

# VU Research Portal

## Shared decision making in mental health care

Metz, M.J.

2018

### **document version**

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

### **citation for published version (APA)**

Metz, M. J. (2018). *Shared decision making in mental health care: the added value for patients and clinicians*.

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

### **E-mail address:**

[vuresearchportal.ub@vu.nl](mailto:vuresearchportal.ub@vu.nl)

# 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.

British Journal of Psychiatry Open 2017; 3: 106-112.

## **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)<sup>1,2</sup> has beneficial effects on achieving response and remission of mental health disorders, such as depression.<sup>1-6</sup> In addition, MBC can enhance effective communication between patients and clinicians and involvement of patients in clinical decision making.<sup>1,5,7-9</sup> Despite these promising prospects of MBC, the progress in the application of outcome measurement in routine mental health care is slow,<sup>10,11</sup> due to the complexity of its implementation.<sup>12-16</sup>

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.<sup>17-20</sup> 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).<sup>5,9,14,21</sup> 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<sup>20</sup> (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.

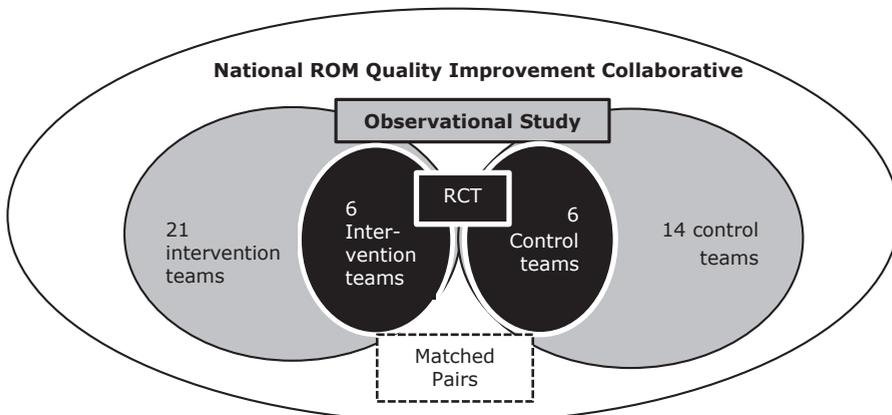


Figure 1. Parallel group design with nested RCT

**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.<sup>17-19</sup> 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.<sup>19,22,23</sup> 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<sup>24</sup> 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.<sup>24</sup> 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.<sup>24</sup> 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<sup>1,5,12,14,16,21</sup> 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.<sup>25</sup>

### **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 the 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:  $M_{\text{post}} - M_{\text{pre}} / SD_{\text{pooled}}$  (because of independent groups).

$SD_{\text{pooled}} = \sqrt{((SD_1^2 + SD_2^2)/2)}$  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.<sup>26</sup> 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).

### ***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.<sup>27</sup>

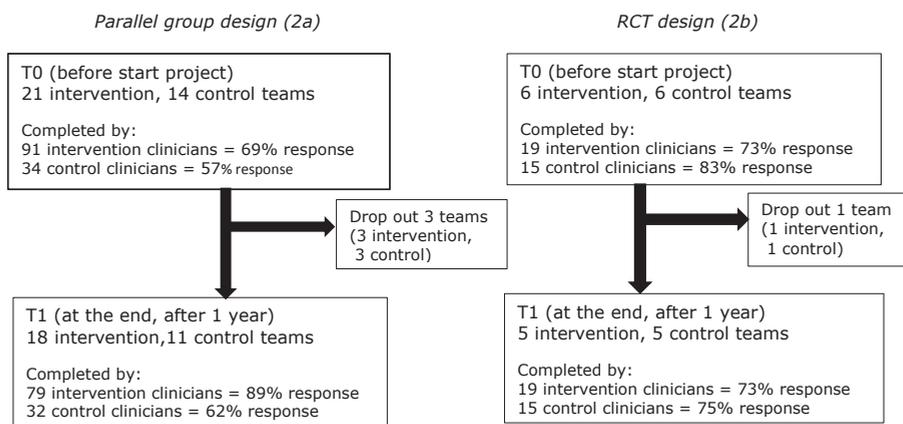
## **Results**

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 , Flowchart 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.



**Figure 2.** Flow chart parallel group design (Flowchart 2a) and RCT design (Flowchart 2b)

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.

