph, the results are first described for the total parallel group design and next for the nested randomised design. Putative differences in effects among types of clinicians are shown for the parallel group design.

Participants

Parallel group design

Twenty-one teams from organisations of specialised mental healthcare across the country participated (see Fig. 2, flowchart 2a). In 14 of them, two similar teams were included. Flowchart 2a shows that, during the Collaborative, three teams dropped out between T_0 and T_1 , mainly because of reorganisations and personnel changes in the participating teams.

At T_0 , 69% of the clinicians in the intervention group and 75% in the control group responded to the survey. The types of clinicians responding to the survey in terms of profession were 11% physicians, 53% psychologists and 36% nurses in the intervention group, and 21% physicians, 43% psychologists and 36% nurses in the control group. The composition between intervention and control teams did not differ significantly.

At T_1 , 89% of the clinicians in the intervention group and 62% in the control group responded to the survey. The composition of the clinicians responding to the survey was 25% physicians, 44% psychologists and 31% nurses in the intervention group, and 17% physicians, 40% psychologists and 43% nurses in the control group. As with T_0 , the differences in composition at T_1 between intervention and control groups were not significant.

Cluster randomised control design

In Fig. 2, flowchart 2b shows loss of data over time in the randomised teams. In total, clinicians of six intervention teams and six control teams filled out the survey. Between T_0 and T_1 , one team dropped out because of reorganisation and personnel changes. At T_0 , 73% of the clinicians in the intervention group and 83% in the control group responded to the survey. The composition of the group of clinicians responding to the survey in terms of profession was 0% physicians, 65% psychologists and 35% nurses in the intervention teams, and 13% physicians, 67% psychologists and 20% nurses in the control teams.

At T_1 , 73% of the clinicians in the intervention group and 75% in the control group responded to the survey. At T_1 , the composition of these groups of clinicians responding to the survey was 9% physicians, 58% psychologists and 33% nurses in the intervention group, and 0% physicians, 54% psychologists and 46% nurses in the control group. Both at T_0 and T_1 , there were no significant differences in the composition of clinicians between intervention and control groups.

Results of the survey

To demonstrate the changes in the actual use and perceived clinical utility of ROM in the teams which participated in the Collaborative, first the difference between first (T_0) and final measurements (T_1) of the intervention group is described. Second, we looked at the differences between intervention and control groups at the end of the Collaborative (T_1). The results are demonstrated for both the parallel groups as the nested randomised design.

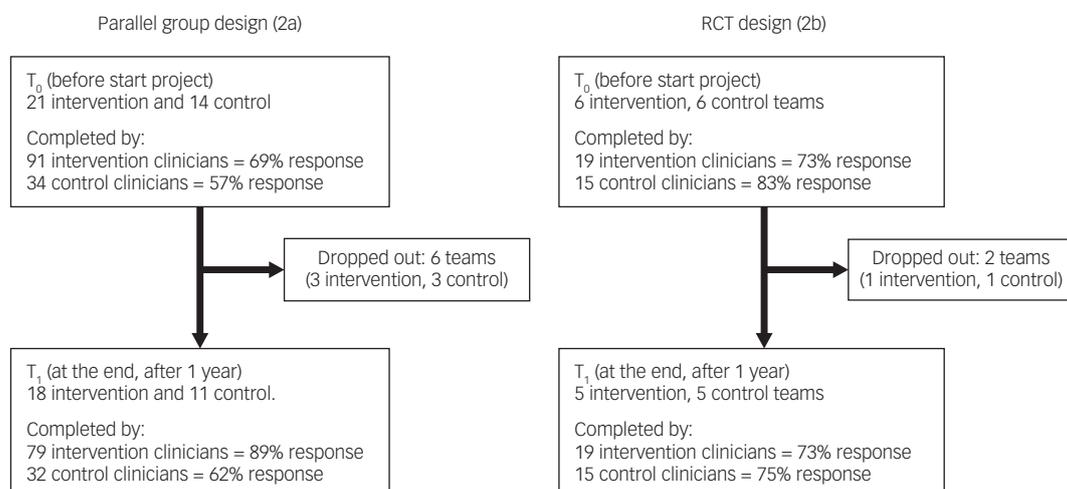


Fig. 2 Flow chart parallel group design (flowchart 2a) and randomised controlled trial (RCT) design (flowchart 2b).

