Chapter 4

A National Quality Improvement Collaborative for the clinical use of outcome measurement in specialized mental health care: results from a parallel group design and a nested cluster randomized controlled trial

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

Abstract

Background
Although the importance and advantages of measurement based care in mental health care 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 specialized mental health care.

Method
To investigate the effects of this initiative, we combined a matched pair parallel group design (21 teams) with a cluster randomized controlled trial (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.
Introduction

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 health care is slow,\(^10\,\,^11\) due to the complexity of its implementation.\(^12\,\,^16\)

In order to promote outcome measurement in routine clinical practice in Dutch specialized mental health care, a government sponsored National Quality Improvement Collaborative 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 Quality Improvement Collaborative, aimed at accelerating the implementation of ROM in clinical practice (for details see paragraph ‘Intervention’). The study included a parallel group design with matched pairs of participating teams, in which a cluster randomized controlled trial (RCT) was embedded (Figure 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 if there were differences among three groups of clinicians (physicians, psychologists and nurses).

The participating specialized mental health care providers were each requested to enrol two similar teams. In total, twenty one 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. Six of the fourteen matched pairs were randomly, and eight pairs non-randomly assigned to either the intervention or the control condition. The randomisation of six matched pairs was conducted by an independent data manager\textsuperscript{20} (Dutch Trial Register, NTR5262) (Figure 1). The fourteen control teams conducted ROM ‘as usual’ and implemented the best practice, only after the ending of the study.

In the teams not participating in the randomised trial, 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. In any case, both teams were treating similar patient groups in the same geographical catchment area, just as the matched pairs of the randomized 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 outpatient treatment). For the study no patient involvement was required; thus no informed consent was needed.

\textbf{Figure 1.} Parallel group design with nested RCT
Intervention: National ROM Quality Improvement Collaborative
The collaborative promoted the routine use of clinical outcome questionnaires or rating scales at the beginning, during and at the end of 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 Quality Improvement Collaborative (QIC) program during one year. A QIC is a multi-faceted implementation strategy. 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 organized 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. 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 survey for clinicians at two moments: at the beginning (T0) and at the end (T1) of the QIC (after one year). Data collection took place independent of the Collaborative, by a data management team. Clinicians were invited and received a reminder by e-mail to fill out the survey. The results were processed anonymously, respondents were only labeled by team.

The survey had previously been developed by the Trimbos Institute, Netherlands Institute of Mental Health and Addiction. 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. Exploratory factor analysis demonstrated a four factor structure of the instrument:

- Individual use and perceived utility of ROM in daily practice, consisting of 8 items, for example ‘I use the ROM scores to evaluate the course of treatment’.
- Use of ROM in the team and organizational preconditions (7 items), for example ‘ROM scores are used in multidisciplinary consultations’.
- Usefulness of the ROM questionnaires (4 items), for example ‘The questionnaires are suitable for measuring change’.
- Accessibility of ROM for patient and clinician (3 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 $\alpha = .93$ and $\alpha = .91$. The Cronbach alphas of two other domains are good: ‘Use of ROM in the team and organizational preconditions’ $\alpha = .86$ and ‘Usefulness of the ROM questionnaires’ $\alpha = .86$. The internal consistency of the domain ‘Accessibility ROM for patient and clinician’ is less adequate ($\alpha = .51$). However, this scale was maintained in the survey, firstly because of the importance of the content of these items. According to implementation literature 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). Secondly, a Cronbach’s alpha $>.5$ is deemed just acceptable, with a minimum of three items contributing to the domain.

**Statistical Analysis**

Analysis was performed on the four subdomains of the survey and the total scale score. Data were analysed using SPSS for Windows, version 22. Firstly,
the number of teams, the drop outs of the study, response to the survey and
the composition of teams who responded to the survey, were described. Chi-
square tests were used to test potential differences in the composition of teams
between de intervention and control group. To calculate differences between
T0 and T1 and the difference at T1 between the intervention and control group
independent-sample t-tests were used, since clinicians of the participating
teams, who filled-out the survey at T0 and T1, were not always the same. Mean,
SD, Confidence Intervals and Effect sizes were computed. The effect sizes
were calculated by the following formula: Mpost-Mpre/SDpooled (because of
independent groups).