***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.

SURVEY DOMAINS	PARALLEL GROUP DESIGN INTERVENTION TEAMS							CLUSTER RANDOMIZED CONTROL TRIAL INTERVENTION TEAMS							
	N	Mean	s.d.	Effect size	Sig. t-tailed	95% CI of the diff.		N	Mean	s.d.	Effect size	Sig. t-tailed	95% CI of the diff.		
						Lower	Upper						Lower	Upper	
<b>Individual use and perceived utility of ROM in daily practice</b>	T0	91	3,28	1,01	0,62	,000	-0,84	-0,29	19	3,22	1,07	1,11	,002	-1,47	-0,36
	T1	79	3,84	0,80					19	4,14	0,47				
<b>Use of ROM in the team and organizational preconditions</b>	T0	91	2,59	0,85	1,02	,000	-1,05	-0,57	19	2,59	0,80	1,14	,001	-1,36	-0,37
	T1	79	3,40	0,74					19	3,45	0,71				
<b>Usefulness of the ROM questionnaires</b>	T0	91	2,95	0,85	0,55	,000	-0,77	-0,22	19	3,07	0,96	0,97	,005	-1,35	-0,26
	T1	79	3,45	0,95					19	3,87	0,68				
<b>Accessibility ROM for patient and clinician</b>	T0	91	2,95	0,72	0,88	,000	-0,86	-0,42	19	2,96	0,97	1,07	,002	-1,39	-0,33
	T1	79	3,59	0,74					19	3,82	0,59				
<b>Total score ROM in daily practice</b>	T0	91	2,94	0,64	0,99	,000	-0,82	-0,43	19	2,96	0,83	1,25	,000	-1,31	-0,41
	T1	79	3,57	0,63					19	3,82	0,51				

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)

SURVEY DOMAINS	PARALLEL GROUP DESIGN T1 INTERVENTION AND CONTROL						CLUSTER RANDOMIZED CONTROL TRIAL T1 INTERVENTION AND CONTROL						
	N	Mean T1	s.d.	Sig. t-tailed	95% CI of the diff.		N	Mean T1	s.d.	Sig. t-tailed	95% CI of the diff.		
					Lower	Upper					Lower	Upper	
<b>Individual use and perceived utility of ROM in daily practice</b>	I	79	3,84	0,80	,000	0,52	1,20	19	4,14	0,47	,000	0,74	1,72
	C	32	2,98	0,87				15	2,91	0,81			
<b>Use of ROM in the team and organizational preconditions</b>	I	79	3,40	0,74	,000	0,44	1,08	19	3,45	0,71	,005	0,24	1,25
	C	32	2,64	0,86				15	2,70	0,74			
<b>Usefulness of the ROM questionnaires</b>	I	79	3,45	0,95	,008	0,15	0,95	19	3,87	0,68	,011	0,19	1,32
	C	32	2,90	0,99				15	3,12	0,94			
<b>Accessibility ROM for patient and clinician</b>	I	79	3,59	0,74	,000	0,28	0,94	19	3,82	0,59	,001	0,45	1,55
	C	32	2,98	0,92				15	2,82	0,97			
<b>Total score ROM in daily practice</b>	I	79	3,57	0,63	,000	0,42	0,97	19	3,82	0,51	,000	0,50	1,36
	C	32	2,88	0,73				15	2,89	0,72			

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

Table 3. Results T1 compared to T0 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 diff.	N	Mean	Std. Dev.	Effect size	Sig. t-tailed	95% CI of the diff.	N	Mean	s.d.	Effect size	Sig. t-tailed	95% CI of the diff.				
<b>Individual use and perceived utility of ROM in daily practice</b>	T0	26	3,00	1,23	0,98	,001	-1,52	-0,40	39	3,50	0,85	0,57	,023	-0,85	-0,07	8	2,89	1,13	,200	-1,47	0,33	
	T1	21	3,95	0,63			30	3,96	0,76				17	3,46	0,96							
<b>Use of ROM in the team and organizational preconditions</b>	T0	26	2,49	1,05	1,11	,001	-1,61	-0,48	39	2,56	0,69	1,17	,000	-1,15	-0,48	8	2,13	0,85	1,51	,001	-1,84	-0,51
	T1	21	3,54	0,82			30	3,37	0,70				17	3,30	0,70							
<b>Usefulness of the ROM questionnaires</b>	T0	26	2,90	0,79	0,68	,024	-1,11	-0,08	39	3,00	0,87	0,59	,018	-0,94	-0,09	8	2,63	1,11	,148	-1,70	0,27	
	T1	21	3,50	0,96			30	3,52	0,90				17	3,34	1,11							
<b>Accessibility ROM for patient and clinician</b>	T0	26	2,58	0,82	1,28	,000	-1,44	-0,52	39	3,15	0,64	0,87	,001	-0,87	-0,25	8	2,75	0,64	,155	-1,21	0,20	
	T1	21	3,56	0,70			30	3,71	0,64				17	3,25	0,86							
<b>Total score ROM in daily practice</b>	T0	26	2,74	0,76	1,28	,000	-1,31	-0,48	39	3,05	0,54	1,13	,000	-0,84	-0,33	8	2,60	0,74	0,97	,037	-1,43	-0,05
	T1	21	3,64	0,64			30	3,64	0,49				17	3,34	0,80							