Table 1 Changes in the intervention teams: T_1 compared with T_0 in parallel group design and nested RCT

| Survey domains | Parallel group design intervention teams | | | | | | | Cluster randomised control trial intervention teams | | | | | | | |
|---|--|------|------|-------------|------------------|--------------------------|-------|---|------|------|-------------|------------------|--------------------------|-------|-------|
| | N | Mean | s.d. | Effect size | Sig. t -tailed | 95% CI of the difference | | N | Mean | s.d. | Effect size | Sig. t -tailed | 95% CI of the difference | | |
| | | | | | | Lower | Upper | | | | | | Lower | Upper | |
| Individual use and perceived utility of ROM in daily practice | T_0 | 91 | 3.28 | 1.01 | 0.62 | 0.000 | -0.84 | -0.29 | 19 | 3.22 | 1.07 | 1.11 | 0.002 | -1.47 | -0.36 |
| | T_1 | 79 | 3.84 | 0.80 | | | | | 19 | 4.14 | 0.47 | | | | |
| Use of ROM in the team and organisational preconditions | T_0 | 91 | 2.59 | 0.85 | 1.02 | 0.000 | -1.05 | -0.57 | 19 | 2.59 | 0.80 | 1.14 | 0.001 | -1.36 | -0.37 |
| | T_1 | 79 | 3.40 | 0.74 | | | | | 19 | 3.45 | 0.71 | | | | |
| Usefulness of the ROM questionnaires | T_0 | 91 | 2.95 | 0.85 | 0.55 | 0.000 | -0.77 | -0.22 | 19 | 3.07 | 0.96 | 0.97 | 0.005 | -1.35 | -0.26 |
| | T_1 | 79 | 3.45 | 0.95 | | | | | 19 | 3.87 | 0.68 | | | | |
| Accessibility ROM for patient and clinician | T_0 | 91 | 2.95 | 0.72 | 0.88 | 0.000 | -0.86 | -0.42 | 19 | 2.96 | 0.97 | 1.07 | 0.002 | -1.39 | -0.33 |
| | T_1 | 79 | 3.59 | 0.74 | | | | | 19 | 3.82 | 0.59 | | | | |
| Total score of the ROM in daily practice | T_0 | 91 | 2.94 | 0.64 | 0.99 | 0.000 | -0.82 | -0.43 | 19 | 2.96 | 0.83 | 1.25 | 0.000 | -1.31 | -0.41 |
| | T_1 | 79 | 3.57 | 0.63 | | | | | 19 | 3.82 | 0.51 | | | | |

CI, confidence interval; RCT, randomised controlled trial; ROM, routine outcome monitoring; Sig., significance.

Differences between first and final measurements of the intervention group

Parallel group design In the intervention group, significant positive differences were shown between T_0 and T_1 on the total scale and all subscales of the survey 'ROM in daily practice' with medium to large effect sizes (between 0.55 and 1.02, with an effect size of 0.99 on the total scale) (Table 1). The control group showed no significant differences between T_0 and T_1 .

Cluster randomised control design The randomised group showed comparable results in the application of ROM in daily practice (Table 1). The effect sizes in the randomised intervention group were even larger (between 0.97 and 1.25, with an effect size of 1.25 on the total scale) than in the intervention group of the parallel group design (Table 1). Also in this design, the control group showed no significant differences between first and final measurements.

Differences in final measurements between intervention and control group

Parallel group design When the differences in T_1 between the intervention and control groups were tested, the final measurement of the intervention group scored significantly higher than the control group (Table 2). This means that at the end of the improvement year, ROM in daily practice is better implemented and used in clinical practice by respondents in the intervention group compared with respondents in the control group.

