This is a repository copy of 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.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/133490/

Version: Published Version

Article:

https://doi.org/10.1192/bjpo.bp.116.004366

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.
Measurement-based care (MBC)\(^1\) has beneficial effects on achieving response and remission of mental health disorders, such as depression.\(^1\) In addition, MBC can enhance effective communication between patients and clinicians and involvement of patients in clinical decision-making.\(^1\) Despite these promising prospects of MBC, the progress in the application of outcome measurement in routine mental healthcare is slow,\(^9\) because of the complexity of its implementation.\(^9\)

To promote outcome measurement in routine clinical practice in Dutch-specialised mental healthcare, a government-sponsored National Quality Improvement Collaborative was initiated in Dutch-specialised mental healthcare.\(^5\)

### Method

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

**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.
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. 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. After the Collaborative’s ending, the control teams are all offered the intervention.

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 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. 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 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 multidisciplinary 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, α=0.93 and α=0.91. The Cronbach’s alphas of the two other domains are good: ‘Use of ROM in the team and organisational preconditions’ α=0.86 and ‘Usefulness of the ROM questionnaires’ α=0.96. The internal consistency of the domain ‘Accessibility ROM for patient and clinician’ is less adequate (α=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 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.

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 T0 and T1 and the difference at T1 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: \( M_{post} - M_{pre} / SD_{pooled} \) (because of independent groups).

\[
SD_{pooled} = \sqrt{\frac{(SD_1^2 + SD_2^2)}{2}}
\]

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 \( \Delta = 0.5 \) 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.\(^2^7\)

### 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.
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 1** Changes in the intervention teams: $T_0$ compared with $T_1$ in parallel group design and nested RCT

<table>
<thead>
<tr>
<th>Survey domains</th>
<th>Parallel group design intervention teams</th>
<th>Cluster randomised control trial intervention teams</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$T_0$ Mean</td>
<td>s.d.</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>------------</td>
<td>------</td>
</tr>
<tr>
<td>Individual use and perceived</td>
<td>91</td>
<td>3.28</td>
</tr>
<tr>
<td>utility of ROM in daily practice</td>
<td>79</td>
<td>3.84</td>
</tr>
<tr>
<td>Use of ROM in the team and</td>
<td>91</td>
<td>2.59</td>
</tr>
<tr>
<td>organisational preconditions</td>
<td>79</td>
<td>3.40</td>
</tr>
<tr>
<td>Usefulness of the ROM</td>
<td>91</td>
<td>2.95</td>
</tr>
<tr>
<td>questionnaires</td>
<td>79</td>
<td>3.45</td>
</tr>
<tr>
<td>Accessibility ROM for</td>
<td>91</td>
<td>2.95</td>
</tr>
<tr>
<td>patient and clinician</td>
<td>79</td>
<td>3.59</td>
</tr>
<tr>
<td>Total score of the ROM in</td>
<td>91</td>
<td>2.94</td>
</tr>
<tr>
<td>daily practice</td>
<td>79</td>
<td>3.57</td>
</tr>
</tbody>
</table>

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

---

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

<table>
<thead>
<tr>
<th>Survey domains</th>
<th>Parallel group design $T_1$ intervention and control</th>
<th>Cluster randomised control trial $T_1$ intervention and control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean $T_2$</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>---</td>
<td>------------</td>
</tr>
<tr>
<td>Individual use and perceived</td>
<td>I</td>
<td>79</td>
</tr>
<tr>
<td>utility of ROM in daily practice</td>
<td>C</td>
<td>32</td>
</tr>
<tr>
<td>Use of ROM in the team and</td>
<td>I</td>
<td>79</td>
</tr>
<tr>
<td>organisational preconditions</td>
<td>C</td>
<td>32</td>
</tr>
<tr>
<td>Usefulness of the ROM</td>
<td>I</td>
<td>79</td>
</tr>
<tr>
<td>questionnaires</td>
<td>C</td>
<td>32</td>
</tr>
<tr>
<td>Accessibility ROM for</td>
<td>I</td>
<td>79</td>
</tr>
<tr>
<td>patient and clinician</td>
<td>C</td>
<td>32</td>
</tr>
<tr>
<td>Total score of the ROM in</td>
<td>I</td>
<td>79</td>
</tr>
<tr>
<td>daily practice</td>
<td>C</td>
<td>32</td>
</tr>
</tbody>
</table>

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

109
Table 3

<table>
<thead>
<tr>
<th>Survey domains</th>
<th>Psychologists</th>
<th>Nurses</th>
<th>Physicians</th>
<th>Effect size</th>
<th>Sig.</th>
<th>95% CI of the difference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Individual use and perceived utility</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Access to the ROM for patient</td>
<td>26</td>
<td>3.56</td>
<td>0.70</td>
<td>0.000</td>
<td></td>
<td>-1.31</td>
</tr>
<tr>
<td>and clinician</td>
<td>26</td>
<td>3.71</td>
<td>0.64</td>
<td>0.000</td>
<td></td>
<td>-1.30</td>
</tr>
<tr>
<td><strong>Usefulness of the ROM</strong></td>
<td>26</td>
<td>3.28</td>
<td>0.76</td>
<td>0.000</td>
<td></td>
<td>-1.44</td>
</tr>
<tr>
<td>and clinician</td>
<td>26</td>
<td>3.45</td>
<td>0.76</td>
<td>0.000</td>
<td></td>
<td>-1.52</td>
</tr>
<tr>
<td><strong>Accessibility ROM in daily practice</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total score of the ROM in daily practice</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**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. The teams experienced ownership of their improvement process and were facilitated by the National QIC. 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 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. 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.

References
9 van der Feltz-Cornelis CM, Andres H, Keesels E, Oudewinkeloord HI, Bemans 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: 376–84.
10 Van Der Wees PJ, Nijhuis van der Sanden MWG, Ayanian JZ, Black N, Wessert GP, Schneider EC. Integrating the use of patient-reported outcomes for both clinical performance and practice measurement: views of experts from 3 countries. Milbank Q 2014; 92: 754–75.