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.<sup>24</sup> 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 life.<sup>20</sup> 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

1. Fortney JC, Unützer J, Wren G, Pyne JM, Smith GR, Schoenbaum M, et al. A tipping point for measurement based care. *Psychiatr Serv.* 2016; September 01: 1-10. <http://dx.doi.org/10.1176/appi.ps.201500439>
2. 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(1) 28-40.
3. 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(10): 1004-13.
4. 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 (1): 15-22.
5. Carlier IVE, Meuldijk D, Vliet van IM, Fenema van EM, Wee van der 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.
6. Davidson K, Perry A, Bell L. Would continuous feedback of patient's clinical outcomes to clinicians improve NHS psychological therapy services? Critical analysis and assessment of quality of existing studies. *Psychol Psychother.* 2015; 88(1): 21-37.
7. Valenstein M, Adler DA, Berlant J, Dixon LB, Dulit RA, Goldman B, et al. Implementing standardized assessments in clinical care: now's the time. *Psychiatric Services.* 2009; 60: 1372-75.
8. 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.
9. Feltz van der-Cornelis C, Andrea H, Kessels E, Duivenvoorden H, Biemans H, Metz M. Shared Decision Making in combinatie met ROM bij patiënten met gecombineerde lichamelijke en psychische klachten; een klinisch empirische verkenning. *Tijdschr Psychiatr.* 2014; 56(6): 375-84.
10. Wees van der PH, 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(4): 754-75.
11. Delespaul PEG. Routine Outcome Measurement in the Netherlands: A focus on benchmarking. *Int Rev Psychiatry.* 2015; 27(4): 320-28. <http://dx.doi.org/10.3109/09540261.2015.1045408>
12. 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. <http://dx.doi.org/10.1080/10503307.2013.817696>
13. Jong de K, Sluis van P, Nugter AM, 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(4): 464-74.
14. Jong de 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(6): 629-39.
15. Jong de K. Challenges in the implementation of measurement feedback systems. *Adm. Policy. Ment. Health.* 2016; 43: 467-70.
16. 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. <http://dx.doi.org/10.1186/1472-6963-12-96>
17. Schouten LMT, Hulscher MEJL, Everdingen van JJE, Huijsman R, Grol RPTM. Evidence for the impact of quality improvement collaboratives: systematic review. *BMJ.* 2008; 28: 336(7659):1491-95.
18. 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. Utrecht, 2012.
19. Øvretveit J, Bate P, Cleary P, Cretin S, Gustafson D, McInnes K, et al. Quality collaboratives: lessons from research. *Qual Saf Health Care.* 2002; 11(4): 345-51.

20. Metz MJ, Franx GC, Veerbeek MA, Beurs de E, Van der Feltz-Cornelis CM, Beekman ATF. 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. <http://dx.doi.org/10.1186/s12888-015-0696-2>
21. Beurs de E, Hollander den-Gijsman ME, Rood van YR, Wee van der NJ, Giltay EJ, Noorden van 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): 1-12. <http://dx.doi.org/10.1002/cpp.696>
22. Berwick DM. Developing and testing changes in delivery of care. *Ann Intern Med*. 1998; 15(128): 651-57.
23. Splunteren van P, Everdingen van 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. Assen: Koninklijke van Gorcum, 2003 (in Dutch).
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-instituut. Utrecht 2014 [www.trimbos.nl/producten-en-diensten/webwinkel/product/?prod=af1383](http://www.trimbos.nl/producten-en-diensten/webwinkel/product/?prod=af1383)
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: 1181-1209.
27. Lipsey MW. *Design sensitivity: statistical power for experimental research*: 137. Newbury Park: Sage, 1990.