Cluster randomised control design While comparing the final measurements (T_1), the above-mentioned positive significant results in favour of the intervention teams were also shown in the RCT (Table 2).

Differences between clinicians

When comparing the first and final measurements in the intervention group of the parallel group design (Table 3), nurses and psychologists in the intervention group demonstrated at T_1 a significantly higher score on all the survey domains with large effect sizes (nurses between 0.68 and 1.28; psychologists between 0.57 and 1.17). Physicians in the intervention group scored at T_1 , compared with T_0 , significantly higher on the total score and the subdomain 'Use of ROM in the team and organisational preconditions', with large effect sizes on these scales (1.51 and 0.97). The three groups of clinicians participating in the control group showed no significant increase of T_1 relative to T_0 .

At T_0 , compared with the psychologists of the intervention group, nurses of this group showed a significantly lower score on the domain 'Accessibility ROM for patient and clinician' ($P=0.006$ and $CI=-1.020$ to -0.134). During the Collaborative year, the differences between these groups of clinicians were reduced. At T_1 , no significant differences were shown between the groups of clinicians in the intervention group.

Discussion

This paper presents the findings from the government-sponsored National QIC aimed to accelerate the implementation of ROM in

Table 2 Differences between intervention and control groups at T_1 in parallel group design and nested RCT

| Survey domains | Parallel group design T_1 intervention and control | | | | | | | Cluster randomised control trial T_1 intervention and control | | | | | | |
|---|--|---------|------|------------------|--------------------------|-------|------|---|------|------------------|--------------------------|-------|------|--|
| | N | Mean T2 | s.d. | Sig. t -tailed | 95% CI of the difference | | N | Mean T2 | s.d. | Sig. t -tailed | 95% CI of the difference | | | |
| | | | | | Lower | Upper | | | | | Lower | Upper | | |
| Individual use and perceived utility of ROM in daily practice | I | 79 | 3.84 | 0.80 | 0.000 | 0.52 | 1.20 | 19 | 4.14 | 0.47 | 0.000 | 0.74 | 1.72 | |
| | C | 32 | 2.98 | 0.87 | | | | 15 | 2.91 | 0.81 | | | | |
| Use of ROM in the team and organisational preconditions | I | 79 | 3.40 | 0.74 | 0.000 | 0.44 | 1.08 | 19 | 3.45 | 0.71 | 0.005 | 0.24 | 1.25 | |
| | C | 32 | 2.64 | 0.86 | | | | 15 | 2.70 | 0.74 | | | | |
| Usefulness of the ROM questionnaires | I | 79 | 3.45 | 0.95 | 0.008 | 0.15 | 0.95 | 19 | 3.87 | 0.68 | 0.011 | 0.19 | 1.32 | |
| | C | 32 | 2.90 | 0.99 | | | | 15 | 3.12 | 0.94 | | | | |
| Accessibility ROM for patient and clinician | I | 79 | 3.59 | 0.74 | 0.000 | 0.28 | 0.94 | 19 | 3.82 | 0.59 | 0.001 | 0.45 | 1.55 | |
| | C | 32 | 2.98 | 0.92 | | | | 15 | 2.82 | 0.97 | | | | |
| Total score of the ROM in daily practice | I | 79 | 3.57 | 0.63 | 0.000 | 0.42 | 0.97 | 19 | 3.82 | 0.51 | 0.000 | 0.50 | 1.36 | |
| | C | 32 | 2.88 | 0.73 | | | | 15 | 2.89 | 0.72 | | | | |

C, control group; CI, confidence interval; I, intervention; RCT, randomised controlled trial; ROM, routine outcome monitoring; Sig., significance.