SD pooled = \sqrt{(SD1^2 + SD2^2)/2} using the effect size calculator for separate
groups of L. Becker, University of Colorado: http://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.\textsuperscript{26} We repeated the analysis
described above for the randomized 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, nurses).
Independent sample t-tests were used to calculate differences between T0 and
T1 for each group of clinicians separately. Differences between the groups of
clinicians on T0 and T1 were tested with analysis of variance (ANOVA) and post
hoc tests (Bonferroni).

\textit{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 T1
with T0. With $\alpha = 0.05$ and a power $\beta = 0.80$, the required sample size was 65
clinicians in the intervention group.\textsuperscript{27}

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

Parallel group design:
Twenty one teams from organizations of specialized mental health care across the country participated (see Figure 2a). In fourteen of them two similar teams were included. Flowchart 2a shows that, during the Collaborative, three teams dropped out between T0 and T1, mainly due to reorganizations and personnel changes in the participating teams.

At T0, in the intervention group 69% and in the control group 57% of the clinicians responded to the survey. The type of clinicians responding to the survey were in the intervention group: 11% physicians, 53% psychologists and 36% nurses, and in the control group: 21% physicians, 43% psychologists and 36% nurses. The composition between intervention and control teams did not differ significantly.

At T1, in the intervention group 89% and in the control group 62% of the clinicians responded to the survey. The composition of these groups, responding to the survey was in the intervention group: 25% physicians, 44% psychologists and 31% nurses and in the control group: 17% physicians, 40% psychologists and 43% nurses. As with T0, the differences in composition at T1 between intervention and control group were not significant.

Figure 2. Flow chart parallel group design (Flowchart 2a) and RCT design (Flowchart 2b)
Cluster Randomized Control design:
Figure 2, Flowchart 2b shows loss of data over time in the randomized teams. In total, clinicians of six intervention and six control teams filled out the survey. Between T0 and T1 one team dropped out, because of reorganization and personnel changes. At T0, in the intervention group 73% of the clinicians responded to the survey and in the control group 83% of the clinicians. The composition of the group of clinicians responding to the survey in terms of profession was in the intervention teams: 0% physicians, 65% psychologists and 35% nurses, and in the control teams: 13% physicians, 67% psychologists and 20% nurses.

At T1, in the intervention group 73% of the clinicians responded to the survey and in the control group 75%. The composition of these groups of clinicians at T1, responding to the survey was in the intervention group: 9% physicians, 58% psychologists and 33% nurses and in the control group: 0% physicians, 54% psychologists and 46% nurses. Both at T0 and T1, there were no significant differences in composition of clinicians between intervention and control group.

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, firstly the difference between first (T0) and final measurements (T1) of the intervention group is described. Secondly, we looked at the differences between intervention and control group at the end of the Collaborative (T1). The results are demonstrated for both the parallel group as the nested randomized design.

Differences between first and final measurements of the intervention group
Parallel group design:
In the intervention group significant positive differences were shown between T0 and T1 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 T0 and T1.
Cluster Randomized Control design:
The randomized group showed comparable results in the application of ROM in daily practice (Table 1). The effect sizes in the randomized 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 T1 between the intervention and control group 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 to respondents in the control group.

Cluster Randomized Control design:
While comparing the final measurements (T1), the above mentioned positive significant results in favor 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 T1 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 T1, compared to T0, significantly higher on the total score and the subdomain ‘Use of ROM in the team and organizational 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 T1 relative to T0.
Table 1. Changes in the intervention teams: T1 compared to T0 in parallel group design and nested RCT.

<table>
<thead>
<tr>
<th>SURVEY DOMAINS</th>
<th>PARALLEL GROUP DESIGN INTERVENTION TEAMS</th>
<th>CLUSTER RANDOMIZED CONTROL TRIAL INTERVENTION TEAMS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
</tr>
<tr>
<td>Individual use and perceived utility of ROM in daily practice</td>
<td>T0 91</td>
<td>3.28</td>
</tr>
<tr>
<td></td>
<td>T1  79</td>
<td>3.84</td>
</tr>
<tr>
<td>Use of ROM in the team and organizational preconditions</td>
<td>T0 91</td>
<td>2.59</td>
</tr>
<tr>
<td></td>
<td>T1  79</td>
<td>3.40</td>
</tr>
<tr>
<td>Usefulness of the ROM questionnaires</td>
<td>T0 91</td>
<td>2.95</td>
</tr>
<tr>
<td></td>
<td>T1  79</td>
<td>3.45</td>
</tr>
<tr>
<td>Accessibility ROM for patient and clinician</td>
<td>T0 91</td>
<td>2.95</td>
</tr>
<tr>
<td></td>
<td>T1  79</td>
<td>3.59</td>
</tr>
<tr>
<td>Total score ROM in daily practice</td>
<td>T0 91</td>
<td>2.94</td>
</tr>
<tr>
<td></td>
<td>T1  79</td>
<td>3.57</td>
</tr>
</tbody>
</table>