Table 3 Results T_1 compared with T_0 in the parallel group design for nurses, psychologists and physicians in the intervention group

| Survey domains | Nurses | | | | | | | | | | Psychologists | | | | | | | | | | Physicians | | | | | | | | | |
|---|----------|------|------|-------------|---------------|--------------------------|-------|----|------|------|---------------|---------------|--------------------------|-------|----|------|------|-------------|---------------|--------------------------|------------|--|--|--|--|--|--|--|--|--|
| | N | Mean | s.d. | Effect size | Sig. t-tailed | 95% CI of the difference | | N | Mean | s.d. | Effect size | Sig. t-tailed | 95% CI of the difference | | N | Mean | s.d. | Effect size | Sig. t-tailed | 95% CI of the difference | | | | | | | | | | |
| | | | | | | Lower | Upper | | | | | | Lower | Upper | | | | | | Lower | Upper | | | | | | | | | |
| Individual use and perceived utility of ROM in daily practice | T_0 26 | 3.00 | 1.23 | 0.98 | 0.001 | -1.52 | -0.40 | 39 | 3.50 | 0.85 | 0.57 | 0.023 | -0.85 | -0.07 | 8 | 2.89 | 1.13 | 0.200 | 0.200 | -1.47 | 0.33 | | | | | | | | | |
| | T_1 21 | 3.95 | 0.63 | | | | | 30 | 3.96 | 0.76 | | | | | 17 | 3.46 | 0.96 | | | | | | | | | | | | | |
| Use of ROM in the team and organisational preconditions | T_0 26 | 2.49 | 1.05 | 1.11 | 0.001 | -1.61 | -0.48 | 39 | 2.56 | 0.69 | 1.17 | 0.000 | -1.15 | -0.48 | 8 | 2.13 | 0.85 | 1.51 | 0.001 | -1.84 | -0.51 | | | | | | | | | |
| | T_1 21 | 3.54 | 0.82 | | | | | 30 | 3.37 | 0.70 | | | | | 17 | 3.30 | 0.70 | | | | | | | | | | | | | |
| Usefulness of the ROM questionnaires | T_0 26 | 2.90 | 0.79 | 0.68 | 0.024 | -1.11 | -0.08 | 39 | 3.00 | 0.87 | 0.59 | 0.018 | -0.94 | -0.09 | 8 | 2.63 | 1.11 | 0.148 | 0.148 | -1.70 | 0.27 | | | | | | | | | |
| | T_1 21 | 3.50 | 0.96 | | | | | 30 | 3.52 | 0.90 | | | | | 17 | 3.34 | 1.11 | | | | | | | | | | | | | |
| Accessibility ROM for patient and clinician | T_0 26 | 2.58 | 0.82 | 1.28 | 0.000 | -1.44 | -0.52 | 39 | 3.15 | 0.64 | 0.87 | 0.001 | -0.87 | -0.25 | 8 | 2.75 | 0.64 | 0.155 | 0.155 | -1.21 | 0.20 | | | | | | | | | |
| | T_1 21 | 3.56 | 0.70 | | | | | 30 | 3.71 | 0.64 | | | | | 17 | 3.25 | 0.86 | | | | | | | | | | | | | |
| Total score of the ROM in daily practice. | T_0 26 | 2.74 | 0.76 | 1.28 | 0.000 | -1.31 | -0.48 | 39 | 3.05 | 0.54 | 1.13 | 0.000 | -0.84 | -0.33 | 8 | 2.60 | 0.74 | 0.97 | 0.037 | -1.43 | -0.05 | | | | | | | | | |
| | T_1 21 | 3.64 | 0.64 | | | | | 30 | 3.64 | 0.49 | | | | | 17 | 3.34 | 0.80 | | | | | | | | | | | | | |

CI, confidence interval; ROM, routine outcome monitoring; Sig., significance.

Dutch-specialised mental healthcare. The study included a parallel group design with matched pairs of participating teams, in which a cluster RCT was nested. In both intervention and control teams, the actual use of ROM in routine clinical practice and the perceived clinical utility of outcome measurements were investigated at the beginning and end of the Collaborative.

In both the parallel group design and the nested RCT, the intervention teams reported much better results with respect to the actual use and the perceived clinical utility of ROM (Tables 1 and 2). In the parallel group design, which included 21 intervention teams across the country, the overall effect was large ($d=0.99$). Notably, the effect size in the nested RCT was even bigger ($d=1.25$) than in the study with the parallel groups. This is probably because of the more rigorous research design and implementation protocol that was used in the RCT. Considering putative differences among specific groups of clinicians, psychologists and nurses participating in the intervention group demonstrated a large improvement on both the overall scale and all the subdomains, measuring different aspects of ROM implementation. The physicians taking part in the study showed a similar large improvement in the overall scale. Looking at the specific subscales, their improvement was restricted to the domain ‘Use of ROM in the team and organisational preconditions’. This may be explained by the tasks physicians have in the teams, which are less focused on the execution of the ROM measures and more on the team supervision and organisation of care. Their assessments of the usefulness of ROM may have been more driven by the ROM-related activities they noticed in the team, represented by the subscale ‘Use of ROM in the team and organisational preconditions’. The other three subdomains showed practical and executive functions in the application of ROM. The baseline difference among psychologists and nurses on the subdomain ‘Accessibility ROM for patient and clinician’ might be related to the background of psychologists who are generally more inclined to use measurement instruments in daily practice. It is encouraging to see that this targeted intervention succeeded in reducing the difference between psychologists and nurses, implying that the intervention was successful in engaging nursing personnel in this area that is so important for their work.

Strengths

In this study, we had the unique opportunity to nest a rigorous experimental study (RCT) design within a government-sponsored national initiative to improve mental healthcare. We built on previous work in which the survey was developed.²⁴ The teams experienced ownership of their improvement process and were facilitated by the National QIC. A variety of teams with a multi-disciplinary composition of clinicians treating different patient groups (age, diagnoses and setting) participated in the study. Independent data collection took place by a data management team, which processed the results anonymously. Thus, the likelihood of socially desirable answers and influence of the research team on the results were diminished. To prevent possible influence of confounding, the results were shown for both the parallel group design and the nested cluster randomised design separately. Strength of the parallel group design was the large external validity because of the number and variation of the participating teams. The randomised group included fewer teams, but the risk of confounding was reduced, and in this design, we conducted a strict research and implementation protocol.

Limitations

The study also had some limitations which may have influenced the results. First, the clinicians were aware of the objective of the National Collaborative, which may have affected their answers on the survey. Second, there may have been cross-over effects of knowledge and experiences from the intervention to control group. Third, the survey could be seen as a process evaluation, focusing on the implementation of ROM seen by clinicians who participated in the Collaborative. To get insight in the experiences of patients and the effectiveness of the intervention at patient level, an additional study is underway, which will research the effects on decisional conflict of patients, working alliance, treatment adherence, clinical outcome and quality of life.²⁰ Finally, the follow-up is restricted, and it is unknown how the teams fared with ROM over a longer time. Given the large effect sizes between the final and first measurements and the attention that was given during the Collaborative to the continuity of the implementation afterwards, we expect the intervention teams will maintain the positive effects of the Collaborative. Nevertheless, it is still important to ensure that the teams continue the intervention by organising follow-up and booster sessions.

Given the above limitations, our overall conclusion is that the implementation of outcome measurement in clinical practice was highly successful and appreciated by the multidisciplinary teams that were involved. All the three groups of clinicians participating in the intervention group take advantage of the ROM implementation and showed, at the end of the Collaborative, an equal level in the actual use and the perceived utility of ROM in clinical practice. Successful in the ROM implementation is the bottom-up approach, in which multidisciplinary teams were facilitated to complete their own improvement cycle. This study is unique in that we combined a National Collaborative of Quality Improvement in mental healthcare with an evaluation study in two designs, a parallel group design and a nested RCT. The results have both internal (with regard to the rigorous design and implementation) and external (given the nationwide implementation and evaluation) validity. Given the established advantages of MBC and the difficulties previously encountered in implementing the use of ROM in routine care, these results are encouraging and call for more implementation efforts along these lines.