CI, confidence interval; RCT, randomised controlled trial; ROM, routine outcome monitoring; Sig., significance.
Table 2. Differences between intervention and control group at T1 in parallel group design and nested RCT (I = intervention and C = control group)

<table>
<thead>
<tr>
<th>SURVEY DOMAINS</th>
<th>PARALLEL GROUP DESIGN T1 INTERVENTION AND CONTROL</th>
<th>CLUSTER RANDOMIZED CONTROL TRIAL T1 INTERVENTION AND CONTROL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean T1</td>
</tr>
<tr>
<td>Individual use and perceived utility of ROM in daily practice</td>
<td>I</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>32</td>
</tr>
<tr>
<td>Use of ROM in the team and organizational preconditions</td>
<td>I</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>32</td>
</tr>
<tr>
<td>Usefulness of the ROM questionnaires</td>
<td>I</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>32</td>
</tr>
<tr>
<td>Accessibility ROM for patient and clinician</td>
<td>I</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>32</td>
</tr>
<tr>
<td>Total score ROM in daily practice</td>
<td>I</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>32</td>
</tr>
</tbody>
</table>

C, control group; CI, confidence interval; I, intervention group; RCT, randomised controlled trial; ROM, routine outcome monitoring; Sig., significance.
<table>
<thead>
<tr>
<th>SURVEY DOMAINS</th>
<th>NURSES</th>
<th>PSYCHOLOGISTS</th>
<th>PHYSICIANS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>s.d.</td>
</tr>
<tr>
<td>Individual use and perceived utility of ROM in daily practice</td>
<td>T0</td>
<td>26</td>
<td>3,00</td>
</tr>
<tr>
<td></td>
<td>T1</td>
<td>21</td>
<td>3,95</td>
</tr>
<tr>
<td>Use of ROM in the team and organizational preconditions</td>
<td>T0</td>
<td>26</td>
<td>2,49</td>
</tr>
<tr>
<td></td>
<td>T1</td>
<td>21</td>
<td>3,54</td>
</tr>
<tr>
<td>Usefulness of the ROM questionnaires</td>
<td>T0</td>
<td>26</td>
<td>2,90</td>
</tr>
<tr>
<td></td>
<td>T1</td>
<td>21</td>
<td>3,50</td>
</tr>
<tr>
<td>Accessibility ROM for patient and clinician</td>
<td>T0</td>
<td>26</td>
<td>2,58</td>
</tr>
<tr>
<td></td>
<td>T1</td>
<td>21</td>
<td>3,56</td>
</tr>
<tr>
<td>Total score ROM in daily practice</td>
<td>T0</td>
<td>26</td>
<td>2,74</td>
</tr>
<tr>
<td></td>
<td>T1</td>
<td>21</td>
<td>3,64</td>
</tr>
</tbody>
</table>

CI, confidence interval; ROM, routine outcome monitoring; Sig., significance.
At T0, compared to 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 = .006 and CI = -1.020 to -0.134). During the collaborative year the differences between these groups of clinicians were reduced. At T1, 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 Quality Improvement Collaborative aimed to accelerate the implementation of ROM in Dutch specialized mental health care. The study included a parallel group design with matched pairs of participating teams, in which a cluster randomized controlled trial (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 (Table 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 due to 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 organizational 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 organization 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 organizational 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 design (RCT) within a government sponsored national initiative to improve mental health care. We built on previous work in which the survey was developed. The teams experienced ownership of their improvement process and were facilitated by the National Quality Improvement Collaborative. A variety of teams with a multidisciplinary 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 randomized design separately. Strength of the parallel group design was the large external validity due to the number and variation of the participating teams. The randomized group included less 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 investigate the effects on decisional conflict of patients, working alliance, treatment adherence, clinical outcome and quality of live. 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 measurement 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.

Conclusions
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 health care 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 Measurement Based Care 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.
References