Margot J. Metz, MSc, GGZ Breburg, Tilburg, The Netherlands; Trimbos Institute, Utrecht, The Netherlands; VU University, Amsterdam, The Netherlands; **Marjolein A. Veerbeek**, PhD, Trimbos Institute, Utrecht, The Netherlands; **Gerdien C. Franx**, PhD, Trimbos Institute, Utrecht, The Netherlands; **Christina M. van der Feltz-Cornelis**, MD, PhD, Professor of Social Psychiatry, GGZ Breburg, Tilburg, The Netherlands; Tilburg University, Tilburg, The Netherlands; **Edwin de Beurs**, PhD, Professor of Routine Outcome Monitoring and Benchmarking, University of Leiden, Leiden, The Netherlands; Stichting Benchmark GGZ, Bilthoven, The Netherlands; **Aartjan T. F. Beekman**, MD, PhD, Professor of Psychiatry, GGZ inGeest, Amsterdam, The Netherlands; VU University Medical Centre, Amsterdam, The Netherlands

Correspondence: Margot J. Metz, GGZ Breburg, Postbus 770, 5000 AT Tilburg, The Netherlands. Email: mmetz@trimbos.nl

First received 29 Nov 16, final revision 21 Mar 2017, accepted 30 Mar 2017

Funding

The project is funded by the National Network for Quality Development in mental healthcare and conducted by the Trimbos Institute, Netherlands Institute of Mental Health and Addiction.

Acknowledgements

The authors would like to thank Mr P. van Splunteren MSc and Mrs H. Sinnema PhD for the project management of the National Collaborative. They also thank the data management team of Trimbos Institute for their support in the data collection. Finally, they thank the clinicians of the participating organisations for responding to the survey.

References

- Fortney JC, Unützer J, Wren G, Pyne JM, Smith GR, Schoenbaum M, et al. A tipping point for measurement based care. *Psychiatr Serv* 2016; **68**: 1–10.
- Trivedi MH, Rush AJ, Wisniewski SR, Nierenberg AA, Warden D, Ritz L, 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.
- Guo T, Xiang YT, Xiao L, Hu CQ, Chiu HFK, Ungvari GS, et al. Measurement-based care versus standard care for major depression: a randomized controlled trial with blind raters. *Am J Psychiatry* 2015; **172**: 1004–13.
- Knaup C, Koesters M, Schoefer D, Becker T, Puschner B. Effect of feedback of treatment outcome in specialist mental healthcare: meta-analysis. *Br J Psychiatry* 2009; **195**: 15–22.
- Carlier IVE, Meuldijk D, van Vliet IM, van Fenema EM, van der Wee NJA, Zitman FG. Routine outcome monitoring and feedback on physical or mental health status: evidence and theory. *J Eval Clin Pract* 2012; **18**: 104–10.
- Davidson K, Perry A, Bell L. Would continuous feedback of patient's clinical outcomes to practitioners improve NHS psychological therapy services? Critical analysis and assessment of quality of existing studies. *Psychol Psychother* 2015; **88**: 21–37.
- Valenstein M, Adler DA, Berlant J, Dixon LB, Dulit RA, Goldman B, et al. Implementing standardized assessments in clinical care: now's the time. *Psychiatr Serv* 2009; **60**: 1372–5.
- Eisen SV, Dickey B, Sederer LI. A self-report symptom and problem rating scale to increase inpatients' involvement in treatment. *Psychiatr Serv* 2000; **51**: 349–53.
- van der Feltz-Cornelis CM, Andrea H, Kessels E, Duivenvoorden H, Biemans H, Metz M. Does routine outcome monitoring have a promising future? An investigation into the use of shared decision-making combined with ROM for patients with a combination of physical and psychiatric symptoms [in Dutch]. *Tijdschr Psychiatr* 2014; **56**: 375–84.
- Van Der Wees PJ, Nijhuis-van der Sanden MWG, Ayanian JZ, Black N, Westert GP, Schneider EC. Integrating the use of patient-reported outcomes for both clinical practice and performance measurement: views of experts from 3 countries. *Milbank Q* 2014; **92**: 754–75.
- Delespaul PEG. Routine outcome measurement in the Netherlands – a focus on benchmarking. *Int Rev Psychiatry* 2015; **27**: 320–8.
- Boswell JF, Kraus DR, Miller SD, Lambert MJ. Implementing routine outcome monitoring in clinical practice: benefits, challenges and solutions. *Psychother Res* 2015; **1**: 6–19.
- de Jong K, van Sluis P, Nugter MA, Heiser WJ, Spinhoven P. Understanding the differential impact of outcome monitoring: therapist variables that moderate feedback effects in a randomised clinical trial. *Psychother Res* 2012; **22**: 464–74.
- de Jong K, Timman R, Hakkaart-van Royen L, Vermeulen P, Kooiman K, Passchier J, et al. The effect of outcome monitoring feedback to clinicians and patients in short and long-term psychotherapy: a randomised controlled trial. *Psychother Res* 2014; **24**: 629–39.
- de Jong K. Challenges in the implementation of measurement feedback systems. *Adm Policy Ment Health* 2016; **43**: 467–70.
- Duncan EAS, Murray J. The barriers and facilitators to routine outcome measurement by allied health professionals in practice: a systematic review. *BMC Health Serv Res* 2012; **12**: 96.
- Schouten LMT, Hulscher ME, van Everdingen JJ, Huijsman R, Grol RP. Evidence for the impact of Quality Improvement Collaboratives: systematic review. *BMJ* 2008; **336**: 1491–5.
- Franx GC. *Quality Improvement in Mental Health care: The Transfer of Knowledge into Practice*. Scientific Institute for Quality of Healthcare and Netherlands Institute of Mental Health and Addiction, 2012.
- Øvreveit J, Bate P, Cleary P, Cretin S, Gustafson D, McInnes K, et al. Quality collaboratives: lessons from research. *Qual Saf Health Care* 2002; **11**: 345–51.
- Metz MJ, Franx GC, Veerbeek MA, de Beurs E, Van der Feltz-Cornelis CM, Beekman AT. Shared decision making in mental health care using routine outcome monitoring as a source of information: a cluster randomised controlled trial. *BMC Psychiatry* 2015; **15**: 313–23.
- de Beurs E, den Hollander-Gijsman ME, van Rood YR, van der Wee NJ, Giltay EJ, van van Noorden MS, et al. Routine outcome monitoring in the Netherlands: practical experiences with a web-based strategy for the assessment of treatment outcome in clinical practice. *Clin Psychol Psychother* 2011; **18**: 1–12.
- Berwick DM. Developing and testing changes in delivery of care. *Ann Intern Med* 1998; **15**: 651–7.
- van Splunteren P, van Everdingen J, Janssen S, Minkman M, Rouppe van de Voort M, Schouten L, et al. *Breaking Through with Results: Improvement of*

- Patient Care using the Breakthrough Method* [in Dutch]. Koninklijke van Gorcum, 2003.
- 24** Nuijen J, Wijngaarden B, Veerbeek M, Franx G, Meeuwissen J, Bon van-Martens M. *Implementatie van ROM in de dagelijkse zorgpraktijk. Resultaten van enquêtes onder behandelaren van GGZ instellingen en vrijevestigde behandelaren*. Trimbo-institute, 2014.
- 25** Vet de HCW, Terwee CB, Mokkink LB, Knol DL. *Measurement in Medicine: A Practical Guide*: 65–95. Cambridge University Press, 2015.
- 26** Lipsey MW, Wilson DB. The efficacy of psychological, educational and behavioural treatment: confirmation from meta-analysis. *Am Psychol* 1993; **48**: 1181–209.
- 27** Lipsey MW. *Design Sensitivity: Statistical Power for Experimental Research*: 137. Sage, 1990.



OPEN
ACCESS

